

Goal-Setting and Self-Control

By

Alice Hsiaw

August 2014

COLLEGE OF THE HOLY CROSS, DEPARTMENT OF ECONOMICS
FACULTY RESEARCH SERIES, PAPER NO. 14-04*



Department of Economics
College of the Holy Cross
Box 45A
Worcester, Massachusetts 01610
(508) 793-3362 (phone)
(508) 793-3708 (fax)

<http://www.holycross.edu/departments/economics/website>

* All papers in the Holy Cross Working Paper Series should be considered draft versions subject to future revision. Comments and suggestions are welcome.

Goal-Setting and Self-Control

By
Alice Hsiaw[†]
College of the Holy Cross

Originally written: April 2012

Posted: August 2014

Abstract

This paper addresses the role of non-binding goals to attenuate time inconsistency. Present-biased agents have linear reference-dependent preferences and endogenously set a goal that is the reference point. They face an infinite horizon, optimal stopping problem in continuous time, where there exists an option value of waiting due to uncertainty. When there is sufficient commitment to expectation-based goals, goal-setting attenuates the time-inconsistent agent's tendency to stop too early, and may even lead an agent to wait longer than the first-best. In particular, reference dependence is strictly worse for a time-consistent agent. Notably, none of the effects of goal-setting require any form of loss aversion.

Acknowledgements: I would like to thank my adviser, Roland B´enabou, for guidance and encouragement. I also thank Wei Xiong, Stephen Morris, Jos´e Scheinkman, Matthew Rabin, Nageeb Ali, Botond K´oszegi, Eldar Shafir, Ing-Haw Cheng, Dustin Tingley, Sambuddha Ghosh, Alisdair McKay, Konstantin Milbradt, Stephanie Wang, and seminar participants at Princeton University, Kellogg School of Management (MEDS), University of Southern California, University of California - Santa Barbara, Texas A&M University, Cornell University, and College of the Holy Cross, for helpful comments and suggestions. All remaining errors are my own.

JEL Classification Codes:

Keywords:

[†]Alice Hsiaw, Department of Economics and Accounting, College of the Holy Cross, 1 College Street, Worcester, MA 01610; Phone: (508) 793-2687; Email: ahsiaw@holycross.edu

1 Introduction

Personal development is a burgeoning, multi-billion dollar industry that focuses on self-improvement on a variety of levels, from career aspirations to lifestyle choices to spiritual well-being. In response to consumer demand, this market offers books, motivational speakers, and personal coaching, as well as innumerable weight-loss programs. Institutional demand has resulted in consulting and employee development programs whose purpose is to raise productivity in the workplace. A central tenet of this industry is that goal-setting is a vital instrument for improving any aspect of life. For example, the “**S**(pecific) **M**(easurable) **A**(ttainable) **R**(elevant) **T**(imely) Goals” mnemonic is ubiquitous in project management, education, and self-help programs.¹ This prescription is supported by extensive empirical evidence that *non-binding* goals tend to increase effort, attention, and persistence (Klein, 1991; Latham and Locke, 1991; Locke and Latham, 2002), and that satisfaction is tied to achievement relative to such a goal, not just final outcomes (Mento et al., 1992).

Although much attention has been paid to the problem of present-biased preferences and intrapersonal conflict by individual decision makers, the role of self-set goals in attenuating the self-control problem has been relatively unexplored by economists thus far. In standard economic models, only binding goals, involving external enforcement of punishment and rewards or pre-commitment, can affect motivation and behavior. For example, an employer can motivate a worker’s effort choice by prescribing goals if the worker is evaluated and compensated accordingly. Present-biased agents can also enforce personal motivation by using binding precommitments or externally enforced contracts.

However, the prevalence of non-binding goal-setting, and the success of personal development, strongly suggests that less drastic measures may be successful regulatory mechanisms. For example, dieters often set weight targets, and writers set self-imposed page targets and deadlines. Individuals may also set goals for themselves in situations where there exists an option value of waiting due to uncertainty. A student may thus continue his education to achieve a target level of starting salary upon graduation. A person trying to save may set a target level of accumulated wealth upon retirement.

In this paper, I propose that a goal provides internal motivation by acting as the reference point for a present-biased agent who possesses reference-dependent preferences. I consider agents who derive utility from both standard consumption, arising from material outcomes, and comparison utility, a psychological payoff derived from the comparison of these outcomes to a self-set goal. At each point in time, the agent sets a goal determined by his expectations about the outcome of future behavior, which his next “self” will inherit as a reference point

¹There exist many variations on this theme, which is generally attributed to Drucker (1954).

in his utility function. In this way, the agent provides a degree of internal motivation that is desirable if he exhibits time inconsistency due to present-biased time preferences.

The economic setting that I consider is an infinite-horizon, optimal stopping problem in continuous time, where there exists an option value of waiting due to uncertainty. Present-biasedness leads the agent to stop too early because he undervalues this option relative to his time-consistent counterpart. In the preceding examples, this means accumulating too little human capital or retiring with too little wealth. I show that if the agent has reference dependent preferences, even of the simplest, *linear* kind, he can potentially induce more patient behavior by setting a goal to be evaluated upon stopping. Formally, I solve for the unique stationary Markov equilibrium of the intrapersonal game. I first consider the case of full commitment, in which the agent's goal is fully constrained to satisfy rational expectations, and show that greater reference dependence leads to later exercise of the option. The presence of a goal increases the agent's incentive to wait because he wishes to avoid incurring comparative disutility from falling short of it. For any degree of present-biasedness, there exists a level of reference dependence such that the agent can achieve the first-best from an ex-ante perspective, which coincides with the time-consistent solution. On the other hand, too much reference dependence causes him to wait *too long*. Reference dependence always decreases the welfare of a time-consistent agent by causing him to wait longer than the first-best, by distorting his incentives at the margin. Thus, goal setting *per se* can itself be a source of intrapersonal conflict. Another key result, which stands in contrast to previous literature, is that goal-setting does not require any form of loss aversion or curvature in the comparison utility function to regulate behavior.

Next, I extend the model to partial commitment, where the agent's goal is not fully constrained to satisfy rational expectations. Here, the goal that an agent evaluates does not necessarily coincide with what he expects to achieve, but can instead be the goal that maximizes his present-biased preferences. This leads to a trade-off between the agent's material payoff and his psychological payoff. When the expectation-based goal is sufficiently likely to be evaluated, an appropriate degree of reference dependence again allows the agent to achieve the ex-ante optimum by counteracting the tendency to stop too early due to present-biasedness. Surprisingly, if commitment to the expectation-based goal is sufficiently weak, the threshold for which *is itself dependent on the degree of present-biasedness*, both present-biasedness and any degree of reference-dependence are actually welfare-improving.

The paper proceeds as follows. Section 2 links this paper to related lines of research. Section 3 describes the model. Section 4 derives and characterizes the stationary Markov equilibrium and discusses the welfare implications of goal-setting. Section 5 considers the case of partial commitment. Section 6 concludes. Proofs are gathered in the Appendix.

2 Related Literature

This paper lies at the intersection of several lines of research. It links the work on reference dependence with that on self-control, which have each been studied quite separately, by considering the role of reference dependent preferences as an instrument to countervail a self-control problem. It also analyzes behavior and welfare in the context of optimal stopping under uncertainty.

Goal-setting theory posits that goals serve as a reference standard in a cognitive comparison process of self-evaluation and satisfaction (Latham and Locke, 1991; Locke and Latham, 2002; Bandura, 1989). Heath et al. (1999) explicitly argue that a goal acts as the reference point in the prospect theory (Kahneman and Tversky, 1979) value function, while Loewenstein (2007) discusses goal-setting in the context of mountaineering, but neither formally endogenizes goals. In parallel and independent work, Suvorov and van de Ven (2008) and Koch and Nafziger (2011, 2009) also propose models of goals as reference points to remedy self-control. While sharing a similar concept, the papers are quite complementary. Both Suvorov and van de Ven (2008) and Koch and Nafziger (2011, 2009) consider a three-period problem where costly effort on a task results in a delayed benefit and a sophisticated, quasi-hyperbolic agent sets a goal regarding both effort and the task benefit. There, loss aversion is required to affect behavior and outcome uncertainty plays a key role. In contrast to related work on reference-dependent preferences, including Koch and Nafziger (2011, 2009), Suvorov and van de Ven (2008), and Kőszegi and Rabin (2006, 2009), none of my results require either loss aversion or uncertainty over final payoffs.

Kőszegi and Rabin (2006) endogenize reference point formation through rational expectations. This framework has been extended to a dynamic setting where agents derive utility from the revision of beliefs (Kőszegi and Rabin, 2009; Matthey and Dwenger, 2007; Matthey, 2008). A recent body of experimental (Abeler et al., 2011; Ericson and Fuster, 2010; Sprenger, 2010) and empirical (Pope and Schweitzer, 2011; Card and Dahl, 2011) studies find support for rational expectations as endogenous reference points. There is also empirical evidence, though debated, that people use reference points in the form of income targets (Camerer et al., 1997; Farber, 2005, 2008; Crawford and Meng, 2008; Fehr and Goette, 2007; Goette and Huffman, 2005).

