# Yield to Temptation
# A Comparison of Present-Biased Preferences in Continuous Time

Patrick Schneider*

May 7, 2025

**Abstract**

I compare two ways of modelling present bias in dynamic consumption-saving problems: quasi-hyperbolic discounting (QHD) and temptation preferences. I show how to implement temptation preferences in continuous time in a general way, where the temptation is to adopt an alternative discount function. Doing so allows for a direct comparison to results for the continuous-time limit of QHD ('instantaneous gratification' preferences (Harris and Laibson, 2013; Maxted, 2024)). I show that QHD consumers are behaviourally equivalent to naively tempted consumers, but that welfare is not equivalent. The models differ in various ways when consumers are sophisticated, and variation in the degree of sophistication offers opportunities for identification between the two models. Temptation preferences are more flexible than QHD, and so offer a promising path forward for incorporating present–bias into macro models.

*Imperial Business School: p.schneider@imperial.ac.uk

# 1 Introduction

A growing body of evidence suggests behavioural biases are important for capturing consumption dynamics within macroeconomic models. Among these biases, a tendency to over–emphasise immediate gratification (present bias) is particularly prominent. Present bias is typically modelled with competing selves frameworks in one of two approaches.

The first and most prevalent approach is quasi-hyperbolic discounting, formalised by Laibson (1997) but stemming from earlier time–inconsistency models (Phelps and Pollak, 1968; Strotz, 1955). Here, consumers disproportionately value immediate rewards compared to future ones, causing short–term over consumption. While intuitive, quasi-hyperbolic discounting poses practical challenges, mainly due to the built in time inconsistency which complicates solutions. Perhaps due to these challenges, they have not been widely adopted in macro models, despite substantial evidence that present–bias is a relevant force (one notable exception is Maxted et al., 2024).

The second approach is temptation preferences, formalised by Gul and Pesendorfer (2001) and implemented in a consumption–saving setting in various papers (e.g. Attanasio et al., 2024; Gul and Pesendorfer, 2004; Kovacs et al., 2021). Here, consumers face costly temptations toward irresponsible consumption, and can mitigate these costs by giving in a little, leading again to short–term over consumption. This method is inherently more flexible and analytically convenient than the former, but has been used less in practice.

Despite their longstanding coexistence, there have been few comparisons of these two approaches, potentially because both are complex, but in different ways. Recent work has improved the tractability of quasi-hyperbolic discounting models by recasting them into continuous–time, termed 'Instantaneous Gratification' (IG) preferences (Harris and Laibson, 2013; Maxted, 2024). This work has demonstrated that, under some important assumptions, (1) IG consumers' policy functions are scale multiples of their rational equivalents, and (2) their value functions are positive affine transformations of their rational equivalents. These results make it much easier to work with and understand these preferences.

In this paper, I make the parallel extension of temptation preferences into continuous–time. The formulation I use is very general, and unifies various discrete–time temptation models by specifying that consumers are tempted to adopt an alternative (less responsible) discount functions. I use the new formulation to make a direct comparison of tempted and IG consumers' behaviour and welfare.

I show that IG preferences are behaviourally equivalent to *naive* temptation preferences, irrespective of the tempting discount function assumed.[1] Temptation preferences are therefore the more general option, because (a) they will deliver identical behaviour

[1] A similar equivalence is noted in O'Donoghue and Loewenstein (2004) and Fudenberg and Levine (2006), albeit in the much more limited setting of two–period models with restricted temptation preferences.

under naivete, and (b) relaxing the assumption of naivete leads to qualitatively different, and observable, behaviour that differs to consumers with IG preferences.

Temptation preferences are also more flexible than IG because the existence of solutions does not depend on functional form assumptions. All of the recent results in IG models are based on the assumption of constant relative risk aversion (CRRA) utility. While IG may be possible with other utility functions, these have not yet been developed, and it is not clear that they can be. As a result, IG preferences can only currently be used in a very narrow setting that excludes, for example, additive labour supply disutility. By contrast, temptation preferences can work with *any* utility function, and as a result are more easily implemented in a range of macroeconomic models.

While the IG model is nested within temptation in behavioural terms, they are not welfare equivalent. The welfare criterion in IG is the long–run preferences over biased behaviour. As Maxted (2024) shows, this will be a positive affine transformation of the rational agent's welfare function under his identifying assumptions. By contrast, temptation preferences work because the presence of temptation *harms* consumers—they are not biased in the sense that they're wrong about what they feel, just relative to what a rational agent would do. These preferences are tractable because they're time–consistent (sophisticated tempted consumers never regret their behaviour), but they achieve this consistency by distorting welfare instead. Because of this, whilst the two models can be behaviourally equivalent if consumers are naive, they will yield very different policy implications, and they will also yield different behaviour if consumers are sophisticated. This last fact presents opportunities for distinguishing between the two empirically.

The rest of the paper is structured as follows. In Section 2 I give an overview of the research on consumption's sensitivity to current income, in order to motivate the importance of behavioural preferences in macroeconomic models. In Section 3 I detail the basic consumption–saving model that will be the core of our analysis in the paper, and solve it for the rational benchmark case. In Section 4 I give an overview of IG preferences, the continuous time equivalent of the most widespread model of present–biased behaviour: quasi–hyperbolic discounting. And in Section 5 I introduce temptation preferences and show how to derive them in continuous–time in a general way. In Section 6 I use these results to compare behaviour and welfare under IG and temptation preferences, establishing naive behavioural equivalence and the different roles that sophistication plays. Section 7 concludes.

# 2 The case for present–bias in macro models

The aggregate marginal propensity to consume (MPC) is a central object in macro–policymaking, and an important target for models seeking to analyse stabilisation policy. The MPC measures the sensitivity of consumption to current income and so governs how responsive consumption will be to fiscal transfers, one of the key stimulus tools available to governments.[2] The magnitude of this MPC has been a topic of contention, but recent evidence suggests that incorporating behavioural bias into theoretical models of consumption is necessary to match what we observe.

In the initial Keynesian analysis of government multipliers, the MPC was imposed by assuming consumption was a declining fraction of income, and the aggregate was an average of this across the population. In the 1950s through 1970s, the assumption was upset by theoretical contributions by Friedman (1957) and Modigliani and Brumberg (1954) which micro–founded consumption as the result of a forward looking problem by rational agents. This work led to the permanent income–life cycle hypothesis (PIH) that people will use savings to smooth through income fluctuations, and so current consumption should be a function of expected lifetime wealth. As a result, current consumption should be effectively independent from changes in current income, and more so if these changes were temporary, anticipated, or expected to be offset by future taxes with equivalent present value. The theoretical result implied that consumption should follow a random walk, supported empirically in Hall (1978). The PIH led to a strong policy prescription against Keynesian stimulus programs: efforts to juice the economy by redistributing from people saving resources to those who would spend them were futile, because *no–one* would spend them.

And yet they do. Starting in the 1980s, a vast literature has documented the 'excess sensitivity' of consumption to current income, in violation of the PIH. Time–series analyses showed that Hall (1978)'s result was upended once better econometric models were used (Campbell and Mankiw, 1989; Flavin, 1981). But the main evidence against the PIH has come from micro–data analysis of individuals' consumption sensitivity to changes in current income. The survey in Jappelli and Pistaferri (2010) gives a good overview of much of this literature, which tends to find current consumption *is* responsive to current income, whether it was predictable or not, but that there is substantial variation across the population in how responsive it is. This variation provides some clues as to what might be driving the apparently sub–optimal behaviour.

One of the most influential results is that individual MPCs are related to liquid resources, with less liquid people having higher MPCs (Baker et al., 2023; Fagereng et al., 2019; Gelman et al., 2022; Jappelli and Pistaferri, 2014; Karger and Rajan, 2020; Toczynski, 2023; Zeldes, 1989). This observation motivated the buffer–stock models of consumption and precautionary savings, starting in the 1990s (Aiyagari, 1994; Carroll et al., 2021; Carroll and Kimball, 1996; Deaton, 1989). In these models, households are subject to severe idiosyncratic risk that they can only partially insure due to borrowing constraints. The fact of incomplete markets leads households that are close to their

[2] MPCs can be measured in response to expected income at any horizon, termed inter–temporal MPCs in Auclert et al. (2024b). I will focus here on the MPC out of current income for brevity.

constraints to ration the resources they do have to avoid a sudden drop in consumption if they receive a bad shock. This rationing reflects short–term concern about hitting the constraint overwhelming long–term concerns about inter–temporal substitution or consumption smoothing i.e. their decisions are driven by a 'precautionary' motive. And in doing so it makes consumption much more sensitive to current income.

These models are the core of a recent wave of macro–models that seek to match aggregate MPCs (Auclert et al., 2024a; Bayer et al., 2019; Kaplan et al., 2018). They tend to feature households that are close to borrowing constraints for two reasons—some are poor, and some are wealthy but have most of their assets in illiquid form like housing or retirement accounts (Kaplan et al., 2014). These models rationalise investments in the illiquid asset with return differences: the illiquid asset pays such a good premium that rational households invest heavily in good times, only regretting it if they receive a bad shock or uncertainty increases. While these two–asset heterogeneous agent models have become the work–horse model for generating endogenous realistic MPCs, they are not a panacea. One issue, for example, is that the calibrated premium the illiquid account pays is implausibly high when compared to the return differences households actually face between genuinely liquid and illiquid options.[3]

The liquidity explanation of high MPCs has also been challenged by a range of recent papers identifying behaviour that is not consistent with precautionary saving. These include the observations that: measured MPCs can only be rationalised by implausibly low exponential discount factors (Gelman, 2022; Gerard and Naritomi, 2021; Hamilton et al., 2024; Laibson, 1997; Shapiro, 2005), people with substantial liquidity often have high MPCs anyway (Baugh et al., 2021; Boehm et al., 2025; Kueng, 2018; Olafsson and Pagel, 2018), consumption responds to current predictable income *losses* (which are insurable with saving) as well as gains (Baugh et al., 2021; Ganong and Noel, 2020; Gerard and Naritomi, 2021; Ni and Seol, 2014), people simultaneously carry high interest credit card debt and contribute to illiquid savings (Laibson et al., 2024), (lack of) liquidity is too predictable to be merely circumstantial (Parker, 2017), and measured MPCs are higher for credit that expires earlier, despite being fungible (Boehm et al., 2025).

This wealth of evidence suggests behavioural biases may be a necessary to addition to our toolbox if we want to explain household decisions. While a number of biases may be operating, much of the above behaviour can be explained with distortions to discounting captured by models of present–bias, the focus of the present paper.[4] In what follows

[3]Alternative theories for why people pile into illiquid but low return options appeal to other convenience–yields these assets might offer: owning housing allows one to escape the landlord risk, or over–withholding taxes might guard against the penalties that come with under–withholding when income is uncertain (Gelman et al., 2022). One such convenience yield I deal with in this paper is that illiquid assets don't tempt us to over–consume (Kaplan and Violante, 2022).

[4]Others in contention are distortions in expectation formation, and mental accounting or bounded rationality. Each of these, as well as temptation preferences, has in common that some part of decision marking is difficult, and the consumer must both make optimal decisions as well as exert effort over

I detail the two main approaches to modelling present bias: quasi–hyperbolic discounting and temptation preferences. My contribution is to derive temptation preferences in continuous time in a general way for the first time, and use these results to do a direct comparison with Instantaneous Gratification preferences, the continuous time limit of quasi–hyperbolic discounting.

I am not the first to consider such a comparison. Krusell et al. (2010) build a model of temptation in which the tempting alternative is to adopt quasi–hyperbolic discounting such that the latter is the special case as the cost of temptation becomes infinite. My approach is much more general, and admits the Krusell et al. (2010) model as a special case. But by reframing the model in continuous time I am able to show that the temptation model nests quasi–hyperbolic discounting under *any* assumption about the tempting alternative if (a) the consumer is completely naive, and (b) the identifying assumptions in Maxted (2024) hold.