The concept of intrapersonal conflict due to intertemporal differences in preferences within the self was first studied by Strotz (1956) and Schelling (1984), and more broadly developed by Ainslie (1992) and Laibson (1997). Self-control can be improved through self-imposed, binding commitments (Brocas and Carrillo, 2005; Carrillo, 2005; Gul and Pendorfer, 2001; Bisin and Hyndman, 2009; Ariely and Wertenbroch, 2002; Fudenberg and

Levine, 2006). Bénabou and Tirole (2004) develop a theory of internal regulation through self-enforcing personal rules based on a mechanism of self-reputation. The intrapersonal problem with time inconsistent preferences shares clear parallels with a principal-agent problem with moral hazard, where optimal compensation schemes have been studied (Ou-Yang, 2003; Kadan and Swinkels, 2008). However, belief constraints and welfare interpretations can differ markedly between the two settings.

The real options approach to investment under uncertainty was pioneered by Brennan and Schwartz (1985) and McDonald and Siegel (1986), and has been built upon extensively (Dixit, 1993; Dixit and Pindyck, 1994). Grenadier and Wang (2007) extend this framework to model the investment decisions of hyperbolic entrepreneurs, while Miao (2008) studies agents with Gul and Pesendorfer's (2001) temptation utility.

3 The Model

I first describe the economic environment, followed by the agent's preferences, which may include both present-biasedness and reference dependence. I focus on an optimal stopping problem, where the self-control problem arises purely from the tension between waiting and stopping today. This framework applies to many economic situations, such as those described above - the student pursuing his education, or the person saving a nest egg for retirement.²

3.1 Optimal Stopping

I consider the standard, continuous-time optimal stopping problem, in which an infinitely lived agent has a non-tradeable option to invest in a project. The problem can also be interpreted as a project termination decision - the agent currently holds a project that has a fixed cost of disinvesting and an uncertain payoff or resale value.

At any time t , the agent knows the current value of the project's payoff $x_t \in [0, \infty)$ and decides whether to stop or to wait. In the latter case, the project's payoff evolves as a geometric Brownian motion:

$$dx_t = \mu x_t dt + \sigma x_t dz, \tag{1}$$

where z is a standard Wiener process, μ the average growth rate of x_t , and σ its standard deviation per unit time. At the stopping time \bar{t} , the project yields the lump-sum terminal payoff $x_{\bar{t}}$. The cost of stopping at any time is $I > 0$, and is incurred only at the stopping time. Thus, there is no intertemporal separation of costs and benefit, which would certainly

²For example, a student who is deciding how long to remain in school will incur job search costs when he stops, and will generally obtain a better job outcome from staying in school longer.

exacerbate the self-control problem, but is not necessary to produce intrapersonal conflict. Without loss of generality, there is no interim flow payoff, nor any direct cost incurred prior to stopping. Due to the stochastic nature of the payoff process, there exists an option value of waiting, in the hope that a higher project value will be realized at a later date.

A more realistic setting might include an observable constant or stochastic flow payoff. In the previous examples, a student might incur some small positive or negative flow payoff from going to school; saving might induce negative flows due to decreased consumption before retirement; and a person searching for a spouse might incur negative flow utility until he finds a match. Such an inclusion has no qualitative effect on the results³, so I exclude flow payoffs henceforth.

3.2 Time Preferences

The agent may have present-biased preferences, creating a self-control problem. I model this present-biasedness by following Harris and Laibson (2004), who formulate a continuous-time version of quasi-hyperbolic preferences. At any time s , an agent's preferences are divided into a "present," which lasts from time s to time $s + \tau_s$, and a "future," which arrives at time $s + \tau_s$ and persists forever. The length of the present, τ_s , is stochastic and distributed exponentially with parameter $\lambda \in [0, \infty)$. While the agent knows the distribution governing the length of the present, he is unaware of when the future will arrive ex ante.⁴ When self s 's future arrives at time $s + \tau_s$, he is replaced by a new self who takes control of decision-making. Likewise, the preferences of this self $s + \tau_s$ are divided into a "present" of length $\tau_{s+\tau_s}$ and a "future" that arrives at time $(s + \tau_s) + \tau_{s+\tau_s}$. Hence, when each self's "future" arrives, it "dies" and is replaced by a new self.

Each self s has a stochastic discount function $D_s(t)$:

$$D_s(t) = \begin{cases} e^{-\rho(t-s)} & \text{if } t \in [s, s + \tau_s) \\ \beta e^{-\rho(t-s)} & \text{if } t \in [s + \tau_s, \infty). \end{cases} \quad (2)$$

where $\beta \in [0, 1]$ and $\rho > 0$. To ensure that the agent never finds it optimal to wait forever in the optimal stopping problem, let $\rho > \mu$. The function $D_s(t)$ decays exponentially at the rate ρ throughout, but drops discontinuously at time $s + \tau_s$ to a fraction β of its prior level.

³More precisely, I can include a constant flow payoff $y \in (y, \infty)$, where $y < 0$ is the minimum flow payoff such that the agent stops immediately for any $x_t \geq 0$. Likewise, incorporating an observable stochastic flow payoff that follows a known process leads to the same qualitative results.

⁴The assumption of a stochastic arrival time of the future allows for a stationary solution to the stopping problem, but is not necessary to obtain qualitative results. One can obtain stationarity by imposing a deterministic arrival time instead, though comparative statics with respect to differing time preferences would be less general.

Here, there are two parameters that determine the extent of the self-control problem. First, the parameter β retains the same role it plays in the discrete-time version, measuring how much the future is valued relative to the present. Second, the parameter λ determines the arrival rate of the future, and thus how often preferences change. When $\beta = 1$ or $\lambda = 0$, the preferences described by Equation (2) are equivalent to those of an exponential discounter with discount rate ρ .

3.3 Goals

The agent’s preferences are assumed to be reference-dependent: his utility is composed of both consumption utility, which is based on absolute levels, and of comparison utility, which is concerned with gains and losses relative to a reference point, which here corresponds to his goal. There is much evidence that people react to goals in ways consistent with prospect theory, such as reporting less or more satisfaction with the same outcome due to differing goal levels or prior expectations (Heath et al., 1999; Medvec et al., 1995; McGraw et al., 2005). Equivalently, an agent may incur psychological costs and benefits from performance relative to his goal, such as embarrassment and pride.

In optimal stopping with zero flow payoffs, the agent’s consumption utility upon stopping at time \bar{t} is simply his net terminal payoff, $x_{\bar{t}} - I$. As in Kőszegi and Rabin (2006; 2009), the agent’s comparison utility is closely related to his consumption utility, and is derived by comparing his net terminal payoff at time \bar{t} against his goal at that time, $r_{\bar{t}}$.⁵ Because the payoff process x_t is bounded below at zero and unbounded above, it is impossible for the agent to receive a net terminal payoff below $-I$. Thus, the domain of r_t is $r_t \in [-I, \infty)$. In the language of personal development, the goal is thus “Specific,” in that it is well-defined, and “Measurable,” in that it can be clearly evaluated. His comparison utility increases with the degree to which he exceeds his goal and decreases with the degree to which he falls short of it, a property consistent with empirical evidence (Mento et al., 1992). A key difference is that comparison utility here is simply a linear function, given by

$$\eta(x_{\bar{t}} - I - r_{\bar{t}}), \tag{3}$$

where $\eta \geq 0$, implying that the agent exhibits neither loss aversion nor diminishing sensitivity to gains and losses. Although the absence of these two features deviates from Kahneman and Tversky’s (1979) value function, it will demonstrate that *neither is necessary for goals*

⁵Alternatively, he could compare his gross terminal payoff $x_{\bar{t}}$ against his goal for the gross terminal payoff at that time, or separately compare the terminal payoff and cost using this comparison utility function, and the results would be unchanged.

to affect behavior in a meaningful way. Indeed, one of the main points of the paper is that the effects of goal-setting can be understood completely independently of loss aversion.

The parameter η can be interpreted as the degree to which the agent is reference dependent. It can also be seen as a measure of salience or “goal commitment,” which is broadly defined in psychology as the degree to which a person is determined to achieve a goal. A central concept of goal-setting theory is that goal difficulty has little effect on behavior if goal commitment is not present (Locke and Latham, 2002).⁶ The absence of goal commitment ($\eta = 0$) corresponds to the absence of reference dependent preferences, when the existence of a goal has no effect on utility, and consequently, his behavior. We can also interpret $\eta = 0$ as the case in which the agent has no goal or an ill-defined goal. Without a well-defined basis against which to make a comparison, it follows that an agent cannot incur comparison utility in this case. Such an interpretation accords with results regarding the ineffectiveness of vague goals on motivation and effort (Latham and Locke, 1991; Mento et al., 1992). Here, I treat η as a fixed parameter and analyze the agent’s subsequent behavior and welfare. The demand for personal development services and products can be interpreted as an attempt by individuals to improve self-regulation and welfare by changing η .⁷ This interpretation is not inconsistent with this assumption, insofar as an agent must determine how he would fare for any given η to determine the value of attempting to change his initial η . In this vein, I later consider the agent’s preferences over η from an ex-ante perspective.