Toussaert (2018) tests for whether biased consumers suffer temptation or hyperbolic discounting in the lab. She identifies the difference by noting that temptation preferences lead people to seek commitment against strictly dominated options, whereas quasi–hyperbolic discounters will only seek commitment from options they would choose. She establishes that a substantial portion of the population is likely to experience temptation, providing evidence that this is the better model of present–bias to adopt. My work is theoretical, showing how these preference structures alter decisions in a consumption–saving framework. My behavioural equivalence result suggests that it will be difficult to tell the behaviour of these agents apart unless they are sophisticated. Identification schemes like the ones used in Toussaert (2018) may provide a useful way to approach identification in macro environments in future.

# 3 Consumption–saving model

The remainder of the paper compares different ways of modelling present bias in the same economic environment. This section details the environment—a continuous time consumption–saving problem where households have two assets, one a liquid transaction account in which borrowing is always possible but increasingly costly, and the other an illiquid saving account, and are exposed to idiosyncratic risk. This section starts by detailing the choices and constraints that households face in the general model environment, and I then solve the model for a household with rational preferences to serve as a benchmark.

their decision making architecture—potentially limiting information gathering and processing, aiming for approximate optima, or reducing the pain from temptation.

## 3.1 The model

Households are differentiated by their liquid wealth $b$, illiquid wealth $a$, and labour productivity $z$.[5] Together, these three variables describe a household's 'state', and I sometimes group them together in the vector $x = (b, a, z)$. The model is in continuous time, and changes in these states over time are governed by the following processes:

$$\frac{\partial b}{\partial t} = r(b)b + (1-\xi)wz - c - d - \chi(d,a) \tag{1}$$

$$\frac{\partial a}{\partial t} = r_a a + \xi y + d \tag{2}$$

$$\frac{\partial \ln z}{\partial t} = -\theta_z \ln z + \sigma_z \frac{\partial W_t}{\partial t} \tag{3}$$

That is, additions to the liquid balance come from asset returns $r(b)b$ and labour income $wz$, net of the automatic illiquid contributions $\xi$, and reductions come from consumption $c$, voluntary contributions to the illiquid account ($d$), and the transaction costs these attract $\chi(d, a)$. The liquid asset return $r(b)$ is able to assume different values for different liquid balances. I assume it is fixed for positive balances at $r_b$ and adopt the following assumption for negative balances.

**Assumption 3.1** (*Soft borrowing constraint*)**.** The liquid return is continuous and weakly decreasing for negative balance $-r'(b) \geq 0 \ \forall b < 0$, and becomes very large close to the natural borrowing limit.

Such a return function encodes a 'soft borrowing constraint' by deterring borrowing rather than banning it. Doing so ensures solutions are interior, which is helpful for some of the results below (Maxted, 2024). And the assumption is also realistic. Hard borrowing constraints are rare: extra credit is usually available as long as the borrower is willing to accept increasingly onerous terms e.g. credit card or payday loan rates are followed, in the limit, by risk of violence or criminal liability (Lee and Maxted, 2023).[6]

The illiquid balance evolves with asset returns $r_a a$, mandatory contributions $\xi wz$, and voluntary contributions or withdrawals $d$. And log–labour productivity $z$ follows an Ornstein–Uhlenbeck process around a stationary mean of $\bar{z} = 1$.

Within this environment, households choose a sequence of consumption and transfer plans between accounts $(\mathbf{c}, \mathbf{d})$ to optimise a general value function, which is composed of

[5]The model mirrors the setup in Kaplan et al. (2018) where households have access to a liquid and illiquid account, and face idiosyncratic risk. Relative to the original it is simplified in that the labour income process is an endowment.

[6]Despite being available, it's likely that these extra sources of credit come with greater costs, potentially both fixed and variable. Here I assume that these costs are all encoded in the interest rate, and this is important. If the costs were instead incurred prior to extracting the credit then they may reinstate hard constraints.

the expected discounted integral of the flow of utility

$$v_t(x) = \max_{\mathbf{c},\mathbf{d}} \mathbb{E}_t \int_0^\infty D(s) u(c_s(x_{t+s})) ds$$

Where the policy at a particular point in time $c_s(x)$ is a subset of the sequence of plans $\mathbf{c}$. In the stationary case, policies are the same at any point in time and $(\mathbf{c}, \mathbf{d}) = (c(x), d(x))$. $D(s)$ is the discount function applied to time $s$ into the future. In the next section, I solve for the rational benchmark case.

### 3.2 Rational benchmark

In the rational benchmark, households have CRRA preferences and an exponential discount function $D(s) = e^{-\rho s}$. Hence stationary their value is

$$v(x_t) = \mathbb{E}_t \int_0^\infty e^{-\rho s} \left( \frac{c(x_{t+s})^{1-\sigma}}{1-\sigma} \right) ds$$

And the solution to their problem is defined by the Hamilton–Jacobi–Bellman (HJB) equation[7], state transition equations (1–3), and first order conditions

$$\rho v(x) = u(c(x)) + \partial_b v(x) \cdot \dot{b}(x) + \partial_a v(x) \cdot \dot{a}(x) + \mathcal{A}[v](x) \tag{4}$$

$$c(x)^{-\sigma} = \partial_b v(x) \tag{5}$$

$$\partial_a v(x) = \partial_b v(x)(1 + \chi_d(d(x), a)) \tag{6}$$

Where $\mathcal{A}[\cdot]$ is the infinitesimal generator of the income process. Note that under Assumption 3.1 the solution is interior, and so we do not need to account for the liquid state–constraint binding. In the following sections I detail how quasi–hyperbolic discounting and temptation models distort the value, and how naivete alters households' expectations of these distortions going forward.

# 4 Quasi–hyperbolic discounting preferences

The most widespread model of present–biased behaviour is the quasi–hyperbolic discounting model. This dates back to the time–inconsistency models of Strotz (1955) and Phelps and Pollak (1968), which grapple with problems where individuals or societies place disproportionate weight on the present when making dynamic choices. These were formalised for consumption–savings problems in Laibson (1997), in which the problem is boiled down to a distorted present–value function, which places an extra discount on all discrete entries except the current one. This so called $\beta - \delta$ model discounts a flow of utilities coming from a consumption plan $\{c_0, c_1, ..., c_T\}$ as follows

$$u(c_0) + \beta \cdot \sum_{t=1}^{T} \delta^t u(c_t)$$

Where $\delta < 1$ is a standard exponential discount factor between periods, and $\beta \leq 1$ captures the degree of present–bias. The starting point for these models was to formalise time–inconsistency, but it has been shown that this phenomenon leads to over–consumption and, potentially, commitment–seeking behaviour where these agents anticipate inconsistency and attempt to bind their future selves' hands.

## 4.1 Quasi–hyperbolic discounting in continuous–time

The model can be extended to its continuous–time limit (where the future is the next instant) and doing so yields to greater tractability, as well as clearer welfare results. In this limit, quasi–hyperbolic discounting is referred to as 'Instantaneous Gratification' (IG) preferences. They were first introduced and explored at length in Harris and Laibson (2013), with complementary analysis in Maxted (2024). What follows in this section is an overview of the main results from these papers.

Instantaneous Gratification (IG) preferences capture quasi–hyperbolic discounting in the continuous time limit, where there is a time–dependent discount function (Harris and Laibson, 2013)

$$D^{\beta}(s) = \begin{cases} 1 & \text{if } s = 0 \\ \beta e^{-\rho s)} & \text{if } s > 0 \end{cases}$$

That is, all time beyond the current instant is discounted at a rate of $\beta < 1$ on top of the standard exponential discounting. This discounting of the future makes the preferences time–inconsistent, as in the discrete–time formulation of the model. Assuming sophistication, i.e. the consumer correctly anticipates future bias, the IG consumer's current value $w^{\beta}(x)$ in the stationary solution is

$$\begin{aligned} w^{\beta}(x_t) &= \max_{c^{\beta}(x), d^{\beta}(x)} \mathbb{E}_t \int_0^{\infty} D^{\beta}(s) u(c^{\beta}(x_{t+s})) ds \\ &= \lim_{\Delta \to 0} \max_{c^{\beta}, d^{\beta}} u(c^{\beta})\Delta + \beta e^{-\rho\Delta}\, \mathbb{E}_t\, v^{\beta}(x_{t+\Delta}(c^{\beta}, d^{\beta})) \end{aligned} \tag{7}$$

Where the maximisation is subject to the state transition equations (1–3), and $v^\beta(x)$ is the long–run value the biased consumer places on their *future* expected stream of consumption, i.e. it anticipates but does not adopt future bias. Converting this into an HJB, by taking a second–order approximation around $\Delta = 0$, yields

$$\beta\rho v^\beta(x) = \max_{c,d} u(c) + \beta\left[\partial_b v^\beta(x)\cdot\dot{b} + \partial_a v^\beta(x)\cdot\dot{a} + \mathcal{A}[v^\beta](x)\right] - \frac{w(x_t) - \beta v^\beta(x)}{dt} \tag{8}$$

A solution requires that the current value satisfy $w(x) = \beta v^\beta(x)$, otherwise the final term in Equation 8 will explode. We can see why intuitively in Equation 7, where the first term goes to zero in the continuous–time limit i.e. if your value depends on current pleasure and discounted anticipated future pleasure, then the closer the future is, the more the latter term will dominate. The IG consumer's optimal choices are given by the FOC

$$u'(c^\beta(x)) = \beta\partial_b v^\beta(x) \tag{9}$$
$$\partial_a v^\beta(x) = \partial_b v^\beta(x)(1 + \chi_d(d^\beta, a)) \tag{10}$$

Based on the FOC, we can find the policy functions if we know $v^\beta(x)$. But this is where things get difficult for IG preferences. Due to the time–inconsistency, we cannot use Equation (8) to solve for $v^\beta(x)$ as we usually would.

**The $\hat{u}$ construction** Harris and Laibson (2013) show that while the value for this household does exist, and is unique, it cannot be estimated using any of the above equations. Instead, they show that one can recover $v^\beta(x)$ by solving a different problem for a rational agent (i.e. an exponential discounter) with a distorted felicity function $\hat{u}(\cdot)$. Assuming $u(\cdot)$ is CRRA over consumption, there is a particular $\hat{u}(\cdot)$ function that delivers a solution equal to the long–run value of the biased household i.e. $\hat{v}(x) = v^\beta(x)$. This value can be solved using standard methods because this agent's problem is time–consistent Achdou et al. (2022); Harris and Laibson (2013). And with this value in hand, we can use the IG agent's FOC (Equations 9 and 10) to find their policy functions.

Maxted (2024) builds on this result by showing that under the assumption of a soft borrowing constraint (Assumption 3.1), the $\hat{u}$ function simplifies to a positive affine transformation of the CRRA utility function.

$$\hat{u}(c) = \frac{\psi}{\beta}\frac{\left(\frac{1}{\psi}c\right)^{1-\sigma}}{1-\sigma} \quad \text{where} \quad \psi = \frac{\sigma - (1-\beta)}{\sigma}$$

This leads to two results (1) the $\hat{u}$ agent's value, and therefore $v^\beta(x)$, is a positive affine transformation of the rational value $v(x)$, and (2) the $\hat{u}$ household's behaviour is identical to a rational agent with a standard utility function's behaviour (because affine transformations don't upset ordinal rankings, choices will not differ).

**Relation to rational benchmark** These results allow us to express the IG agent's optimal behaviour as analytical functions of the rational benchmark derived earlier, obviating the need for the $\hat{u}(\cdot)$ gymnastics at all. To see this for the consumption function, for example, note from combining Equations 9 and 5 that the IG consumption policy satisfies

$$u'(c^\beta) = \beta \frac{\partial_b v^\beta}{\partial_b v} u'(c)$$

Given we know the IG agent's long–run value is a positive affine transformation of the rational one, the ratio $\frac{\partial_b v^\beta}{\partial_b v}$ is determined by the scaling factor in this transformation. In this case the scaling factor is $\frac{\partial_b v^\beta}{\partial_b v} = \frac{\psi^\sigma}{\beta} > 1$, and the relationship between the consumption policies is as follows

$$c^\beta(x) = \left(\frac{\sigma - (1-\beta)}{\sigma}\right)^{-1} c(x)$$

Hence, as long as $\sigma > (1-\beta)$ and $\sigma > 1$, there is a well–behaved relationship with the IG consumption increasing, relative to the rational benchmark, as the present bias becomes more severe $\downarrow \beta$, or as the elasticity of inter–temporal substitution $(1/\sigma)$ increases.

Similar logic leads to an even cleaner result for the transfer policy function $d^\beta(x)$, where the scaling factor appears on both sides of Equation 10, and cancels out. As a result, the IG agent makes the same asset allocation decision as the rational benchmark

$$d^\beta(x) = d(x)$$

As discussed in Maxted (2024), this result makes intuitive sense because asset allocation decisions compare different impacts on the future. Quasi–hyperbolic discounting distorts evaluations of decisions the span the future and the present, so it is natural that decisions not involving the present are not distorted.

This result implies that *IG agents do not pursue commitment devices*, a strong result relative to the existing literature, but one that has some empirical support (Laibson, 2015). The lack of commitment seeking rests on Assumption 3.1, which implies no amount of squirreling away funds in an illiquid account can actually bind the biased agent's consumption in future, because there are always other sources of credit to tap. As a result, they don't seek out such devices.

## 4.2 Extension to naivete

All of the discussion so far has presumed that households are sophisticated. That is they (correctly) anticipate that their bias will continue in the future. We can relax this assumption by introducing naivete, where households (incorrectly) expect their future selves to be subject to a degree of bias $\beta^E$ that is less severe than their current level.

All of the above results hold with slight alterations to make a distinction between effects coming from current versus anticipated bias. I use the tilde notation to denote naivete. First, the current value is now

$$\tilde{w}^{\beta}(x_t) = \lim_{\Delta \to 0} \max_{\{\tilde{c}^{\beta}, \tilde{d}^{\beta}\}} u(\tilde{c}^{\beta})\Delta + \beta e^{-\rho\Delta}\, \mathbb{E}_t\, v^{\beta^E}(x_{t+\Delta}(\tilde{c}^{\beta}, \tilde{d}^{\beta})) \tag{11}$$

Where $v^{\beta^E}(x) = \int_0^{\infty} e^{-\rho s} u\left(c^{\beta^E}(x_s)\right) ds$ is the long–run value placed on the consumption plan of an agent with expected future bias $\beta^E$. The transfer policy $\tilde{d}^{\beta}(x)$ is unaffected by bias, real or imagined, and so is unchanged

$$\tilde{d}^{\beta}(x) = d(x) \tag{12}$$

and the consumption policy is

$$\tilde{c}^{\beta}(x) = \left(\frac{\beta}{\beta^E}\right)^{-\frac{1}{\sigma}} \left(\frac{\sigma - (1-\beta^E)}{\sigma}\right)^{-1} c(x) \tag{13}$$

This policy function embeds two interesting extremes. Under sophistication $\beta^E = \beta$ the function reduces to the sophisticated case derived earlier. Under complete naivete $\beta^E = 1$ the IG agent's consumption scales the rational benchmark by $\beta^{-\frac{1}{\sigma}}$.

We can use this formula to quantify the impact of naivete on consumption behaviour. Consider a stylised example where $(\sigma, \beta) = (2, 0.7)$, which are within the range of empirical estimates for the parameters. Using Equation 13 we can see that a sophisticated agent will consume 18% more than the rational benchmark, compared to 20% more for a complete naif. Naivete therefore increases consumption. The effect is quite small in this example, but it is larger the more intense the present bias.[8]

As all of these cases are scale multiples of the rational benchmark policy, they are also isomorphic to each other. Indeed if Assumption 3.1 holds, then any degree of sophistica-

[8]The sign of this effect actually depends on parameters. Per Maxted (2024) "The intuition for this result is that naivete introduces two offsetting effects. On the one hand, the naif is more willing to save because the naif trusts their future selves. On the other hand, the naif is less willing to save because the naif believes that future selves will save enough on their own. The former effect dominates when the agent is relatively more willing to substitute intertemporally ($\sigma < 1$), and vice versa." (footnote 34, p. 28) Empirical estimates of $\sigma$ place it well above 1, and so we can be confident that naivete should lead to lower consumption in the present.

tion is equivalent to a completely naive agent with a reduced discount.

**Lemma 4.1** (*Inverse of Maxted (2024) Corollary 5*)**.** An IG agent with any degree of bias or sophistication parameters

$$(\beta, \beta^E) \in [0, 1] \times [\beta, 1]$$

consumes identically to a completely naive IG agent with a bias parameter $\beta^*$ that satisfies

$$\beta^* = \frac{\beta}{\beta^E} \left( \frac{\sigma - (1 - \beta^E)}{\sigma} \right)^{\sigma}$$

*Proof.* This results from setting the scaling factor in Equation 13 equal to $(\beta^*)^{-\frac{1}{\sigma}}$ □

Lemma 4.1 implies that sophistication cannot be identified in consumption data alone under these preferences.

**Comment on inflexibility** These results mean the IG agent's policy functions can be found in a simple two–step process of (1) solving the rational agent's problem, and (2) backing out the IG agent's policies. This avoids the need to actually use the $\hat{u}$ function itself in the process of solving the model, but the results do rely on this construction.

As a result, dealing with IG preferences necessitates first finding the appropriate $\hat{u}$ function. Harris and Laibson (2013) backward–engineer it to deliver the value function equivalence for CRRA utility over consumption. But a different $\hat{u}$ is needed for any potential $u(\cdot)$. Other $\hat{u}$ constructions have been shown to exist. For example Maxted (2024) derives the new $\hat{u}$ function when the utility function is a CRRA wrapper around a Cobb–Douglas bundle of consumption and housing durable, and shows that the extension is relatively straightforward.

It's currently unclear, however, if these functions exist for more complicated problems—for additive disutility from labour, for example. For the moment, working with IG preferences can be cumbersome unless we either (a) stay within the bounds of CRRA utility wrappers, or (b) assume complete naivete (noting that the isomorphism between naive and sophisticated behaviour described above depends on the assumption of both CRRA utility and non–binding constraints). An alternative approach is to instead use *time–consistent* preferences that also induce present biased behaviour.

# 5 Temptation preferences

Temptation preferences are an alternative way of modelling present bias that are time–consistent. In this model, we imagine people derive utility from their choices, but this satisfaction is coloured by what their alternatives were. It is as though while making their choices, a demon stood on their shoulder whispering to them to be less responsible. While the demon doesn't succeed in driving behaviour, the act of resistance is costly. The cost depends on the gap in temptation utility between what you imagined and what you did, and so in the moment people can alleviate the cost by giving in a little.[9] This 'giving in' can mean over–consuming if the problem is a consumption–saving one. And if people recognise this pattern of behaviour in themselves (i.e. if they're sophisticated), they may take actions to limit what they can be tempted by in the future (i.e. seeking commitment devices), reducing the power these demons have to sway decisions in future.

Whilst temptation is at least as old as humanity, we owe the preference formulation to Gul and Pesendorfer (2001). In temptation preferences, people have rankings over choice sets as well as options within them and, if particular axioms hold, then temptation preferences in their most general form can be rationalised by the following value function

$$V(\mathcal{K}) = \max_{x \in \mathcal{K}} h(x) - \left[ \max_{y \in \mathcal{K}} k(y) - k(x) \right] \tag{14}$$

Where the choices $(x, y)$ are made from the consumer's choice set $\mathcal{K}$, $x$ is the actual action taken and $y$ is the tempting alternative the demon suggests. $h(x)$ represents the utility the path chosen generates, and the term in square brackets represents the cost of temptation: the difference between the temptation utility $k(\cdot)$ experienced by choosing the 'most tempting alternative' $y$ compared to the actual choice $x$, evaluated with the tempting discount function $k(\cdot)$. With this structure, a consumer may prefer the limited choice set $\mathcal{S} \subset \mathcal{K}$ because the latter includes options that, while not chosen, will cause harm by giving the demon some ammunition.[10]

The key to using these preferences is being specific about what exactly the consumer is most tempted by i.e. the form of the temptation utility $k(\cdot)$. In this paper we are concerned with consumption–savings problems in which the choice is over streams of consumption flows and the choice–set is the budget–set. As I will show, it's natural for the temptation to be an alternative discount function.

**Discrete–time example** Before getting to the continuous–time derivation, we can illustrate how temptation works in a consumption–savings setting using the linear version of Dynamic Self–Control (DSC) preferences (Attanasio et al., 2024; Fudenberg and Levine, 2006; Gul and Pesendorfer, 2004; Kovacs et al., 2021; O'Donoghue and Loewen-

[9] Parents of small children may for example resist buying toys in the gift shop at museum exits by promising a smaller, cheaper toy from the supermarket on the way home.

[10] This feature of preferences is called 'set–betweeness', and is one of the new axioms introduced by Gul and Pesendorfer (2001) that permits a preference for commitment. The standard example of this demand for commitment (preferring limited choice–sets) in practice is the harm felt by recovering addicts of being around drug–users *even if* they manage not to relapse, because the willpower required to resist is itself costly. Standard preferences do not feature this because more options always weakly improve welfare.

stein, 2004).

In this situation consumers are tempted to discount the future entirely, and their temptation utility based on a consumption stream $\{c_t\}_{t=0}^{\infty}$ therefore ignores all future consumption $k(c) = \lambda u(c)$. In a simple one–asset setup with interest rate $r$, stochastic income $y$ and discount factor $\delta$, the Bellman equation is

$$\begin{aligned} V(x) &= \max_c u(c) - \lambda\left[u((1+r)b+y) - u(c)\right] + \delta\,\mathbb{E}\left[V(x')\right] \\ b' &= (1+r)(b-c) + y' \end{aligned}$$

Here the most–tempting feasible plan is to consume all available cash on hand $c^* = (1+r)b+y$, presuming no borrowing is allowed. When $\lambda = 0$ these preferences embed the rational case. When $\lambda > 0$ the closer actual consumption is to exhausting resources the less the temptation cost is felt, and so a rational consumer will choose greater consumption than their un–tempted equivalent to relieve this pressure. We can rearrange the Bellman equation slightly to show two separate forces introduced by temptation

$$V(x) = \max_c \,(1+\lambda)u(c) - \lambda u((1+r)b+y) + \delta\,\mathbb{E}\left[V(x')\right] \tag{15}$$

Here we can see that temptation ($\lambda > 0$) introduces two effects—it boosts the value of present–day consumption by $1+\lambda$, and it reduces felicity by an amount that depends on cash–in–hand. This latter force means the disutility from the most–tempting alternative introduces a force akin to negative money in the utility function models (Poterba and Rotemberg, 1986) i.e. it makes liquid assets relatively less attractive for a given financial return by imposing a sort of inconvenience yield. In cases where there are alternative storage options that are not tempting, this force will lead consumers to store less wealth in the transacting asset and more in those alternatives. That is, they will seek out commitment devices.

## 5.1 Temptation in continuous–time

In this section I extend the temptation model to continuous–time in a general way, based on the consumption–saving problem introduced in Section 3 and where I make two assumptions

**Assumption 5.1.** The costs of temptation comparisons is linear.

**Assumption 5.2.** Consumers are tempted by an alternative discount function

$$D(s) \neq e^{-\rho s}$$

The first is a common simplifying assumption (e.g. in Attanasio et al., 2024; Fudenberg and Levine, 2006; Kaplan and Violante, 2022; Kovacs et al., 2021; Krusell et al., 2010) and the second allows us to embed various forms of temptation preferences in a general framework.