I assume that the agent only incurs comparison utility when he stops and receives the net terminal payoff. Although he is aware that he will incur comparison utility upon stopping, he does *not* incur any while waiting. This assumption accords with the notion from mental accounting that individuals do not necessarily “feel” gains and losses until they have been realized (Thaler, 1999; Odean, 1998; Barberis and Xiong, 2012).

For simplicity, overall utility is taken to be additively separable in its two components. The agent’s total utility at the stopping time is thus

$$x_{\bar{t}} - I + \eta(x_{\bar{t}} - I - r_{\bar{t}}). \quad (4)$$

At any time s , the goal r_s is taken as given by self s and cannot be changed during his entire “lifetime,” having been set by his previous self. Similarly, the goal that self $s + \tau_s$ inherits,

⁶Similarly, McClelland (1962) defines those high in “achievement motivation,” or n Ach, as those derive satisfaction from setting achievable and measurable goals for themselves.

⁷For example, the suggestively titled “The Magic Lamp: Goal Setting for People Who Hate Setting Goals” Ellis (1998) purports to help individuals improve their goal commitment. The inside flap declares, “**The Magic Lamp** . . . [combines] the methods of goal setting with the magic of making your wishes come true.” Despite the extravagance of the claim, it suggests that goal commitment is necessary and that individuals may be able to change their existing goal commitment with some effort.

denoted $r_{s+\tau_s}$, is set by self s , where τ_s , the lifespan of self s , is stochastically determined and a priori unknown to self s .

That the agent cannot change an inherited goal implies that it can provide a degree of internal motivation to his (present-biased) future selves. Such “goal stickiness” is necessary: if the agent could simultaneously make the stopping decision and set a goal for himself for the *current* period, this goal would have no effect on his current behavior, since his present bias implies that he has no desire to behave more patiently today. He would also have no means by which to impose ex-ante preferences over time-consistent behavior. In Section 5, I relax this assumption and study the case of partial commitment, in which the agent is not always committed to evaluate himself against an inherited goal.

3.4 Expectations

I assume that each self’s inherited goal is constrained by rational expectations. Each self forms an expectation of his future payoff if he does not stop himself, which his descendant inherits and evaluates himself against if he stops. I consider the case of sophistication, in which the agent is fully aware of future present-biasedness.⁸ Because he observes the current project payoff perfectly and correctly anticipates his actions, he holds no ex-ante uncertainty over the outcome upon stopping. Thus, the goal that he sets for his future self must, in equilibrium, coincide with his net terminal payoff upon stopping.

That the goal must be consistent with rational expectations implies that the agent cannot consistently fool himself about what he can or cannot achieve - he must set goals that are realistic. That is, the agent’s goal must be “Attainable.” The assumption of expectational goals is consistent with a recent body of experimental (Abeler et al., 2011; Ericson and Fuster, 2010; Sprenger, 2010) and empirical (Pope and Schweitzer, 2011; Card and Dahl, 2011) studies, which find support for endogenous reference points constrained by rational expectations (Kőszegi and Rabin, 2006, 2009). Based on lab and field experiments, Latham and Locke (1991) conclude that goal choice is an integration of what one wants and what one believes is possible, suggesting that goals are realistic. Carrillo and Dewatripont (2008) also discuss the tension between foresight and the credibility of promises in intrapersonal games, arguing that agents cannot simultaneously anticipate future behavior and fool future selves. Note, however, that these assumptions do *not* necessarily imply that each self must have the same goal. Each self cannot change the goal that he inherits, but is free to choose a different one for his future self if he so desires, as long as that he believes that it is realistic.

⁸In an earlier version (Hsiaw, 2010b), I also consider the case of naivete, in which the agent is completely unaware of future present-biasedness, and show that the major qualitative results from sophistication apply to naivete.

Because each self inherits his goal from his predecessor, it is necessary to specify the source of the agent's goal when he is first able to stop the project. I assume that there exists a "self 0," an ex-ante self, who learns that the stopping opportunity will present itself and forms an expectation of how he will behave once the option becomes available for exercise.

Given such preferences, at any time s the agent chooses the stopping rule that determines a (random) stopping time \bar{t} to maximize the expected present value of his overall utility:

$$\max_{\bar{t}} E_s \{ D_s(\bar{t}) [x_{\bar{t}} - I + \eta(x_{\bar{t}} - I - r_{\bar{t}})] \}, \quad (5)$$

where E_s denotes the conditional expectation at time s and $D_s(\bar{t})$ is given by Equation (2). When the agent has hyperbolic time preferences, he is prone to stopping too early because he undervalues the future, and thus the option value of waiting (Grenadier and Wang, 2007). The question of interest is to what extent setting goals can attenuate this problem.

4 Full Commitment

When the agent is quasi-hyperbolic and is fully committed to his inherited goal, in that he cannot change and must evaluate himself against it, the problem takes on the nature of a dynamic game between successive selves. While there could exist equilibria in which each self chooses a different stopping strategy, I focus on the most natural equilibrium, namely a stationary Markov equilibrium in which each self employs the same threshold strategy. In Appendix A.4, I verify that the stationary stopping rule is in fact optimal.

Since the geometric Brownian motion is continuous, the project value cannot jump discontinuously from one moment in time to the next. This implies that the sophisticated agent has no uncertainty over the net terminal payoff from stopping: if every self uses the same threshold \bar{x} , then $x_{\bar{t}} = \bar{x}$, so the expected net terminal payoff is $\bar{x} - I$. The source of ex ante uncertainty is the *timing* of this stopping - that is, when he chooses to stop. Furthermore, every self will inherit, set, and meet the same goal in a stationary equilibrium, so the inherited expectational goal satisfies $r_{\bar{t}} = \bar{x} - I$ in equilibrium.

To construct the equilibrium, I solve the intrapersonal game backwards. Each self anticipates that his descendants will employ thresholds that maximize their own current benefit of waiting, so they will face a problem identical to his own. A stationary solution is thus a fixed point such that current and future selves employ a common threshold.

Because each self controls the stopping decision in the present, and cares about - but cannot directly control - those of the future, two value functions are required to describe the intrapersonal problem. Let all descendants inherit the goal \hat{r} . Then the Bellman equation

for the continuation value function $v(x, \hat{r})$ is

$$v(x, \hat{r}) = \max\{x - I + \eta(x - I - \hat{r}), e^{-\rho dt} E[v(x + dx, \hat{r})]\}. \quad (6)$$

This continuation value function describes each self s 's consideration of future selves, following the random arrival of the future at time τ_s . Beyond time τ_s , he discounts future utility flows exponentially. Hence, the continuation value function also describes how the current self *prefers* future selves to evaluate payoff streams - by discounting exponentially. Thus, v describes the agent's preferences over the future from an ex ante perspective, including those of self 0. If the agent were time consistent, then all selves' preferences would coincide and he would choose the optimal strategy to maximize v , given the current project value x and his inherited goal \hat{r} .

However, if the agent is present-biased ($\beta < 1$), he maximizes a different value function, w , that overweights the present relative to the future. Denoting the goal inherited by the current self by \bar{r} , the Bellman equation that describes this current value function $w(x, r)$ is

$$w(x, \bar{r}) = \max\{x - I + \eta(x - I - \bar{r}), (1 - e^{-\lambda dt})e^{-\rho dt} \beta E[v(x + dx, \hat{r})] + (e^{-\lambda dt})e^{-\rho dt} E[w(x + dx, \bar{r})]\}. \quad (7)$$

Given the current x and an inherited goal \bar{r} , and anticipating that his future selves will inherit \hat{r} (with the knowledge that he sets \hat{r} for his immediate descendant), the current self chooses the maximum of the current total utility from stopping and the expected discounted value of waiting for a higher realization of x , where this discounting discontinuously drops by the factor β upon the random arrival of the future. A future self arrives in the next instant dt with probability $1 - e^{-\lambda dt}$, while the current self remains in control with probability $e^{-\lambda dt}$.⁹

4.1 Stationary Equilibrium

In Appendix A.1, I derive the continuation and current value functions that describe the agent's ex-ante and present-biased preferences, respectively. In a stationary equilibrium, the sophisticated agent knows that all current and future selves employ the same threshold, which implies that $\bar{x} = \hat{x}$. To make clear the effect of goal-setting on the optimal stopping rule, I first consider the sophisticate's stopping threshold given a fixed goal level (i.e. $r = \bar{r} = \hat{r}$),

⁹Alternatively, consider a parent who derives utility from the payoff stream of his descendants. The continuation value v describes his evaluation of the stream of his descendants' utilities - he discounts them exponentially. Thus he prefers that every descendant evaluate the payoff streams from his entire family line, including his own, exponentially. When his child becomes the decision-maker, he discounts his own descendants' utilities exponentially, just as his parent did. But he also underweights the stream of his descendants' utilities relative to his own by the factor β , in disagreement with his parent's wishes. Thus, w describes the child's evaluation of the payoff stream from his entire family line, including his own.

denoted \bar{x}^{SE} , provided that this goal is set ex ante and does not change during the stopping decision. Thus, it describes a sophisticate’s response to any externally set goal, which can differ from the agent’s actual final payoff. Next, I derive the sophisticate’s stopping rule when his goals are set internally, denoted \bar{x}^{SI} , and therefore are required to be met in equilibrium.

4.1.1 Exogenous Goals

Letting $r = \bar{r} = \hat{r}$ and imposing the fixed point condition that $\bar{x} = \hat{x} \equiv \bar{x}^{SE}$, the optimal threshold in response to an externally set goal $r \geq 0$, denoted \bar{x}^{SE} , is

$$\bar{x}^{SE} = \left(\frac{\bar{\gamma}}{\bar{\gamma}-1}\right)I + r\left(\frac{\eta}{1+\eta}\right)\left(\frac{\bar{\gamma}}{\bar{\gamma}-1}\right), \quad \text{with } \bar{\gamma} \equiv \beta\gamma_1 + (1-\beta)\gamma_2, \quad (8)$$

where the parameter $\gamma_1 > 1$ is the positive root of the quadratic equation

$$\frac{1}{2}\sigma^2\gamma_1^2 + \left(\mu - \frac{1}{2}\sigma^2\right)\gamma_1 - \rho = 0, \quad (9)$$

reflecting the fact that from an ex-ante perspective, the agent discounts the future exponentially at the rate ρ . The parameter γ_2 is the positive root of the quadratic equation

$$\frac{1}{2}\sigma^2\gamma_2^2 + \left(\mu - \frac{1}{2}\sigma^2\right)\gamma_2 - (\rho + \lambda) = 0. \quad (10)$$

The only difference between the quadratic equation for γ_2 and that of γ_1 is the presence of the parameter λ , the hazard rate for the arrival of the future. It is apparent that $\gamma_2 \geq \bar{\gamma} \geq \gamma_1$, with equality only if the future never arrives ($\lambda = 0$), i.e. preferences never change. The parameter γ_2 reflects the fact that each self’s expected “lifetime” shortens with λ , while the degree to which this feature affects the stopping decision is determined by the degree of present-biasedness, measured by $1 - \beta$. If the agent is time-consistent ($\bar{\gamma} = \gamma_1$) and not reference-dependent ($\eta = 0$), \bar{x}^{SE} equals $x^* \equiv \left(\frac{\gamma_1}{\gamma_1-1}\right)I$, the solution to the standard stopping problem.

Equation (8) makes clear the effect of goal-setting on the optimal stopping rule. The first term is the agent’s stopping threshold in the absence of reference dependence or a goal: if $\eta = 0$, then $\bar{x}^{SE} = \left(\frac{\bar{\gamma}}{\bar{\gamma}-1}\right)I < x^*$.¹⁰ A hyperbolic discounter without reference dependent preferences stops earlier than the standard agent, and this impatience is exacerbated as the present is more overweighted, i.e. as β decreases, and as preferences change more frequently, i.e. as λ increases. Recall that it is the combination of parameters β and λ that determines the degree of the self-control problem. In fact, Equation (8) reveals that $\bar{\gamma}$ is a sufficient

¹⁰If the agent is not reference dependent ($\eta = 0$), then Equation (8) is identical to the sophisticate’s threshold obtained by Grenadier and Wang (2007).

statistic for measuring the sophisticated agent's degree of impulsiveness.¹¹ The second term in Equation (8) is the effect of the goal on the stopping threshold. Because an agent is motivated to avoid settling for a lower project value only if there exists a potential comparative penalty from falling short, the goal r only induces more patient behavior as long as $r > 0$, even in the case of $\bar{\gamma} = \infty$. No potential penalty is imposed when $r = 0$, so behavior is unchanged by this goal and it is equivalent to having no goal.¹²

Proposition 1 *In a stationary equilibrium with exogenous goal r , the sophisticate's stopping threshold exhibits the following properties:*

1. *The threshold decreases with the degree of impulsiveness: $\frac{\partial \bar{x}^{SE}}{\partial \bar{\gamma}} < 0$.*
2. *The threshold increases with the level of the goal: $\frac{\partial \bar{x}^{SE}}{\partial r} > 0$.*
3. *Responsiveness to a goal increases with the degree of reference dependence: $\frac{\partial^2 \bar{x}^{SE}}{\partial r \partial \eta} > 0$.*
4. *Responsiveness to a goal decreases with the degree of impulsiveness: $\frac{\partial^2 \bar{x}^{SE}}{\partial r \partial \bar{\gamma}} < 0$.*

It is intuitive that the agent's threshold should decrease with his degree of impulsiveness. First, he undervalues the future himself. Second, being sophisticated, he anticipates that future selves will undervalue their own futures as well, decreasing the value of waiting even further. A second intuitive result is that, regardless of his degree of impulsiveness (including if he is not), the agent's threshold increases with the level of the goal if he is reference dependent. Raising r increases the potential cost of settling for a lower project value. Thus, a goal induces more patient behavior by providing an additional incentive to wait for a higher realization of the project value. This result is consistent with experimental evidence that task performance increases with the goal difficulty, whether externally- or self-set. Locke and Latham (2002) even find a positive linear relationship between goal difficulty and performance.

Consistent with empirical evidence (Klein et al., 1999), responsiveness to a goal is increasing in the agent's degree of reference dependence. An agent who cares more about falling short of his goal is more motivated to avoid such an outcome than one who cares less about this comparison. Finally, an agent's responsiveness to a goal is decreasing in his degree of impulsiveness, illustrating the interaction between present-biasedness and reference dependence. A more impulsive agent values the future less, so he has a weaker incentive to avoid incurring a comparative penalty assessed in the future. Thus, he not only exhibits more impatience in the absence of a goal, but is also less responsive to a given goal.

¹¹Since $\bar{\gamma}$ is decreasing in β and increasing in λ , comparative statics on each parameter separately yield the same qualitative results.

¹²This feature makes clear that η in itself does not simply act as a subsidy to reaching a higher net x .

4.1.2 Endogenous Goals

When the goal is internally set, rational expectations implies that each self correctly anticipates his descendant's threshold strategy, so that $\bar{r} = \bar{x} - I$ and $\hat{r} = \hat{x} - I$. But since $\bar{x} = \hat{x}$, this implies that $r = \bar{r} = \hat{r} = \bar{x} - I$. Each self correctly expects his descendants to use the same stopping rule as he does, so every self inherits and meets the same goal in equilibrium. That is, whether goals are externally or internally set, sophistication implies that $\bar{x} = \hat{x}$ and $\bar{r} = \hat{r}$. Thus, the optimal threshold with internally set goals, denoted \bar{x}^{SI} , must satisfy $r = \bar{x}^{SI} - I$ and can be derived by imposing this condition on Equation (8), yielding

$$\bar{x}^{SI} = \left(\frac{\bar{\gamma}}{\bar{\gamma} - 1 - \eta} \right) I, \quad \begin{array}{l} \text{with } \bar{\gamma} \equiv \beta\gamma_1 + (1 - \beta)\gamma_2 \\ \text{and } \eta < \bar{\gamma} - 1, \end{array} \quad (11)$$

and the equilibrium value functions w^{SI} and v^{SI} :

$$w^{SI}(x, r = \bar{x}^{SI} - I) = \begin{cases} \beta(\bar{x}^{SI} - I)\left(\frac{x}{\bar{x}^{SI}}\right)^{\gamma_1} + (1 - \beta)(\bar{x}^{SI} - I)\left(\frac{x}{\bar{x}^{SI}}\right)^{\gamma_2} & \text{if } x < \bar{x}^{SI} \\ x - I + \eta(x - \bar{x}^{SI}) & \text{if } x \geq \bar{x}^{SI}, \end{cases} \quad (12)$$

$$v^{SI}(x, r = \bar{x}^{SI} - I) = \begin{cases} (\bar{x}^{SI} - I)\left(\frac{x}{\bar{x}^{SI}}\right)^{\gamma_1} & \text{if } x < \bar{x}^{SI} \\ x - I + \eta(x - \bar{x}^{SI}) & \text{if } x \geq \bar{x}^{SI}. \end{cases} \quad (13)$$

The value of Equation (12) in its wait region is the project's expected present value to the current self, given the current value of the project's payoff, $x < \bar{x}^{SI}$, and the optimal threshold \bar{x}^{SI} . This is essentially the weighted average of two time-consistent option values, where the first, weighted by β , uses the discount rate ρ , which is reflected in γ_1 . The second, weighted by $1 - \beta$, uses the discount rate $\rho + \lambda$, which is reflected in γ_2 . The value of Equation (13) in its wait region is the expected present value of the option to stop, using only the discount rate ρ , reflected in γ_1 . Thus, v^{SI} also represents the option value that each current self would prefer his future selves to use from an ex ante perspective (or would like to commit them to), though he is resigned to the knowledge that they will maximize w rather than v .

Figure 1 depicts the equilibrium value functions when goals are self-set. The presence of reference dependence increases the marginal value of waiting, increasing the slope of both the continuation and current value functions upon stopping, and thus the incentive to wait. The fact that v lies above its respective w reflects the fact that an exponential discounter values the option more than a present-biased agent. Independently of η , the slope of v relative to that of w at the equilibrium stopping threshold reflects the fact that the exponential discounter prefers to wait longer than the present-biased agent, and thus, that the present-biased agent prefers his future selves to wait longer than they actually do.