To start, define the inputs into the temptation value (Equation 14). The choice set contains all the feasible consumption and deposit plans for each point in the state–space. The standard value placed on a consumption and deposit plan $(\mathbf{c}, \mathbf{d})$ that respects state–transitions (where bold text represents a stream of consumption plans into the future) is

$$h(\mathbf{c}, \mathbf{d}, x) = \mathbb{E} \int_0^\infty e^{-\rho s} u(c(x_{t+s}, s)) ds$$

and the temptation value place on the same plan under our two assumptions is

$$k(\mathbf{c}, \mathbf{d}, x) = \lambda \, \mathbb{E} \int_0^\infty D(s) u(c(x_{t+s}, s)) ds$$

The temptation value differs from the standard in two ways: consumption flows may be discounted differently to standard $D(s) \neq e^{-\rho s}$, and they receive a linear boost $\lambda \geq 0$. Combining these, we can define the value function for a general tempted consumer as

$$v^\lambda(x) = \max_{\mathbf{c},\mathbf{d},\hat{\mathbf{c}},,\hat{\mathbf{d}}} \mathbb{E} \int_0^\infty e^{-\rho s} u(c(x_{t+s}, s)) ds - \lambda \, \mathbb{E} \int_0^\infty D(s) \left[ u(\hat{c}(x_{t+s}, s)) - u(c(x_{t+s}, s)) \right] ds$$

This setup embeds various different approaches to temptation preferences that differ in the discount function the consumer finds tempting.

**Rational benchmark** First note that these preferences embed rationality as a special case. The linear boost $\lambda$ determines how intensely temptation is felt, and can be used to switch it off entirely: in the case where $\lambda = 0$, the above resolves to the rational value.

Alternatively, if consumers are tempted by their options in the same way as they actually evaluate them, i.e. they set the tempting discount function to the exponential one $D(s) = e^{-\rho s}$, then the above resolves to the rational value as well because the optimal $c(s) = \hat{c}(s)$ and so there is no difference between the tempting plan and the actual one.

For temptation to have any effect, there must therefore be some alternative way of viewing options $k(c) \neq \lambda \cdot h(c)$ *and* the cost of resisting temptation must not be zero $\lambda > 0$.

**Greater discounting** In Kaplan and Violante (2022) people are tempted to adopt a much higher discount rate. We can encode that here with $\lambda > 0$ and the exponential discount function with $\hat{\rho} > \rho$

$$D(s) = e^{-\hat{\rho} s}$$

These preferences yield the following HJB equation, derived in Appendix A.2

$$
\begin{aligned}
\rho v^\lambda(x) = \max_{c^\lambda, d^\lambda} \, &(1+\lambda)u(c^\lambda) - \lambda u(\hat{c}(x)) + (\hat{\rho} - \rho)\left(\hat{k}(x) - k(\mathbf{c}^\lambda, \mathbf{d}^\lambda, x)\right) \\
&+ \partial_b v^\lambda(x) \cdot \dot{b} + \partial_a v^\lambda(x) \cdot \dot{a} + \mathcal{A}\left[v^\lambda\right](x)
\end{aligned}
$$

Where $\hat{k}(x) = k(\hat{\mathbf{c}}, \hat{\mathbf{d}}, x)$ is the temptation value of the tempting consumption and deposit plan, and the latter are solved in a separate problem with the discount rate $\hat{\rho}$. Solving the general problem amounts to treating it as a time–consistent problem with the distorted felicity function

$$
(1+\lambda)u(c(x)) - \lambda u(\hat{c}) + (\hat{\rho} - \rho)\left(\hat{k}(x) - k(\mathbf{c}^\lambda, \mathbf{d}^\lambda, x)\right)
$$

There are two distortions here. The first is from currently felt temptation utility, and the second is from the anticipated future temptation and the clash of discount rates used to evaluate this future. In a stationary solution, $k(\mathbf{c}^\lambda, \mathbf{d}^\lambda, x)$ is simply a function of the chosen policies. In a dynamic solution, this object will depend on the sequence of consumption plans going forward. Similarly, $\hat{k}(x)$ depends on the solution to the tempting problem, which will depend on the anticipated sequence of prices and shocks in dynamic solutions. From the consumer problem's perspective, these are based on future behaviour and so they are not taken into account in solving the problem for the present.[11] As a result, they do not enter the FOC and so optimal choices will satisfy

$$
\begin{aligned}
(1+\lambda)u'(c^\lambda(x)) &= \partial_b v^\lambda(x) \\
\partial_a v^\lambda(x) &= \partial_b v^\lambda(x)(1 + \chi_d(d^\lambda(x), a))
\end{aligned}
$$

The distortions do impact the value itself, however, and so the optimal choices will be distorted by the anticipation of future temptation. With the effect of current temptation reducing the relative value of the liquid asset and so driving more consumption and more transfers out of the liquid account.

**Hyperbolic discounting** In Krusell et al. (2010) consumers are tempted to be quasi–hyperbolic discounters. We can encode continuous–time equivalent here with $\lambda > 0$ and the IG discount function discussed in Section 4

$$
D(s) = \begin{cases} 1 & \text{if } s = 0 \\ \beta \cdot e^{-\rho s} & \text{else} \end{cases}
$$

[11]As the derivation in Appendix A.2 shows they are the result of taking the limit $\lim_{\Delta \to 0} k(\mathbf{c}_\Delta, \mathbf{d}_\Delta, x_{t+\Delta})$

These preferences reduce to the following HJB equation (see Appendix A.2)

$$\rho v^{\lambda}(x) = \max_{c^{\lambda},d^{\lambda}} (1+\lambda)u(c^{\lambda}) - \lambda u(\hat{c}(x)) + \partial_b v^{\lambda}(x) \cdot \dot{b} + \partial_a v^{\lambda}(x) \cdot \dot{a} + \mathcal{A}\left[v^{\lambda}\right](x)$$

And where $\hat{c}(x)$ is the solution to the sophisticated IG problem discussed in the previous section i.e. it is some multiple of the rational agent's consumption policy. With this tempting consumption policy in hand $(v^{\lambda}, c^{\lambda}, d^{\lambda})$ can be solved the usual way because the above is a time–consistent problem for an agent with a distorted felicity function. These preferences embed *IG* in the limit as resistance to temptation becomes infinitely costly $\lambda \to \infty$ (Krusell et al., 2010).

**Ignoring the future** Most implementations of temptation preferences in the consumption–saving literature have used the DSC preferences introduced by Gul and Pesendorfer (2004) in discrete–time. This is where people are tempted to ignore the future entirely (Attanasio et al., 2024; Fudenberg and Levine, 2006; Gul and Pesendorfer, 2004; Kovacs et al., 2021; O'Donoghue and Loewenstein, 2004).

We can encode this in continuous–time with $\lambda > 0$ and the piecewise discount function

$$D(s) = \begin{cases} e^{-\rho s} & \text{if } s \leq \Delta^{\lambda} \\ 0 & \text{if } s > \Delta^{\lambda} \end{cases}$$

That is, future utility flows are discounted as normal, but only over a limited span $\Delta^{\lambda}$. This is a little different to DSC preferences because continuous–time doesn't admit a clean distinction between now and the future. In the limit as $\Delta^{\lambda}$ goes to zero, the tempting action is to ignore all but the current instant, the tempting consumption rate explodes, and so actual consumption would explode to meet it if $\lambda > 0$. To keep the most tempting alternative finite it's therefore necessary to fix a positive span of time over which the tempted consumer's imagination wanders. For example, a natural assumption would be to mirror the discrete time literature and set $\Delta^{\lambda}$ equal to one unit of the model frequency e.g. one year.

As well as having a positive $\Delta^{\lambda}$ it's necessary to impose a terminal condition that rules out Ponzi schemes for the tempting consumption plan, particularly if there is no limit to potential credit, as in Assumption 3.1. To this end, I assume that consumers are not tempted to incur debt when they start with a positive liquid balance, nor are those already in debt tempted to increase it.[12]

$$\hat{b}(\Delta^\lambda) \geq \begin{cases} 0 & \text{if } b(0) \geq 0 \\ b(0) & \text{else} \end{cases}$$

Note that there is a relationship between $\Delta^\lambda$ and $\lambda$. If the tempting party is over a longer time–span ($\uparrow \Delta^\lambda$), the most–tempting flow is lower as the starting stock of assets is spread over a longer time, so some adjustment $\uparrow \lambda$ that restores the correct intensity is needed. Knowing the duration that a consumer is tempted by is an impossible identification challenge. Given this, we should expect to recover different values of $\lambda$ from the same empirical exercise measuring behaviour at different frequencies, unless the most–tempting alternative is spread out over the same span.[13]

For given assumptions about the temptation parameters $(\lambda, \Delta^\lambda)$ these preferences reduce to the following HJB equation (see Appendix A.2)

$$\rho v^\lambda(x) = \max_{c,d} \, (1+\lambda)u(c) - \lambda u(\hat{c}) + \partial_b v^\lambda(x) \cdot \dot{b} + \partial_a v^\lambda(x) \cdot \dot{a} + \mathcal{A}\left[v^\lambda\right](x) \tag{16}$$

Where $\hat{c}(x)$ is the solution to the cake–eating problem over the limited span $\Delta^\lambda$, funded in part by the most tempting voluntary transfer function $\hat{d}(x)$. As with the hyperbolic discounting case, once we know this most–tempting alternative consumption policy, solving for $(v^\lambda, c^\lambda, d^\lambda)$ is simple.[14] The FOC yield the policy functions, and the HJB provides the solution for the value.

$$\begin{aligned} (1+\lambda)u'(c^\lambda(x)) &= \partial_b v^\lambda(x) \\ \partial_a v^\lambda(x) &= \partial_b v^\lambda(x)(1 + \chi_d(d^\lambda(x), a)) \end{aligned}$$

The HJB Equation 16 is the continuous–time analogue to Equation 15. The force of temptation boosts the utility experienced in the current moment, but also drags it down because of the tempting consumption alternative $\hat{c}$. This alternative will be closely related to cash on hand, and so this second force acts like negative money–in–utility function, making the liquid asset relatively less attractive for non–pecuniary reasons and so increasing (decreasing) the voluntary contributions to (withdrawals from) the illiquid account.

The second and third of the above cases have the same equations describing the

[12] This assumption can be justified with reference to mental accounting—people only daydream about spending money they can see, for example, or at least their demons don't tempt them with becoming criminals to fund their short party.

[13] The empirical exercises in Attanasio et al. (2024) and Kovacs et al. (2021) both use a frequency of one year, and find similar estimates of the temptation parameter.

[14] In practice, I have found a finite–difference scheme using the nested–drift algorithm (Sabet and Schneider, 2024) works well, but that it is important to start with a guessed value that assumes no temptation effect on utility, and to let the algorithm discover this force as it updates.

solution: the HJB is in the same form and policy functions determined by the same FOC. The only difference is the most–tempting consumption plan $\hat{c}(x)$. This will be the case for any tempting alternative discount function that discounts future increments at the rate $\rho$. With tempting discount rates that don't match this, the extra adjustment term in the first case is necessary, complicating the solution. In what follows, I will focus on the simpler case.

**Relation to rational benchmark** These results allow us to express the tempted agent's optimal behaviour relative to the rational benchmark policies defined earlier. Focusing on the consumption function, the FOC implies

$$u'(c^\lambda(x)) = \frac{1}{1+\lambda}\frac{\partial_b v^\lambda(x)}{\partial_b v(x)}u'(c(x))$$

Marginal utility of the tempted consumer is therefore distorted away from their rational equivalent by two forces—first, it is lower due to the pressure of current temptation $1/(1+\lambda) \leq 1$ and, second, it is lower due to the anticipation of future temptation reducing the value of carrying resources forward $\frac{\partial_b v^\lambda(x)}{\partial_b v(x)} \leq 1$. Both forces drive the consumer to over–consume relative to the rational benchmark.