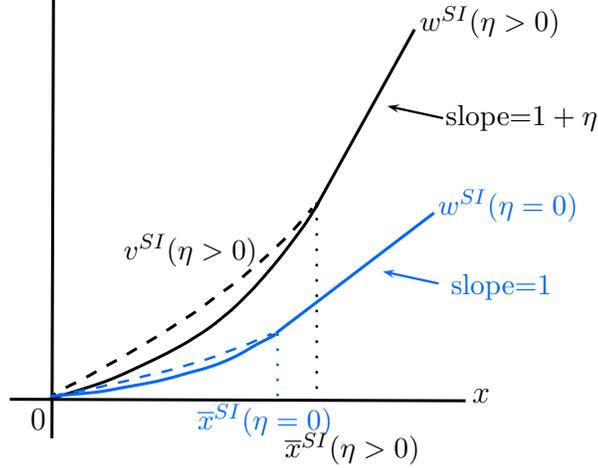


Figure 1: Option values with endogenous goals when $\beta < 1$ for varying degrees of η . The black lines represent the value functions when $\eta > 0$. The blue lines represent the value functions when $\eta = 0$. The solid lines depict the current value functions w , which are smooth and continuous everywhere. The dashed lines depict the continuation values v , which are kinked at the equilibrium thresholds \bar{x}^{SI} and coincide with their respective current value functions w thereafter.

Proposition 2 *By inducing more patient behavior, reference dependence attenuates impulsiveness in a stationary equilibrium with endogenous goals: $\frac{\partial \bar{x}^{SI}}{\partial \eta} > 0$.*

From Equation (11), it is clear that the equilibrium threshold increases with the degree of reference dependence for any given degree of impulsiveness if λ is finite. An agent with greater reference dependence has a stronger incentive to meet his goal, since he puts more weight on the comparative disutility from falling short. It is only in the case of $\bar{\gamma} = \infty$ that the agent, with infinite impatience and finite reference dependence, is unaffected by goal-setting. As noted in Equation (8), the agent *is* responsive to a goal even in this case, as long as $r > 0$. But when the goal is self-set, his anticipation of extreme impatience makes him unable to set a realistic penalty to improve his patience, so his behavior is unchanged by reference dependence.

4.2 Welfare

From an ex-ante perspective, the agent, no matter how severe his degree of present-biasedness or reference dependence, prefers that his future selves behave according to a time-consistent strategy. Therefore, I use the preferences of self 0, which coincide with the ex ante optimum, to evaluate the agent's welfare. Since all selves possess reference-dependent preferences, self 0 must take both consumption and comparison utility into account when evaluating

welfare. Such an analysis allows us to evaluate the welfare consequences for a heterogeneous population of individuals, with varying degrees of present-biasedness and goal commitment. It also allows self 0 to determine the value of adjusting his degree of goal commitment η to improve future behavior, since this choice may be costly.

Sophistication implies that the agent always anticipates that he will meet his self-set goal in equilibrium. He is thus aware that once he actually stops at some point in the future, his comparative utility will equal zero in equilibrium, so his overall utility will consist only of the net terminal payoff, just as in the $\eta = 0$ case. Therefore, the ex-ante optimum maximizes $v^{SI}(x, r)$, the same objective of a standard agent who is dynamically consistent and not reference dependent. Thus, the first-best threshold for any sophisticated agent, no matter his degree of present-biasedness or goal commitment, coincides with the threshold that a standard agent chooses: $x^* = (\frac{\gamma_1}{\gamma_1 - 1})I$. This implies that for any degree of impulsiveness, there exists a degree of countervailing goal commitment that enables the agent to employ the first-best threshold strategy (i.e. $\bar{x}^{SI}(\eta^*) = x^*$). Unsurprisingly, this optimal goal commitment η^* is increasing in $\bar{\gamma}$ - as impulsiveness increases, the degree of reference dependence required to attenuate it increases.

Proposition 3 *For every $\bar{\gamma} \in (1, \infty)$, there exists an η^* such that the reference-dependent agent with $\eta = \eta^*$ achieves the first-best, given by*

$$\eta^* = \frac{\bar{\gamma} - \gamma_1}{\gamma_1}.$$

This analysis makes clear the differential effect of increasing the degree of reference dependence η versus increasing the stopping cost I . While both have the same qualitative effect of increasing \bar{x}^{SI} , only the former improves welfare by closing the gap between equilibrium and first-best behavior if $\eta < \eta^*$. In contrast, the latter has no effect on the self-control problem, since both the equilibrium *and* first-best thresholds increase with I .

On the other hand, the degree of reference dependence can also be sufficiently high that the agent waits *longer* than is first-best. Consider now an agent who is dynamically consistent ($\bar{\gamma} = \gamma_1$) and reference dependent ($\eta > 0$). Although reference dependence does not distort final overall utility directly, it changes the *marginal* value of stopping. Because the agent has an incentive to avoid incurring comparative disutility, the marginal value of waiting at the first-best threshold x^* exceeds the marginal value of stopping, so the agent waits longer. Although he achieves a payoff that exceeds x^* , its ex-ante discounted value is lower than the first-best. Hence, reference dependence distorts his behavior *away from the first-best* and causes him to be overly patient.¹³ Since a time-consistent agent has no

¹³In contrast, Kőszegi and Rabin (2009) find that in the absence of uncertainty over final payoffs, reference

self-control problem, reference dependence offers no beneficial value, so $\eta^* = 0$ if $\bar{\gamma} = \gamma_1$. The same argument can clearly be applied to present-biased agents, if η overcompensates for the conflict in time preferences between current and future selves. Moreover, reference dependence can be so high that even an impulsive agent would be better off in the absence of goal-setting. That is, when $\eta > \bar{\eta}$, he would wait so long under goal-setting that he is actually *better off* if *he cannot set goals for himself* (i.e. $\eta = 0$) and behaves impatiently. Thus, goal-setting can itself be a source of intrapersonal conflict, since it can cause an agent to wait longer than is optimal from an ex ante perspective. The level of reference dependence required to be detrimental to the agent’s welfare is increasing in his degree of impulsiveness.

Corollary 1 *For every $\bar{\gamma} \in (1, \infty)$, there exists a range of η such that the sophisticated agent waits longer than the first-best: $\eta^* < \eta < \bar{\gamma} - 1$. Moreover, he is strictly worse off under goal-setting if $\bar{\eta} < \eta < \bar{\gamma} - 1$, where $\bar{\eta} \geq \eta^*$ is increasing in $\bar{\gamma}$ and is defined by the following condition:*

$$\left(\frac{\bar{\gamma} - 1}{\bar{\gamma} - 1 - \bar{\eta}}\right)^{\gamma_1 - 1} \left(\frac{1}{1 + \bar{\eta}}\right) - 1 = 0.$$

In particular, a time-consistent agent stops at $\bar{x}^{SI} = \left(\frac{\gamma_1}{\gamma_1 - 1 - \eta}\right)I$ and is strictly worse off if he is reference-dependent ($\eta > 0$).

Although the sophisticate’s self-set goal level is pinned down by rational expectations, he may further regulate his behavior by seeking to adjust his degree of reference dependence, behavior that can be strictly welfare-improving. Beyond providing strategies to determine and set well-defined goals, personal development attempts to aid such endeavors by suggesting concrete ways to increase goal salience, such as writing down and reviewing goals daily and continually visualizing their achievement (Tracy, 2003). Prescriptively, welfare improvement can be achieved purely through programs or services such as those offered by personal development, which educate individuals about appropriate goal-setting.

While there is much focus on the need for reference dependence, there is evidence that the converse problem is recognized as well. In psychology, dysfunctional perfectionism is defined as “overdependence of self-evaluation on the determined pursuit of personally demanding, self-imposed, standards in at least one highly salient domain, despite adverse consequences” (Shafran et al., 2002). Shu (2008) finds that decision-makers can search too long when they have an ideal focal outcome in mind. Consistent with the idea that η can be detrimentally high, Goldsmith (2008) discusses the prevalence of “goal obsession” among successful executives, many of whom regret sacrificing health or family life in the pursuit of their careers.

dependence can lead to apparently impatient behavior, even without present-biased preferences. There, the reference point is a vector of beliefs over time, and the agent derives comparison utility in each period of a consumption-savings problem. If he values contemporaneous comparison utility more than future comparison utility, then he may overconsume relative to the ex-ante optimum.

5 Partial Commitment

Thus far, I have assumed that the agent’s goal is inherited from his immediate predecessor, and that this inherited goal is fully determined by rational expectations. Because there is no uncertainty over the realized outcome, regardless of whether *beliefs* are correct or not, it is unclear how the agent could simultaneously anticipate a particular outcome with certainty and knowingly set a goal that differs from that expectation ex ante.

However, it is possible that having reference dependent preferences may be only a partial commitment device, if the present-biased self is not fully constrained to evaluate an inherited goal. If the current self can set a goal for himself that is not constrained by his expectations, then he chooses it to maximize his current value function, which reflects present-biasedness. Such a goal can represent an individual’s aspiration or motivated belief (Caplin and Leahy, 2001; Köszegi, 2006; Brunnermeier and Parker, 2005), rather than an expectation about what he will attain. Thus, it may arguably be divorced from expectations without logical conflict, so that he can strategically set goals even if he is aware that they are unrealistic.

Let r_w denote the goal that maximizes the current value function, $w(\cdot)$. Since the project payoff process is bounded below by zero, the lowest net terminal payoff the agent can possibly receive is $-I$, if he stops the process when $x = 0$. Thus, the domain of r_w is again $r_w \in [-I, \infty)$. His decision involves a trade-off between consumption and comparison utility. Choosing a more ambitious goal increases his material payoff by incentivizing him to wait for a higher project payoff, but reduces his comparison utility.¹⁴ Let r_{RE} denote the goal that the agent inherits from his immediate predecessor. As in the full commitment model, the inherited goal must satisfy rational expectations: $r_{RE} = \bar{x} - I$.

When the agent inherits the realistic goal r_{RE} , the probability with which he evaluates himself against r_{RE} upon stopping is given by $p \in [0, 1]$. But with probability $1 - p$, the agent instead selects and evaluates himself against the goal r_w upon stopping. To obtain a stationary solution, I assume that the agent is aware of the probability of evaluating r_w instead of r_{RE} , but does not know which goal he will actually evaluate himself against when he makes the decision to stop. Thus, when he decides to stop, he expects to evaluate himself against r_{RE} with probability p , and against r_w with probability $1 - p$. If the current self does not stop within his own “lifetime,” he sets the realistic goal r_{RE} for his future self, because he is fully aware of future misbehavior ex ante. The probability p represents the degree to which the agent’s goal evaluation is constrained by his expectations, rather than affected by

¹⁴Brunnermeier and Parker (2005) consider the optimal choice of subjective beliefs when the agent faces a trade-off between material outcomes and belief-based utility. In their setting, there is no incentive to distort beliefs in order to change actions. The only motivation for belief distortion is the benefit for anticipatory utility flows, so any belief distortion decreases material benefits and pessimistic beliefs can only hurt.

motivated beliefs which are subject to present bias. The full commitment case explored in Section 4 corresponds to the $p = 1$ case; when $p = 0$, the agent's goal evaluation is entirely unconstrained by rational expectations and is driven by present bias. Thus, each self has a limited ability, increasing in p , to influence the behavior of future selves through the goal r_{RE} . The equilibrium value of r_{RE} varies with p , since the agent's expectations must adjust for the fact that his stopping behavior will be affected by the possibility of evaluating the motivated goal r_w .

5.1 Stationary Equilibrium

The goal r_w that maximizes the agent's current value function is the lowest possible: $r_w = -I$ (proof provided in Appendix A.5). Intuitively, the agent weighs a higher goal's positive effect on consumption utility through improving patience against its negative effect on comparison utility. While the present-biased agent prefers that *future* behavior be more patient than it is currently, he sees no need to change current behavior. Since r_w does not affect future behavior, he sees no need to set a goal to raise consumption utility, so he prefers to set the lowest goal possible to maximize comparison utility. Although this distorts consumption utility, the value of obtaining higher comparison utility earlier dominates.

Let \bar{x}^{PI} denote the agent's stopping threshold under partial commitment, which is

$$\bar{x}^{PI} = \left(\frac{\bar{\gamma}}{\bar{\gamma} - 1 - \eta(1 - \bar{\gamma}(1 - p))} \right) I, \quad \begin{array}{l} \text{with } \bar{\gamma} \equiv \beta\gamma_1 + (1 - \beta)\gamma_2 \\ \text{and } \eta < \bar{\gamma} - 1. \end{array} \quad (14)$$

When the agent is more impulsive, he undervalues the future himself and anticipates that he will also do so in the future, leading him to stop earlier. This result carries over from the full commitment case intuitively. He is more patient when realistic goal evaluation is more likely to occur. Just as in the full commitment case, evaluation against a higher goal induces patience because it provides an additional incentive to wait, to avoid negative comparison utility. Since r_w is unrealistically low, the agent waits longer when the likelihood of evaluating the higher, realistic goal r_{RE} is higher.

Proposition 4 *In a stationary equilibrium with partial commitment, the sophisticate's stopping threshold (\bar{x}^{PI}) exhibits the following properties:*

1. *The threshold decreases with the degree of impulsiveness: $\frac{\partial \bar{x}^{PI}}{\partial \bar{\gamma}} < 0$.*
2. *The threshold increases with the likelihood of realistic goal evaluation: $\frac{\partial \bar{x}^{PI}}{\partial p} > 0$.*
3. *If $p > \frac{\bar{\gamma}-1}{\bar{\gamma}}$, the threshold increases with the degree of reference dependence: $\frac{\partial \bar{x}^{PI}}{\partial \eta} > 0$.
If $p < \frac{\bar{\gamma}-1}{\bar{\gamma}}$, the threshold decreases with the degree of reference dependence: $\frac{\partial \bar{x}^{PI}}{\partial \eta} < 0$.*

In contrast to the full commitment case, the effect of reference dependence depends on the interaction between the likelihood of realistic evaluation and the extent of the self-control problem. Since r_w is unrealistically low, the stopping threshold decreases with its likelihood of being evaluated, because it presents a weaker incentive to wait relative to the realistic goal, r_{RE} . Thus, when evaluating r_w is sufficiently likely, placing more weight on comparison utility leads the agent to stop earlier, to obtain this positive comparison utility. Note that in this case, it is possible for the agent to obtain a negative terminal payoff ($\bar{x}^{PI} < I$) because obtaining positive comparison utility outweighs it¹⁵. When evaluating r_w is sufficiently unlikely, placing more weight on comparison utility leads the agent to stop later, since the incentive to avoid negative comparison utility from falling short of the realistic goal r_{RE} dominates. The threshold at which the prospect of evaluating r_w changes the effect of reference dependence depends on the degree of impulsiveness, which weakens an agent's responsiveness to a goal. A more impulsive agent values the future less, so he has a weaker incentive to avoid incurring the comparative penalty posed by a realistic goal, which is assessed in the future. This implies that as the agent becomes more impulsive, evaluating r_w must be increasingly unlikely in order for reference dependence to improve patience.

5.2 Welfare

From the ex ante, self-0 perspective, the agent prefers a stopping threshold that maximizes the continuation value, $v(x, r)$. In contrast with the full commitment case, however, the agent anticipates that he will exceed his expected goal on average rather than meet it exactly, due to the possibility that he evaluates himself against the excessively low goal r_w . Thus, his overall utility consists of both his net terminal payoff and expected positive comparison utility, where the latter decreases with p . The presence of this positive comparison utility distorts the degree of reference dependence required to achieve the first-best, denoted η^{PI} , since reference dependence now affects both the net terminal payoff, through its effect on when to stop, and the expected comparison utility, since it governs the weight on this psychological payoff. In the full commitment case ($p = 1$), the optimal degree of reference dependence is zero when the agent faces no self-control problem ($\bar{\gamma} = \gamma_1$), because it only distorts the net terminal payoff without affecting comparison utility. But when $p < 1$, it is ex-ante optimal to place positive weight on goal comparison even in the absence of impulsiveness, in order to obtain positive comparison utility. The value of obtaining some positive psychological payoff earlier outweighs the loss of material payoff, which would be received later.

¹⁵Specifically, $\bar{x}^{PI} < I$ when $p < 1 - (\frac{1+\eta}{\eta})(\frac{1}{\bar{\gamma}})$.

Proposition 5 *If $p > \frac{\bar{\gamma}-1}{\bar{\gamma}}$, there exists an η^{PI} such that the reference-dependent agent with $\eta = \eta^{PI}$ achieves the first-best stopping threshold of $\bar{x}^{PI} = (\frac{\gamma_1}{p(\gamma_1-1)})I$, given by*

$$\eta^{PI} = \frac{\bar{\gamma} - \gamma_1 + (1-p)\bar{\gamma}(\gamma_1 - 1)}{\gamma_1(1 - \bar{\gamma}(1-p))}.$$

If $p < \frac{\bar{\gamma}-1}{\bar{\gamma}}$, then $\eta^{PI} = \infty$ and the optimal ex-ante stopping threshold is $\bar{x}^{PI} \rightarrow 0$.

When realistic goal evaluation is sufficiently likely, the incentive to obtain a higher net terminal payoff by waiting longer is stronger than the incentive to obtain positive comparison utility earlier. For this reason, increasing goal commitment to η^{PI} attenuates the effect of impulsiveness when $p > \frac{\bar{\gamma}-1}{\bar{\gamma}}$, but can be detrimental when too high. Unsurprisingly, then, the need for reference dependence to achieve the first-best increases with impulsiveness. As the likelihood of realistic evaluation increases, optimal reference dependence decreases as well, due to two effects. First, the incentive to add weight to increase expected positive comparison utility diminishes. Second, the motivating effect of the more ambitious r_{RE} strengthens, inducing more patience. Since in this case the incentive to avoid falling short of r_{RE} by waiting for a higher payoff dominates, then decreasing reference dependence weakens this effect. Thus, when realistic goal evaluation is sufficiently likely, the first-best stopping threshold \bar{x}^{PI} is decreasing in p , and only coincides with x^* when there is full commitment.