The latter force has the opposite sign in IG preferences, where self–awareness reduces the impact of present bias, and the drive to over–consume. By contrast, with temptation preferences, the sophisticate over–consumes for two reasons. First, to relieve the cost of resisting his own temptation, and second to reduce the same torment for himself in the future. I will revisit this in Section 6 because it offers an opportunity for identifying between the two preference structures.

## 5.2 Extension to naivete

As with IG preferences, we can relax the assumption that tempted consumers correctly anticipate their future temptation by introducing naivete, where households expect their future selves to be subject to some lesser degree of temptation.

Naivete is difficult to define in the temptation model in continuous time. This is because of the potential for a clash between the horizon over which naivete is defined (separating the present from future) and over which the tempting alternative is defined (which depends on the tempting discount function). One approach could be to use a time–dependent temptation intensity.[15]

$$\lambda(s) = \begin{cases} \lambda & \text{if } s = 0 \\ \lambda^E & \text{else} \end{cases}$$

Where $\lambda^E$ captures the degree of temptation experienced in the future. But this on its own leads to pathological results in cases approaching complete naivete i.e. $\lambda^E = 0$, because tempting plans will blow up as the temptation is to entirely ignore the future, leading to infinite $c$ to avoid the cost of that temptation. To avoid this complexity, I make the following assumption

**Assumption 5.3.** A naif of any degree expects to be sophisticated in future with $\lambda^E \leq \lambda$. In the present, their most–tempting actions are those of a sophisticate with their current degree of temptation $\lambda$.

This assumption leads to the general value function

$$\tilde{v}^\lambda(x_t) = \lim_{\Delta\to 0} \max_{\tilde{c}^\lambda, \tilde{d}^\lambda} \left(u(\tilde{c}^\lambda) - \lambda\left[u(\hat{c}) - u(\tilde{c}^\lambda)\right]\right)\Delta + e^{-\rho\Delta}\,\mathbb{E}\,v^{\lambda^E}(x_{t+\Delta})$$

This assumption implies a contradiction in the consumer's thinking: they feel bias in the present and do not expect it to continue, but they *do* expect it to continue in the off–equilibrium path their demon is offering, from which $\hat{c}$ is derived. Under this assumption, the degree of naivete or sophistication does not affect the tempting plan felt in the moment of deciding ($\hat{c}$), just the perceived continuation value. This assumption is useful because it helps arrive at a solution, but it introduces time inconsistency into the calculation of the tempting plan $c(s)$, and so we cannot derive these naive temptation preferences from the general form.

The solution to this problem is the HJB and first order conditions

$$\begin{aligned}
\rho v^{\lambda^E}(x) &= (1+\lambda)u(\tilde{c}^\lambda(x)) - \lambda u(\hat{c}(x)) + \partial_b v^{\lambda^E}(x)\cdot\dot{b}(x) + \partial_a v^{\lambda^E}(x)\cdot\dot{a}(x) + \mathcal{A}[v^{\lambda^E}](x) \\
u'(\tilde{c}^\lambda(x)) &= \frac{1}{1+\lambda}\partial_b v^{\lambda^E}(x) \\
\partial_a v^{\lambda^E}(x) &= \partial_b v^{\lambda^E}(x)(1+\chi_d(\tilde{d}^\lambda(x), a))
\end{aligned}$$

Focusing on the consumption policy, the more naive an agent is, the more marginally valuable liquid resources will seem going forward, and so optimal consumption will reduce. Similarly for voluntary transfers to the illiquid account—naivete about future temptation

[15] O'Donoghue and Loewenstein (2004) introduce naivete in their model in the same way, noting that it could arise from genuine misunderstanding, or alternatively from time–inconsistency akin to quasi–hyperbolic discounting. Under that latter channel, the current consumer feels and responds to temptation, but doesn't care that their future selves will feel this as well when making decisions.

will cause consumers to under–estimate the inconvenience yield of future liquid resources, and they will keep relatively more resources in liquid form going forward.

**Complete naivete** Consider now the special case of complete naivete, where a consumer feels tempted in the moment, but anticipates that they'll be rational in future so $\lambda(s) = 0$ if $s > 0$ and as a result $v^{\lambda^E}(x_{t+\Delta}) = v(x_{t+\Delta})$. In this special case the consumer's optimal choices will be the same under *any* assumption about the tempting discount function.

**Lemma 5.4** (*Tempting naive behavioural equivalence*)**.** If Assumption 5.3 holds, then different models of temptation, i.e. assumptions about $D(s)$, are behaviourally equivalent when the tempted consumer is completely naive.

*Proof.* The value function of completely naive consumers is

$$\tilde{v}^{\lambda}(x_t) = \lim_{\Delta \to 0} \max_{c,d} \left(u(c) - \lambda \left[u(\hat{c}) - u(c)\right]\right) \Delta + e^{-\rho\Delta}\, \mathbb{E}\, v(x_{t+\Delta})$$

Where $v(x_{t+\Delta})$ is the rational valuation of the future states. The cost of temptation in the present – $\lambda u(\hat{c})$ – is sunk. This is the only term in the naif's value that would be affected by the tempting discount function $D(s)$. Therefore the form of this tempting discount function cannot alter optimal choices. □

This result will be useful in comparing the different models of present–bias.

# 6 Identification between present–bias models

How could we tell the difference between these two models? Which is the more appropriate to use in macro–economic models? Now that I have defined both types of preferences in continuous time, I compare results and look for ways they differ. I start with the simplest case of complete naivete, and show the strong result that the two approaches are behaviourally equivalent, albeit not welfare equivalent. Identification between them, therefore, relies on the behaviour of sophisticates. I then extend to partial or full sophistication, showing that behaviour diverges, with tempted consumers seeking commitment devices more, and changing their consumption in different ways to IG consumers.

These results hold under the assumptions in Maxted (2024) that the borrowing constraint never binds, and CRRA utility. If we relax the first then IG consumers will seek commitment devices but only if they constrain behaviour, whereas tempted consumers will seek commitment devices that restrain their demons, and so will place more of a premium on commitment devices than IG consumers. These differences under sophistication offer opportunities for identifying between the two present–bias formulations if sophistication is measurable.

## 6.1 Naivete and identification

Assume that our agents are completely naive. That is, if they are IG consumers then $\beta^E = 1$, and if they're tempted then $\lambda^E = 0$. Under this assumption, there is a relationship between their bias parameters that delivers exactly the same behaviour.

**Proposition 6.1** (*Naive Behavioural Equivalence*)**.** An IG consumer with future discount $\beta < 1$ and a tempted consumer with cost of temptation $\lambda > 0$ and tempting alternative consumption $\hat{c}(x)$ will make the same choices if (a) they are both completely naive, and (b) their parameters are related like this[16]

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

*Proof.* The naive tempted value is a location–scale transformation of the naive IG value. To see this, note the IG consumer's value function for given choices $(c, d)$ is

$$\tilde{w}^{\beta}(x_t) = \lim_{\Delta \to 0} u(c)\Delta + \beta e^{-\rho\Delta}\, \mathbb{E}_t\, v(x_{t+\Delta}(c, d))$$

And the tempted consumer's equivalent is

$$\tilde{v}^{\lambda}(x_t) = \lim_{\Delta \to 0} \left[(1+\lambda)u(c) - \lambda u(\hat{c}(x_t))\right]\Delta + e^{-\rho\Delta}\, \mathbb{E}_t\, v(x_{t+\Delta}(c, d))$$

If $\beta = 1/(1+\lambda)$ then these values are location–scale transformations of each other

$$\tilde{v}^{\lambda}(x) = (1+\lambda)\tilde{w}^{\beta}(x) - \lambda u(\hat{c}(x_t))dt$$

At the choices $(c, d)$ all marginals are the same. This is true of any choices $(c, d)$, and so it is true of the optimal choices as well. Therefore optimal choices will be the same. □

Intuitively, naive agents share a continuation value of resources, and only differ in the source of their present bias. In the moment, their decisions are guided by a boost to momentary utility (if tempted) or a reduced emphasis on the future (if IG). The relative evaluation of the present versus the future is therefore the same: emphasising the present more or the future less are two sides of the same coin.

**Corollary 6.2** (*Nesting*)**.** Adopt the Maxted (2024) assumptions: (a) CRRA utility, and (b) non–binding borrowing constraints. There is always a naive tempted agent that behaves identically to an IG agent with *any* degree of sophistication.

[16] Fudenberg and Levine (2006) and O'Donoghue and Loewenstein (2004) both note a similar same relationship in a two–period setting, where the continuation value is the same by definition. This result is more general, applying to infinite–horizons with exponential discounting, and I also show in following text that it implies nesting of the IG model within the temptation model.

*Proof.* This stems from the combination of Proposition 4.1 that sophistication in IG preferences is isomorphic to naivete with a different discount, and Proposition 6.1. $\square$

Under the Maxted (2024) assumptions, any behaviour (sophisticated or naive) under the IG model is nested by the temptation model as a special case. Note this holds regardless of the form of temptation adopted. This follows from Result 5.4 that temptation models are behaviourally equivalent to each other under the assumption of naivete. Krusell et al. (2010) showed that quasi–hyperbolic discounting can be nested within temptation models when consumers were actually tempted by this discount function. While this remains true, I have shown a stronger result that IG is nested under complete naivete with $\beta = 1/(1+\lambda)$, regardless of the tempting alternative discount function.

**Welfare** While the two approaches are behaviourally equivalent, they are not welfare equivalent. They seem to be from the perspective of the agents due to their naivete, but the appropriate welfare criterion to adopt with naive behavioural preferences erases this naivete. Taking the optimal policies as given $(c^*, d^*)$, the appropriate welfare criteria are

$$
\begin{aligned}
w^\beta(x_t) &= \beta\,\mathbb{E}\int_0^\infty e^{-\rho s} u(c^*(x_{t+s}))ds \\
v^\lambda(x_t) &= (1+\lambda)\,\mathbb{E}\int_0^\infty e^{-\rho s}\left[u(c^*(x_{t+s})) - \frac{\lambda}{(1+\lambda)}u(\hat{c}(x_{t+s}))\right] ds
\end{aligned}
$$

These are not at all the same. The stream of utility coming from the consumption plan are equivalent up to a multiple. But the tempted consumer experiences actual harm from the temptations their budget set offers them, distorting the consumer's welfare, and more so the greater the distance between the tempting and actual consumption $\hat{c}(x)$ and $c(x)$.

**Identification** The non–equivalence in welfare is important for policy–making as it will alter how we evaluate different options. But it is useless for identifying which of the two preferences is the better model to use. If people are genuinely naive, then observable variables (behaviour) will be identical. Identifying the difference between these preferences therefore hinges on the behaviour of sophisticates.

## 6.2 Sophistication and identification

Assume now that our agents are only partially naive. That is, if they are IG consumers then $\beta^E \in [\beta, 1)$ and if they're tempted then $\lambda^E \in (0, \lambda]$. The presence of some awareness of bias changes these consumers' behaviour. In the IG case, it does so in a way that is isomorphic to remaining naive but having a lower discount rate. In the tempted case the effect is more pronounced.

**General Euler equation** The difference is apparent in the general Euler equation describing consumption growth for the tempted consumer.

**Lemma 6.3** (*Tempted Euler equation*)**.** Assume an agent is naive with $\lambda^E \leq \lambda$, and utility is CRRA over consumption with relative risk aversion $\sigma$. Whenever the consumption function is locally differentiable in the liquid asset, it satisfies the Euler

$$E\left[\frac{\dot{\tilde{c}}^{\lambda}(x)}{\tilde{c}^{\lambda}(x)}\right] = \frac{1}{\sigma}\left[r - \rho - \frac{\lambda^E}{1+\lambda}\left(\frac{\hat{c}(x)}{\tilde{c}^{\lambda}(x)}\right)^{-\sigma}\hat{c}_b(x) - \sigma\left(1 - \left(\frac{1+\lambda^E}{1+\lambda}\right)^{\frac{1}{\sigma}}\right)\tilde{c}_b^{\lambda}(x)\right]$$

*Proof.* Derived in Appendix A.3 □

Recall that a rational agent's expected growth rate of consumption will equal

$$E\left[\frac{\dot{c}(x)}{c(x)}\right] = \frac{1}{\sigma}\left[r - \rho\right]$$

Expected consumption growth under temptation preferences is distorted from this rational benchmark by two forces, captured in the two extra terms above. First, expected future temptation and second, naivete. Both are distortions to the discount factor. The first scales with the *tempting* MPC, adjusted by the ratio of marginal utility from the actual to tempting consumption flow. The second scales with the actual MPC. Both forces lead to over–consumption in the present (lower expected consumption growth), and distort behaviour more for people on steeper parts of their consumption function.

This Euler embeds two important special cases. At the two ends of the naivete spectrum, one of these forces is switched off. With full sophistication $\lambda^E = \lambda$, and so only the first force is present. In this case, and assuming DSC preferences hold, the Euler has exactly the same properties as the discrete time sophisticated equivalent that has been used elsewhere (Attanasio et al., 2024; Kovacs et al., 2021). With full naivete $\lambda^E = 0$ the first force drops out and only the second remains. In this case, the Euler is identical to the naive IG agent's (with $\beta = \frac{1}{1+\lambda}$) derived in Maxted (2024).

In one respect this equivalence result reduces our knowledge. It takes a previous identification strategy off the table. That is the Euler equation method in Kovacs et al. (2021) and Huang et al. (2015). These papers estimate the degree of temptation in a life–cycle model, assuming the Euler has the sophisticated form in discrete time above. By showing expected consumption growth is increasing in liquid resources, they estimate the degree of temptation to be around $\lambda = 0.2$. But this result hinges on the assumption of sophistication. If people are at all naive, then there is an omitted variable, the MPC, that is correlated with liquid resources. As a result, the estimates from these papers will be biased if people are at all naive.[17]

[17] The papers also structurally estimate the model parameter and find similar estimate as in their Euler equation empirical estimate. The results in this paper do not invalidate that approach.

If we had an ideal dataset, we could use the Euler equation approach to identify between the two models. We would require individual measures of (a) consumption growth, (b) tempting resources (well proxied by liquid resources, wealth and income), and (c) MPC. If we found that consumption growth was increasing in tempting resources, conditional on MPC, this would be evidence in favour of the temptation model against the null of IG. The data challenge for such an approach is finding a measure of the MPC across individuals.

**The effect of sophistication on consumption** An alternative identification strategy could exploit theoretical differences in the marginal impact of sophistication on consumption. Recalling that the consumption functions in the two types have the following relationship to the rational benchmark, there is a predicted sign difference in the marginal impact of sophistication itself.

$$
\begin{aligned}
u'(\tilde{c}^{\beta}(x)) &= \beta \frac{\partial v^{\beta^E}(x)}{\partial v(x)} u'(c(x)) \\
u'(\tilde{c}^{\lambda}(x)) &= \frac{1}{1+\lambda} \frac{\partial v^{\lambda^E}(x)}{\partial v(x)} u'(c(x))
\end{aligned}
$$

For IG consumers, greater sophistication (holding actual bias fixed) reduces current consumption at all points in the state–space because $\frac{\partial v^{\beta^E}(x)}{\partial v(x)}$ is greater than one, and increasing as $\beta^E$ declines toward $\beta$.[18] By contrast, for tempted consumers, greater sophistication leads to an increase in consumption, as they over–consume in the present as a commitment device.

This raises the possibility that we could discern between the two preferences in a cross–sectional design if we had a measure of sophistication, separate from actual bias. What would this look like? Sophistication means the ability to appreciate one's future self without rose–tinted glasses. Measuring this could mean collecting (a) proxies that correlate with this faculty e.g. cognitive capacity, as in Zhang and Greiner (2021), or the various correlates in Cobb-Clark et al. (2024), (b) self–reported measures of sophistication from surveys[19] (as in Cobb-Clark et al., 2024), or (c) data capturing consumers' forecasting errors about their own behaviour (similar to the experiments in Augenblick and Rabin (2019) or Fedyk (2024)).

An ideal dataset would therefore contain measures of sophistication, consumption, and individual states like wealth and income. The directional predictions for the impact of sophistication on conditional consumption could be used to build a test distinguishing

[18] For reasonable calibrations of the elasticity of inter–temporal substitution $\sigma > 1$.

[19] Measures of self–reported self–control issues, such as the one employed in Chapter **??** likely confound both bias and sophistication: a low self–control score could mean either a sophisticated person with ample self–control, or a tempted naif.

the two approaches to modelling present–bias.

**Commitment seeking** Ever since models of present–bias have been written people have been exploring whether one of the predictions—commitment seeking by sophisticated agents—bears out. A high–level summary of this literature is that while people do like the idea of commitment devices, they seek them out less than we'd expect, and this behaviour is context dependent e.g. people adopt them in lab environments a lot, but not so much in the real world (Bernheim and Taubinsky, 2018; Laibson, 2015; Maxted, 2024).

In the context of our model, commitment–seeking appears in the voluntary contribution policy—deposits into the illiquid account tie both the agents' hands, and also their demons'. If we uphold the Maxted (2024) assumptions, then *any* commitment–seeking behaviour is evidence in favour of the temptation model rather than IG. Recall that in IG under these assumptions, actual behaviour can never be constrained and so commitment is pointless. By contrast, tempted consumers will seek commitment to rob their demons of ammunition even if they continue to over–consume. Under these assumptions, this is the simplest test of the models' predictions.

But the Maxted (2024) assumptions are strong. There may be no actual limit to the resources one *could* get, if willing to take on increasingly onerous financial and social costs. But some of these costs may be more fixed than marginal, and that would be enough to replicate a hard borrowing constraint by inducing procrastination (Maxted et al., 2024). In such a world, both models predict commitment seeking behaviour, but there would still be a difference. As identified in Toussaert (2018), *tempted consumers seek to tie their demon's hands*, not necessarily their own. As a result, they may adopt commitment devices that aren't particularly constraining except by reducing temptation. Finding examples of these outside of the lab would be an interesting opportunity to test between the two theories, and bolster the evidence found there.

# 7 Conclusion

There is substantial evidence that consumers are subject to some degree of present–bias. Building this into macro models is helpful for matching aggregate moments that matter for policy–making. I have covered the two main approaches to modelling present–bias, quasi-hyperbolic discounting (QHD) and temptation preferences, and shown how to implement both in continuous time, allowing for the first direct comparison between the two.

I show that temptation preferences are the more flexible tool. They nest IG preferences as a special case (either under naivete or when the temptation is to adopt IG) and they also require fewer structural assumptions. Unlike IG, temptation preferences can be used with any utility specification, making them practical in richer macro settings with e.g. labour supply.

While the two models are behaviourally equivalent for naive agents, they have very different welfare implications for the agents, and the behavioural equivalence breaks if agents are sophisticated. That first point should give us pause before using one or the other of these preferences in policy settings. The second presents opportunities for identification between the two.

# A Appendix

## A.1 Rational benchmark HJB equation

The following details the steps to derive a standard HJB equation for the rational benchmark. The consumer's value general is

$$v(x_t) = \max_{\mathbf{c},\mathbf{d}} \mathbb{E} \int_0^\infty e^{-\rho s} u(c_s(x_{t+s}))ds$$

Where the state transitions are governed by the policy functions and risk processes. Suppose we're dealing with the stationary solution, and so the stream of consumption and deposit plans resolve to two policy functions $\mathbf{c} = c(x)$ and $\mathbf{d} = d(x)$. To make the HJB out of this, first separate out the present $\Delta$ from the future

$$v(x_t) = \lim_{\Delta \to 0} \max_{c,d} u(c)\Delta + e^{-\rho\Delta} v(x_{t+\Delta}(c,d))$$

Approximate the continuation value with a second order Taylor expansion around $\Delta = 0$

$$v(x_{t+\Delta}) \approx v(x_t) + \left[\partial_b v(x_t) \cdot \dot{b}_t + \partial_a v(x_t) \cdot \dot{a}_t + \mathcal{A}[v](x)\right] \Delta$$

Where $\mathcal{A}[v](x) = -\theta_z \ln z_t \partial_z v(x) + \frac{1}{2}\sigma_z^2 \partial_{zz} v(x)$ is the infinitesimal generator of the idiosyncratic risk process. Substitute this back in and also use the approximation $e^{-\rho\Delta} \approx 1 - \rho\Delta$ (that gets better in the limit as $\Delta \to 0$

$$v(x_t) = \lim_{\Delta \to 0} \max_{c,d} u(c)\Delta + (1-\rho\Delta)\left(v(x_t) + \left[\partial_b v(x_t) \cdot \dot{b}_t + \partial_a v(x_t) \cdot \dot{a}_t + \mathcal{A}[v](x_t)\right] \Delta\right)$$

Finally, rearranging and taking the limit yields the HJB equation

$$\rho v(x) = \max_{c,d} u(c) + \partial_b v(x) \cdot \dot{b} + \partial_a v(x) \cdot \dot{a} + \mathcal{A}[v](x)$$

## A.2 Tempted HJB equation

An HJB equation can be derived for each case of temptation preferences introduced in the main text. In each case we must do a little work to show how the tempting discount function affects the value function, and how this can be re–arranged into a form appropriate for deriving an HJB equation.

**Tempted by greater discounting** Consumers are tempted to adopt the alternative discount function $D(s) = e^{-\hat{\rho}s}$. Their tempting consumption plan $\hat{\mathbf{c}}$ will be the solution to the model with that discount rate, where bold case captures that this is a stream of consumption plans into the future. Their long–run preferences for a consumption and deposit plan $(\mathbf{c},\mathbf{d})$ are

$$h(\mathbf{c},\mathbf{d},x)=\mathbb{E}\int_0^\infty e^{-\rho s}u(c(x_{t+s},s))ds$$

And the tempting utility from the same plan is

$$k(\mathbf{c},\mathbf{d},x)=\mathbb{E}\int_0^\infty \lambda e^{-\hat{\rho}s}u(c(x_{t+s},s))ds$$

Based on these, the sophisticated tempted value is

$$v^\lambda(x)=\max_{\mathbf{c},\mathbf{d}} h(\mathbf{c},\mathbf{d},x)-\left[k(\hat{\mathbf{c}},\hat{\mathbf{d}},x)-k(\mathbf{c},\mathbf{d},x)\right]$$