Proposition 6 *If $p > \frac{\bar{\gamma}-1}{\bar{\gamma}}$, the optimal degree of reference dependence, η^{PI} , exhibits the following properties:*

1. *It increases with the degree of impulsiveness: $\frac{\partial \eta^{PI}}{\partial \bar{\gamma}} > 0$.*
2. *It decreases with the likelihood of realistic goal evaluation: $\frac{\partial \eta^{PI}}{\partial p} < 0$.*

When realistic goal evaluation is sufficiently unlikely, the incentive to maximize positive comparison utility overshadows the value of the material payoff, for which patience is required, *even from an ex-ante perspective*. For this reason, the detrimental effects of impulsiveness, which diminishes the material payoff, do not arise. In this case, the agent prefers to increase reference dependence as much as possible, in order to maximize the positive comparison utility that he derives from comparison to the unrealistically low goal of r_w . At this ex ante first-best, the agent prefers to stop as early as possible to obtain it. But because his reference dependence is limited ($\eta < \bar{\gamma} - 1$), and impulsiveness leads him to stop earlier, then impulsiveness is actually *welfare-improving*, as is any degree of goal commitment.

Corollary 2 *If $p < \frac{\bar{\gamma}-1}{\bar{\gamma}}$, then impulsiveness is welfare-improving.*

Whether impulsiveness is welfare-improving is dependent on its severity relative to the likelihood of realistic goal evaluation. From an ex ante perspective, the anticipation of future impulsiveness further decreases the value of waiting for a higher net terminal payoff, relative to obtaining a psychological payoff earlier. On the other hand, increasing the likelihood of realistic goal evaluation has the opposite effect, increasing the value of waiting for his material payoff relative to his psychological payoff. Thus, the welfare effect of impulsiveness is dependent on the relative strengths of these two countervailing effects.

In the partial commitment setting, the agent's ability to set a goal that maximizes his current value function distorts his behavior and welfare, because he prefers to set an unrealistically low goal to obtain positive comparison utility. From an ex-ante perspective, the agent anticipates this behavior and thus faces a motivation beyond that of maximizing his net terminal payoff, which is the only ex-ante objective in the case of full commitment. The qualitative results from full commitment ($p = 1$) extend to partial commitment when realistic goal evaluation is sufficiently likely. Though the agent prefers and tends to stop earlier to obtain positive comparison utility, his incentive to increase his material payoff is sufficiently strong that he prefers to improve his patience to do so. Thus, increasing reference dependence to η^{PI} can counteract his impulsiveness, though too much can lead him to wait too long. But when realistic goal evaluation is sufficiently unlikely, the incentive to maximize positive comparison utility overshadows the value of the material payoff. For this reason, the detrimental effects of impulsiveness, which diminishes the material payoff, do not arise. In fact, impulsiveness is welfare-improving, because the agent prefers to obtain positive comparison utility as early as possible.

Thus, a self-control problem due to present-biasedness only arises in optimal stopping when the agent is either not reference-dependent, or when he is reference-dependent *and expects to evaluate himself against sufficiently realistic goals*, because he is primarily motivated to increase his material payoff. In this case, reference dependence can attenuate his impulsive impatience, which diminishes his material payoff. But when the agent expects to evaluate himself against unrealistically low goals, he is primarily motivated to increase his comparison utility, a psychological payoff. Since he prefers to receive such a payoff as soon as possible, impulsive impatience is actually welfare-improving to such an agent.

It should be emphasized that the comparative statics and welfare results for partial commitment when realistic evaluation is sufficiently unlikely are not dependent on the linearity of the comparison utility function. More realistically, the comparison utility function is increasing and concave when his outcome exceeds his goal, reflecting diminishing sensitivity to gains (Kahneman and Tversky, 1979). Even this case, he will still choose the lowest goal possible to maximize his current value function. Intuitively, w is dependent on the expected

discounted values of his consumption and comparison utilities, and the two forces affecting each are the utility incurred upon stopping and the time value of waiting for it. As the agent's goal decreases, the decrease in consumption utility reduces the value function, but time discounting counteracts this reduction, since achieving a lower terminal payoff does not require waiting as long. On the other hand, as the goal decreases, both the increase in comparison utility and time discounting positively affect the value function, since he realizes higher comparison utility and stops earlier due to the weaker force on self-discipline. Since the agent is selecting r_w to satisfy his present-biasedness, he has no incentive to increase his consumption utility through a higher goal, and the latter effect dominates even if his comparison utility function is concave.

However, the partial commitment results that the stopping threshold can decrease with reference dependence and that impulsiveness can be beneficial when the likelihood of realistic evaluation is sufficiently low do depend on fact that the lower bound of r_w lies below zero. Due to the presence of comparison utility, the agent may optimally obtain a negative terminal payoff (\bar{x}^{PI}) when p is sufficiently low, because obtaining positive comparison utility outweighs this loss. Thus, specifying the domain r_w as $r_w \in [-I, \infty)$ because the lower bound of the payoff process is $x = 0$ is well-founded. More generally, these results hold even if $r_w > -I$, specifically whenever $r_w < -\frac{pI}{(1-p)(\bar{\gamma}-1)}$.

6 Conclusion

This paper addresses the role of self-set, non-binding goals as a source of internal motivation in an optimal stopping problem. Agents have linear reference-dependent preferences and endogenously set a goal regarding the expected outcome that serves as the reference point. When realistic goal evaluation is sufficiently likely, they can attenuate the self-control problem, and can even achieve the first-best from an ex ante perspective. Too much reference dependence, on the other hand, leads an agent to wait longer than the first-best, and is always detrimental in the absence of present-biasedness. Notably, none of the effects of goal-setting require any form of loss aversion or curvature in the comparison utility function, nor do they rely on ex-ante uncertainty over outcomes. These findings suggest that the demand for goods and services aimed at educating individuals about goal-setting and goal commitment may be a rational response to impulsiveness, and that the need for external enforcement may be overestimated. Surprisingly, if the goal is sufficiently unconstrained by expectations, both present-biasedness and any degree of reference-dependence can actually be welfare-improving. Whether impulsiveness is welfare-improving is dependent on its severity relative to the likelihood of realistic goal evaluation.

The model presented here can be readily distinguished from other proposed explanations of goal-setting behavior by observables. That goals are simply reminders of tasks to do is belied by evidence that people report less or more satisfaction with the same outcome when their goals differ (Heath et al., 1999; Medvec et al., 1995; McGraw et al., 2005). While publicly announced goals may also influence behavior through social censure or praise, that individuals respond to goals outside the public domain (Locke and Latham, 2002) indicates that social reputation concerns are not the sole driver. My model predicts that a reference dependent agent may be motivated by a self-set goal even if it is privately known. An alternative is that a goal can act as a self-monitoring tool for an individual with imperfect information about the extent of his self-control problem (Bénabou and Tirole, 2004). The model presented here can be distinguished from a self-reputation model in several ways. The two models predict very different responses to information. The reference-dependent agent is motivated by a goal regarding a future outcome, but is unresponsive to information about past behavior. In contrast, knowledge about past behavior leads the imperfectly informed agent to update his beliefs about his willpower and behave accordingly. Thus, a reference-dependent agent might keep a list of future accomplishments to motivate their achievement, while an imperfectly informed agent would prefer to remind himself of past accomplishments. Likewise, while external constraints on previous behavior have a detrimental effect on the future behavior of an imperfectly informed agent because they weaken his reputation capital, they should have no such influence on the reference-dependent one.

The current model limits attention to a single stopping decision. A natural extension is to consider goal-setting when the agent faces a multi-stage project. In this richer setting, the agent may also be able to “break down” or aggregate goals to improve his welfare. This decision corresponds to Read et al.’s (1999) discussion of “motivated bracketing,” where an agent with a self-control problem chooses to frame a problem narrowly or broadly to accomplish a goal. In Hsiaw (2010a), I show that in the presence of uncertainty over outcomes and loss aversion, the decision to frame narrowly or broadly involves a trade-off between greater motivation and more expected comparative disutility due to outcome variance when goal evaluation is frequent. Similarly, the agent could perform the same exercise when facing multiple projects. There, an important consideration is the timing of the comparative evaluation if a goal pertains to multiple outcomes realized at different times.

While this paper considers situations in which goals are evaluated upon the endogenous stopping time, one can imagine situations in which the relevant goals are time-dependent, i.e. “Timely.” For example, a dieter might set a target weight to achieve by the end of the year, and feel bad about herself on December 31 if she has not reached it. In social comparisons, attaining a favorable outcome before a peer or rival can provide positive utility

as well. In such settings, an agent might explicitly set time-contingent goals or engage in continual evaluation, relative to expectations about himself or to his peer.

The interaction of hyperbolic discounting and reference-dependent preferences has implications for a number of other contexts. In Hsiaw (2010b), I extend the model of expectation-based goals to study peer effects through social comparison, which suggests that group or team settings may be beneficial to managers or educators who are interested in improving productivity. Contracting between a principal and agent may be underestimating the effect of external incentive schemes if they are reinforced by goal-setting. Further pursuit of these or related lines of inquiry would enrich our understanding of such effects and their interactions with standard mechanisms.