Substituting the definitions, and separating present from future

$$\begin{aligned}
v^\lambda(x)&=\max_{\mathbf{c},\mathbf{d}}\mathbb{E}\int_0^\infty e^{-\rho s}u(c(x_{t+s},s))ds-\mathbb{E}\int_0^\infty \lambda e^{-\hat{\rho}s}\left[u(\hat{c}(x_{t+s},s))-u(c(x_{t+s},s))\right]ds\\
v^\lambda(x)&=\lim_{\Delta\to0}\max_{\mathbf{c},\mathbf{d}}\left[(1+\lambda)u(c(x,0))-\lambda u(\hat{c}(x,0))\right]\Delta\\
&\quad+e^{-\rho\Delta}\,\mathbb{E}\int_\Delta^\infty e^{-\rho s}u(c(x_{t+s},s))ds-e^{-\hat{\rho}\Delta}\,\mathbb{E}\int_\Delta^\infty \lambda e^{-\hat{\rho}s}\left[u(\hat{c}(x_{t+s},s))-u(c(x_{t+s},s))\right]ds\\
v^\lambda(x)&=\lim_{\Delta\to0}\max_{\mathbf{c},\mathbf{d}}\left[(1+\lambda)u(c(x,0))-\lambda u(\hat{c}(x,0))\right]\Delta\\
&\quad+e^{-\rho\Delta}\left[h(\mathbf{c}_\Delta,\mathbf{d}_\Delta,x_{t+\Delta})-\left[k(\hat{\mathbf{c}}_\Delta,\hat{\mathbf{d}}_\Delta,x_{t+\Delta})-k(\mathbf{c}_\Delta,\mathbf{d}_\Delta,x_{t+\Delta})\right]\right]\\
&\quad+\left[e^{-\rho\Delta}-e^{-\hat{\rho}\Delta}\right]\left[k(\hat{\mathbf{c}}_\Delta,\hat{\mathbf{d}}_\Delta,x_{t+\Delta})-k(\mathbf{c}_\Delta,\mathbf{d}_\Delta,x_{t+\Delta})\right]
\end{aligned}$$

And rearranging

$$\begin{aligned}
v^\lambda(x)&=\lim_{\Delta\to0}\max_{c,d}\left[(1+\lambda)u(c)-\lambda u(\hat{c}(x))\right]\Delta+e^{-\rho\Delta}v^\lambda(x_{t+\Delta})\\
&\quad+(e^{-\rho\Delta}-e^{-\hat{\rho}\Delta})\left(k(\hat{\mathbf{c}},\hat{\mathbf{d}},x_{t+\Delta})-k(\mathbf{c},\mathbf{d},x_{t+\Delta})\right)
\end{aligned}$$

This has the form of a standard Bellman with a distorted utility function, where the final term makes an adjustment for the discounting of the temptation cost going forward, which is not felt as strongly the continuation value suggests. To find the HJB we take a second order approximation around $\Delta=0$ and rearrange, as in Appendix A.1.

$$\begin{aligned}
\rho v^\lambda(x)&=\max_{c,d}\,(1+\lambda)u(c)-\lambda u(\hat{c})+(\hat{\rho}-\rho)\left(k(\hat{\mathbf{c}},\hat{\mathbf{d}},x_{t+\Delta})-k(\mathbf{c},\mathbf{d},x_{t+\Delta})\right)\\
&\quad+\partial_b v^\lambda(x)\dot{b}+\partial_a v^\lambda(x)\dot{a}+\mathcal{A}[v^\lambda](x)
\end{aligned}$$

**Tempted by quasi–hyperbolic discounting** Recall the IG discount function

$$D(s) = \begin{cases} 1 & \text{if } s = 0 \\ \beta e^{-\rho s} & \text{else} \end{cases}$$

And note the tempting consumption function $\hat{c}$ for a given degree of present bias will be a scale multiple of the rational equivalent. Their long–run preferences for a consumption and deposit plan $(\mathbf{c}, \mathbf{d})$ are

$$h(\mathbf{c}, \mathbf{d}, x)) = \mathbb{E} \int_0^\infty e^{-\rho s} u(c(x_{t+s}, s)) ds$$

and note the tempting version of these satisfies

$$k(\mathbf{c}, \mathbf{d}, x) = \mathbb{E} \int_0^\infty \lambda D(s) u(c(x_{t+s}, s)) ds = \lambda \beta h(\mathbf{c}, \mathbf{d}, x))$$

Now define the sophisticated tempted value as follows

$$v^\lambda(x_t) = \max_{\mathbf{c},\mathbf{d}} \mathbb{E} \int_0^\infty e^{-\rho s} u(c(x_{t+s}, s)) ds - \mathbb{E} \int_0^\infty \lambda D(s) \left[ u(\hat{c}(x_{t+s}, s)) - u(c(x_{t+s}, s)) \right] ds$$

Separate the present from the future

$$\begin{aligned} v^\lambda(x_t) = &\lim_{\Delta \to 0} \max_{\mathbf{c},\mathbf{d}} \left( u(c(x_t, 0)) - \lambda \left[ u(\hat{c}(x_t, 0)) - u(c(x_t, 0)) \right] \right) \Delta \\ &+ e^{-\rho \Delta} \left( \mathbb{E} \int_\Delta^\infty e^{-\rho s} u(c(x_{t+s}, s)) ds - \lambda \beta \, \mathbb{E} \int_\Delta^\infty e^{-\rho s} \left[ u(\hat{c}(x_{t+s}, s)) - u(c(x_{t+s}, s)) \right] ds \right) \end{aligned}$$

Use the definitions of $h(\cdot)$ for the future parts, and let $\mathbf{c}_\Delta$ denote the consumption plan starting from $t + \Delta$

$$\begin{aligned} v^\lambda(x_t) = &\lim_{\Delta \to 0} \max_{\mathbf{c},\mathbf{d}} \left( u(c) - \lambda \left[ u(\hat{c}(x_t, 0)) - u(c) \right] \right) \Delta \\ &+ e^{-\rho \Delta} \, \mathbb{E} \left( h(\mathbf{c}_\Delta, \mathbf{d}_\Delta, x_{t+\Delta}) - \lambda \beta \left[ h(\hat{\mathbf{c}}_\Delta, \hat{\mathbf{d}}_\Delta, x_{t+\Delta}) - h(\mathbf{c}_\Delta, \mathbf{d}_\Delta, x_{t+\Delta}) \right] \right) \end{aligned}$$

Recognise that $\lambda \beta \cdot h(\mathbf{c}, \mathbf{d}, x)) = k(\mathbf{c}, \mathbf{d}, x))$ and substitute

$$\begin{aligned} v^\lambda(x_t) = &\lim_{\Delta \to 0} \max_{\mathbf{c},\mathbf{d}} \left( u(c(x_t, 0)) - \lambda \left[ u(\hat{c}(x_t, 0)) - u(c(x_t, 0)) \right] \right) \Delta \\ &+ e^{-\rho \Delta} \, \mathbb{E} \left( h(\mathbf{c}_\Delta, \mathbf{d}_\Delta, x_{t+\Delta}) - \left[ k(\hat{\mathbf{c}}_\Delta, \hat{\mathbf{d}}_\Delta, x_{t+\Delta}) - k(\mathbf{c}_\Delta, \mathbf{d}_\Delta, x_{t+\Delta}) \right] \right) \end{aligned}$$

And simplify to arrive at the continuation value function

$$v^\lambda(x_t) = \lim_{\Delta\to 0} \max_{c,d} \left(u(c) - \lambda \left[u(\hat{c}(x_t, 0)) - u(c)\right]\right) \Delta + e^{-\rho\Delta}\, \mathbb{E}\, v^\lambda(x_{t+\Delta})$$

This has a standard form, and so the HJB does as well. Taking a second–order approximation around $\Delta = 0$, as in Appendix A.1, yields

$$\rho v^\lambda(x) = \max_{c,d}\ (1+\lambda)u(c) - \lambda u(\hat{c}) + \partial_b v^\lambda(x)\dot{b} + \partial_a v^\lambda(x)\dot{a} + \mathcal{A}[v^\lambda](x)$$

**Tempted by ignoring the future** This case uses the piecewise tempting discount function detailed in Section 5.1. The steps to derive the HJB are identical to those with quasi–hyperbolic discounting, and the result is the same except that the tempting consumption plan $\hat{c}$ solves a different problem—cake–eating over a limited span, compared with the IG consumption policy.

## A.3 Tempted Euler equation

Here I derive an Euler equation for a tempted consumer with tempting alternative consumption $\hat{c}$ and an arbitrary level of sophistication where $\lambda^E \in [0, \lambda]$. To start, note that these consumers define their choices based on the value function they expect to prevail in future. This is defined by the HJB below

$$\begin{aligned}\rho v^{\lambda^E}(x) &= (1+\lambda^E)u(c^E(x)) - \lambda^E u(\hat{c}^E(x)) \\ &\quad + \partial_b v^{\lambda^E}(x)\dot{b}^E(x) + \partial_a v^{\lambda^E}(x)\dot{a}^E(x) + \mathcal{A}\left[v^{\lambda^E}\right](x)\end{aligned}$$

To derive the Euler we first take the derivative with respect to the liquid asset

$$\begin{aligned}\rho\partial_b v^{\lambda^E}(x) &= (1+\lambda^E)u'(c^E(x))\partial_b c^E(x) - \lambda^E u'(\hat{c}^E(x))\partial_b \hat{c}^E(x) \\ &\quad + \partial_b v^{\lambda^E}(x)(r(b) + r_b(b)b - \partial_b c^E(x)) \\ &\quad + \partial_{bb} v^{\lambda^E}(x)\dot{b}^E(x) + \partial_{ab} v^{\lambda^E}(x)\dot{a}^E(x) + \mathcal{A}\left[\partial_b v^{\lambda^E}\right](x)\end{aligned}$$

Substituting the *realised* FOC $\partial_b v^{\lambda^E}(x) = (1+\lambda)u'(c(x))$ to eliminate $\partial_b v^{\lambda^E}(x)$, and collecting terms

$$\begin{aligned}\left(\rho - r(b) - r_b(b)b\right) u'(c(x)) &= \frac{1+\lambda^E}{1+\lambda} u'(c^E(x))\partial_b c^E(x) - \frac{\lambda^E}{1+\lambda} u'(\hat{c}^E(x))\partial_b \hat{c}^E(x) \\ &\quad - u'(c(x))\partial_b c^E(x) \\ &\quad + \partial_b u'(c(x))\dot{b}^E(x) + \partial_a u'(c(x))\dot{a}^E(x) + \mathcal{A}\left[u'(c(x))\right](x)\end{aligned}$$

Now recognise that the consumer expects their future selves to choose consumption to meet the FOC $(1+\lambda^E)u'(c^E(x)) = \partial_b v^{\lambda^E}(x)$. Combining this expectation with the present consumer's FOC we know $(1+\lambda^E)u'(c^E(x)) = (1+\lambda)u'(c(x))$. Substituting and collecting terms

$$
\begin{aligned}
(\rho - r(b) - r_b(b)b)\, u'(c(x)) = &-\frac{\lambda^E}{1+\lambda} u'(\hat{c}^E(x))\partial_b \hat{c}^E(x) \\
&+ \partial_b u'(c(x))\dot{b}^E(x) + \partial_a u'(c(x))\dot{a}^E(x) + \mathcal{A}\left[u'(c(x))\right](x)
\end{aligned}
$$

The next step is to track the expected path of marginal utility for the consumer in the present. The drift terms at the moment are framed in terms of their expected future policies, rather than those employed right now. We need to add and subtract $\partial_b u'(c(x))\dot{b}(x) + \partial_a u'(c(x))\dot{a}(x)$ and collect terms, substituting the definition for the time derivative of marginal utility $\mathbb{E}\left[du'(c(x))/dt\right] = \partial_b u'(c(x))\dot{b}(x) + \partial_a u'(c(x))\dot{a}(x) + \mathcal{A}\left[u'(c(x))\right](x)$. This yields

$$
\begin{aligned}
(\rho - r(b) - r_b(b)b)\, u'(c(x)) = &-\frac{\lambda^E}{1+\lambda} u'(\hat{c}^E(x))\partial_b \hat{c}^E(x) + \mathbb{E}\left[du'(c(x))/dt\right] \\
&+ \partial_b u'(c(x))\left[\dot{b}^E(x) - \dot{b}(x)\right] + \partial_a u'(c(x))\left[\dot{a}^E(x) - \dot{a}(x)\right]
\end{aligned}
$$

And now substituting for the state transition equations, and recognising that $d^E(x) = d(x)$

$$
\begin{aligned}
(\rho - r(b) - r_b(b)b)\, u'(c(x)) = &-\frac{\lambda^E}{1+\lambda} u'(\hat{c}^E(x))\partial_b \hat{c}^E(x) + \mathbb{E}\left[du'(c(x))/dt\right] \\
&+ \partial_b u'(c(x))\left[c^E(x) - c(x)\right]
\end{aligned}
$$

And then rearranging we have our general Euler equation

$$
\mathbb{E}\left[\frac{du'(c(x))/dt}{u'(c(x))}\right] = \rho - r(b) - r_b(b)b + \frac{\lambda^E}{1+\lambda}\frac{u'(\hat{c}^E(x))}{u'(c(x))}\partial_b \hat{c}^E(x) - \frac{\partial_b u'(c(x))}{u'(c(x))}\left[c^E(x) - c(x)\right]
$$

Here we can see the general Euler equation includes two separate distortions to discounting stemming from temptation and sophistication about that temptation. In the limit where consumers are fully sophisticated, the final term drops out and we're left with just the one distortion. Similarly, when consumers are fully naive the second to last term drops out, and we're left with the other.

The Euler simplifies further when we assume CRRA utility such that $u'(c) = c^{-\sigma}$ and we know from combining FOC that $c^E(x) = \left(\frac{1+\lambda}{1+\lambda^E}\right)^{-\frac{1}{\sigma}} c(x)$ and also substituting for the coefficient of relative risk aversion $-\sigma = \frac{u''(c(x))c(x)}{u'(c(x))}$

$$\mathbb{E}\left[\frac{\dot{c}(x)}{c(x)}\right] = -\frac{1}{\sigma}\left(\rho - r(b) - r_b(b)b + \frac{\lambda^E}{1+\lambda}\left(\frac{\hat{c}^E(x)}{c(x)}\right)^{-\sigma}\partial_b\hat{c}^E(x) + \sigma\left[1-\left(\frac{1+\lambda^E}{1+\lambda}\right)^{\frac{1}{\sigma}}\right]\partial_b c(x)\right)$$

This Euler simplifies to two special cases. With complete sophistication, we recover the continuous–time equivalent of the tempting Euler equation derived in Attanasio et al. (2024). With complete naivete, we recover the equivalent of the naive IG Euler equation in Maxted (2024), noting hat $1/(1+\lambda)$ takes the place of $\beta$.

# References


Achdou, Y., J. Han, J.-M. Lasry, P.-L. Lions, and B. Moll (2022): "Income and wealth distribution in macroeconomics: A continuous-time approach," *The review of economic studies*, 89, 45–86.

Aiyagari, S. R. (1994): "Uninsured idiosyncratic risk and aggregate saving," *The Quarterly Journal of Economics*, 109, 659–684.

Attanasio, O., A. Kovacs, and P. Moran (2024): "Temptation and Commitment: a Model of Hand-to-Mouth Behavior," *Journal of the European Economic Association*, 2025–2073.

Auclert, A., M. Rognlie, and L. Straub (2024a): "Fiscal and Monetary Policy with Heterogeneous Agents," *Unpublished Working Paper*.

——— (2024b): "The Intertemporal Keynesian Cross," *Journal of Political Economy*, 132, 4068–4121.

Augenblick, N. and M. Rabin (2019): "An experiment on time preference and misprediction in unpleasant tasks," *Review of Economic Studies*, 86, 941–975.

Baker, S. R., R. A. Farrokhnia, S. Meyer, M. Pagel, and C. Yannelis (2023): "Income, liquidity, and the consumption response to the 2020 economic stimulus payments," *Review of Finance*, 27, 2271–2304.

Baugh, B., I. Ben-David, H. Park, and J. A. Parker (2021): "Asymmetric Consumption Smoothing," *American Economic Review*, 111, 192–230.

Bayer, C., R. Luetticke, L. Pham-Dao, and V. Tjaden (2019): "Precautionary Savings, Illiquid Assets, and the Aggregate Consequences of Shocks to Household Income Risk," *Econometrica*, 87, 255–290.

Bernheim, B. D. and D. Taubinsky (2018): "Chapter 5 - Behavioral Public Economics," in *Handbook of Behavioral Economics - Foundations and Applications 1*, ed. by B. D. Bernheim, S. DellaVigna, and D. Laibson, North-Holland, vol. 1 of *Handbook of Behavioral Economics: Applications and Foundations 1*, 381–516.

Boehm, J., E. Fize, and X. Jaravel (2025): "Five facts about MPCs: Evidence from a randomized experiment," *American Economic Review*, 115, 1–42.

Campbell, J. Y. and N. G. Mankiw (1989): "Consumption, income, and interest rates: Reinterpreting the time series evidence," *NBER Macroeconomics Annual*, 4, 185–216.

Carroll, C. D., M. B. Holm, and M. S. Kimball (2021): "Liquidity constraints and precautionary saving," *Journal of Economic Theory*, 195, 105276.

Carroll, C. D. and M. S. Kimball (1996): "On the Concavity of the Consumption Function," *Econometrica*, 64, 981–992.

Cobb-Clark, D. A., S. C. Dahmann, D. A. Kamhöfer, and H. Schildberg-Hörisch (2024): "Sophistication about self-control," *Journal of Public Economics*, 238, 105196.

Deaton, A. (1989): "Saving and liquidity constraints," *NBER Working Paper*.

Fagereng, A., M. B. Holm, B. Moll, and G. Natvik (2019): "Saving behavior across the wealth distribution: The importance of capital gains," .

Fedyk, A. (2024): "Asymmetric naivete: Beliefs about self-control," *Management Science*.

Flavin, M. A. (1981): "The adjustment of consumption to changing expectations about future income," *Journal of political economy*, 89, 974–1009.

Friedman, M. (1957): "The Permanent Income Hypothesis," in *A Theory of the Consumption Function*, Princeton University Press, 20–37.

Fudenberg, D. and D. K. Levine (2006): "A dual-self model of impulse control," *American Economic Review*, 96, 1449–1476.

Ganong, P. and P. Noel (2020): "Liquidity versus wealth in household debt obligations: Evidence from housing policy in the great recession," *American Economic Review*, 110, 3100–3138.

Gelman, M. (2022): "The Self-Constrained Hand-to-Mouth," *The Review of Economics and Statistics*, 104, 1096–1109.

Gelman, M., S. Kariv, M. D. Shapiro, and D. Silverman (2022): "Rational Illiquidity and Consumption: Theory and Evidence from Income Tax Withholding and Refunds," *American Economic Review*, 112, 2959–91.

Gerard, F. and J. Naritomi (2021): "Job displacement insurance and (the lack of) consumption-smoothing," *American Economic Review*, 111, 899–942.

Gul, F. and W. Pesendorfer (2001): "Temptation and self-control," *Econometrica*, 69, 1403–1435.

——— (2004): "Self-control, revealed preference and consumption choice," *Review of Economic Dynamics*, 7, 243–264.

Hall, R. E. (1978): "Stochastic implications of the life cycle-permanent income hypothesis: theory and evidence," *Journal of political economy*, 86, 971–987.

Hamilton, S., G. Liu, J. Miranda-Pinto, and T. Sainsbury (2024): "A $100,000 marshmallow experiment: Withdrawal and spending responses to early retirement-savings access," *Unpublished Working Paper.*

Harris, C. and D. Laibson (2013): "Instantaneous Gratification," *The Quarterly Journal of Economics*, 128, 205–248.

Huang, K. X., Z. Liu, and J. Q. Zhu (2015): "Temptation and Self-Control: Some Evidence and Applications," *Journal of Money, Credit and Banking*, 47, 581–615.

Jappelli, T. and L. Pistaferri (2010): "The consumption response to income changes," *Annu. Rev. Econ.*, 2, 479–506.

——— (2014): "Fiscal policy and MPC heterogeneity," *American Economic Journal: Macroeconomics*, 6, 107–136.

Kaplan, G., B. Moll, and G. L. Violante (2018): "Monetary policy according to HANK," *American Economic Review*, 108, 697–743.

Kaplan, G. and G. L. Violante (2022): "The marginal propensity to consume in heterogeneous agent models," *Annual Review of Economics*, 14, 747–775.

Kaplan, G., G. L. Violante, and J. Weidner (2014): "The wealthy hand-to-mouth," .

Karger, E. and A. Rajan (2020): "Heterogeneity in the marginal propensity to consume: evidence from Covid-19 stimulus payments," *FRB of Chicago Working Paper No. WP 2020-15.*

Kovacs, A., H. Low, and P. Moran (2021): "Estimating Temptation and Commitment Over the Life Cycle," *International Economic Review*, 62, 101–139.

Krusell, P., B. Kuruşçu, and A. A. Smith Jr (2010): "Temptation and taxation," *Econometrica*, 78, 2063–2084.

Kueng, L. (2018): "Excess Sensitivity of High-Income Consumers," *The Quarterly Journal of Economics*, 133, 1693–1751.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *The Quarterly Journal of Economics*, 112, 443–478.

——— (2015): "Why Don't Present-Biased Agents Make Commitments?" *American Economic Review*, 105, 267–72.

Laibson, D., S. C. Lee, P. Maxted, A. Repetto, and J. Tobacman (2024): "Estimating Discount Functions with Consumption Choices over the Lifecycle," *The Review of Financial Studies.*

Lee, S. C. and P. Maxted (2023): "Credit Card Borrowing in Heterogeneous-Agent Models: Reconciling Theory and Data," *Unpublished Working Paper.*

Maxted, P. (2024): "Present Bias Unconstrained: Consumption, Welfare, and the Present–Bias Dilemma," *Unpublished Working Paper.*

Maxted, P., D. Laibson, and B. Moll (2024): "Present Bias Amplifies the Household Balance-Sheet Channels of Macroeconomic Policy," *The Quarterly Journal of Economics*, 691–743.

Modigliani, F. and R. Brumberg (1954): "Utility analysis and the consumption function: An interpretation of cross-section data," *Franco Modigliani*, 1, 388–436.

Ni, S. and Y. Seol (2014): "New evidence on excess sensitivity of household consumption," *Journal of Monetary Economics*, 63, 80–94.

Olafsson, A. and M. Pagel (2018): "The Liquid Hand-to-Mouth: Evidence from Personal Finance Management Software," *The Review of Financial Studies*, 31, 4398–4446.

O'Donoghue, T. and G. Loewenstein (2004): "Animal Spirits: Affective and Deliberative Processes in Economic Behavior," *Unpublished Working Paper*.

Parker, J. A. (2017): "Why Don't Households Smooth Consumption? Evidence from a $25 Million Experiment," *American Economic Journal: Macroeconomics*, 9, 153–83.

Phelps, E. S. and R. A. Pollak (1968): "On Second-Best National Saving and Game-Equilibrium Growth1," *The Review of Economic Studies*, 35, 185–199.

Poterba, J. M. and J. J. Rotemberg (1986): "Money in the utility function: An empirical implementation," .

Sabet, S. and P. Schneider (2024): "The Nested–Drift Algorithm," *Unpublished Working Paper*.

Shapiro, J. M. (2005): "Is there a daily discount rate? Evidence from the food stamp nutrition cycle," *Journal of Public Economics*, 89, 303–325.

Strotz, R. H. (1955): "Myopia and Inconsistency in Dynamic Utility Maximization," *The Review of Economic Studies*, 23, 165–180.

Toczynski, J. (2023): "Spend or invest? Analyzing MPC heterogeneity across three stimulus programs," *Unpublished Working Paper*.

Toussaert, S. (2018): "Eliciting temptation and self-control through menu choices: A lab experiment," *Econometrica*, 86, 859–889.

Zeldes, S. P. (1989): "Consumption and liquidity constraints: an empirical investigation," *Journal of political economy*, 97, 305–346.

Zhang, Q. and B. Greiner (2021): "Time inconsistency, sophistication, and commitment: An experimental study," *Economics Letters*, 206, 109982.