References

- ABELER, J., A. FALK, L. GOETTE, AND D. HUFFMAN (2011): “Reference Points and Effort Provision,” *American Economic Review*, 101, 470–492.
- AINSLIE, G. (1992): *Picoeconomics: The Interaction of Successive Motivational States within the Person*, Cambridge University Press.
- ARIELY, D. AND K. WERTENBROCH (2002): “Procrastination, Deadlines, and Performance: Self-Control by Precommitment,” *Psychological Science*, 13, 219–24.
- BANDURA, A. (1989): “Self-Regulation of Motivation and Action Through Internal Standards and Goal Systems,” in *Goal Concepts in Personality and Social Psychology*, ed. by L. A. Pervin, Lawrence Erlbaum Associates, Publishers, 19–63.
- BARBERIS, N. C. AND W. XIONG (2012): “Realization Utility,” *Journal of Financial Economics*, 104, 251–271.
- BÉNABOU, R. AND J. TIROLE (2004): “Willpower and Personal Rules,” *Journal of Political Economy*, 112, 848–886.
- BISIN, A. AND K. HYNDMAN (2009): “Procrastination, Self-Imposed Deadlines and Other Commitment Devices,” Working paper.
- BRENNAN, M. J. AND E. S. SCHWARTZ (1985): “Evaluating Natural Resource Investments,” *The Journal of Business*, 58, 135–157.
- BROCAS, I. AND J. D. CARRILLO (2005): “A theory of haste,” *Journal of Economic Behavior & Organization*, 56, 1–23.
- BRUNNERMEIER, M. K. AND J. A. PARKER (2005): “Optimal Expectations,” *The American Economic Review*, 95, 1092–1118.
- CAMERER, C., L. BABCOCK, G. LOEWENSTEIN, AND R. THALER (1997): “Labor Supply of New York City Cabdrivers: One Day at a Time,” *The Quarterly Journal of Economics*, 112, 407–441.
- CAPLIN, A. AND J. LEAHY (2001): “Psychological Expected Utility Theory and Anticipatory Feelings,” *Quarterly Journal of Economics*, 116, 55–79.
- CARD, D. AND G. B. DAHL (2011): “Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior,” *Quarterly Journal of Economics*, 126, 103–143.

- CARRILLO, J. D. (2005): “To be consumed with moderation,” *European Economic Review*, 49, 99–111.
- CARRILLO, J. D. AND M. DEWATRIPONT (2008): “Promises, Promises, . . .,” *The Economic Journal*, 118, 1453–1473.
- CRAWFORD, V. P. AND J. MENG (2008): “New York City Cabdrivers’ Labor Supply Revisited: Reference-Dependence Preferences with Rational-Expectations Targets for Hours and Income,” UCSD Discussion Paper 2008-03, University of California, San Diego.
- DIXIT, A. (1993): *Art of Smooth Pasting, The (Fundamentals of Pure and Applied Economics)*, Routledge.
- DIXIT, A. K. AND R. S. PINDYCK (1994): *Investment under Uncertainty*, Princeton University Press.
- DRUCKER, P. F. (1954): *The Practice of Management*, Harper & Row.
- ELLIS, K. (1998): *The Magic Lamp: Goal Setting for People Who Hate Setting Goals*, Three Rivers Press.
- ERICSON, K. M. AND A. FUSTER (2010): “Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments,” *SSRN eLibrary*.
- FARBER, H. S. (2005): “Is Tomorrow Another Day? The Labor Supply of New York City Cabdrivers,” *Journal of Political Economy*, 113, 46–82.
- (2008): “Reference-Dependent Preferences and Labor Supply: The Case of New York City Taxi Drivers,” *American Economic Review*, 98, 1069–1082.
- FEHR, E. AND L. GOETTE (2007): “Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment,” *American Economic Review*, 97, 298–317.
- FUDENBERG, D. AND D. K. LEVINE (2006): “A Dual-Self Model of Impulse Control,” *American Economic Review*, 96, 1449–1476.
- GOETTE, L. F. AND D. HUFFMAN (2005): “Affect as a Source of Motivation in the Workplace: A New Model of Labor Supply, and New Field Evidence on Income Targeting and the Goal Gradient,” IZA Discussion Paper Series 1890, IZA.
- GOLDSMITH, M. (2008): “Are You Too Obsessed with Your Goals?” Harvard Business Online, [online] http://discussionleader.hbsp.com/goldsmith/2008/03/goal_obsession.html.

- GRENADIER, S. AND N. WANG (2007): “Investment Under Uncertainty and Time-Inconsistent Preferences,” *Journal of Financial Economics*, 84, 2–39.
- GUL, F. AND W. PESENDORFER (2001): “Temptation and Self-Control,” *Econometrica*, 69, 1403–1435.
- HARRIS, C. AND D. LAIBSON (2004): “Instantaneous Gratification,” Working paper, Harvard University.
- HEATH, C., R. P. LARRICK, AND G. WU (1999): “Goals as Reference Points,” *Cognitive Psychology*, 38, 79–109.
- HSIAW, A. (2010a): “Goal Bracketing and Self-Control,” Working paper.
- (2010b): “Goal-Setting, Social Comparison, and Self-Control,” Working paper.
- KADAN, O. AND J. M. SWINKELS (2008): “Stocks or Options? Moral Hazard, Firm Viability, and the Design of Compensation Contracts,” *Review of Financial Studies*, 21, 451–482.
- KAHNEMAN, D. AND A. TVERSKY (1979): “Prospect Theory: An Analysis of Decision under Risk,” *Econometrica*, 47, 263–291.
- KŐSZEKI, B. (2006): “Ego Utility, Overconfidence, and Task Choice,” *Journal of the European Economic Association*, 4, 673–707.
- KŐSZEKI, B. AND M. RABIN (2006): “A Model of Reference-Dependent Preferences,” *Quarterly Journal of Economics*, 121, 1133–1165.
- (2009): “Reference-Dependent Consumption Plans,” *American Economic Review*, 99, 909–936.
- KLEIN, H. J. (1991): “Further evidence on the relationship between goal setting and expectancy theories,” *Organizational Behavior and Human Decision Processes*, 49, 230–257.
- KLEIN, H. J., M. J. WESSON, J. R. HOLLENBECK, AND B. J. ALGE (1999): “Goal Commitment and the Goal-Setting Process: Conceptual Clarification and Empirical Synthesis,” *Journal of Applied Psychology*, 84, 885–896.
- KOCH, A. K. AND J. NAFZIGER (2011): “Self-Regulation through Goal Setting,” *Scandinavian Journal of Economics*, 113, 212–227.

- LAIBSON, D. (1997): “Golden Eggs and Hyperbolic Discounting,” *Quarterly Journal of Economics*, 112, 443–477.
- LATHAM, G. P. AND E. A. LOCKE (1991): “Self-Regulation through Goal Setting,” *Organizational Behavior and Human Decision Processes*, 50, 212–247.
- LOCKE, E. A. AND G. P. LATHAM (2002): “Building a practically useful theory of goal setting and task motivation: A 35-year odyssey.” *American Psychologist*, 57, 705–717.
- LOEWENSTEIN, G. (2007): “Because It Is There: The Challenge of Mountaineering . . . for Utility Theory,” *Kyklos*, 52, 315–343.
- MATTHEY, A. (2008): “Yesterday’s Expectation of Tomorrow Determines What You Do Today: The Role of Reference-Dependent Utility from Expectations,” Jena Economic Research Papers 2008-003.
- MATTHEY, A. AND N. DWENGER (2007): “Don’t Aim Too High: The Potential Costs of High Aspirations,” Jena Economic Research Papers 2007-097.
- MCCLELLAND, D. C. (1962): “Business Drive and National Achievement,” *Harvard Business Review*, 40, 99–112.
- MCDONALD, R. AND D. SIEGEL (1986): “The Value of Waiting to Invest,” *The Quarterly Journal of Economics*, 101, 707–728.
- MCGRAW, A. P., B. A. MELLERS, AND P. E. TETLOCK (2005): “Expectations and emotions of Olympic athletes,” *Journal of Experimental and Social Psychology*, 41, 438–446.
- MEDVEC, V. H., S. F. MADEY, AND T. GILOVICH (1995): “When less is more: Counterfactual thinking and satisfaction among Olympic medalists,” *Journal of Personality and Social Psychology*, 69, 603–610.
- MENTO, A. J., E. A. LOCKE, AND H. J. KLEIN (1992): “Relationship of Goal Level to Valence and Instrumentality,” *Journal of Applied Psychology*, 77, 395–405.
- MIAO, J. (2008): “Option exercise with temptation,” *Economic Theory*, 34, 473–501.
- ODEAN, T. (1998): “Are Investors Reluctant to Realize Their Losses?” *The Journal of Finance*, 53, 1775–1798.
- OU-YANG, H. (2003): “Optimal Contracts in a Continuous-Time Delegated Portfolio Management Problem,” *Review of Financial Studies*, 16, 173–208.

- POPE, D. AND M. SCHWEITZER (2011): “Is Tiger Woods Loss Averse? Persistence Bias in the Face of Experience, Competition, and High Stakes,” *American Economic Review*, 101, 129–157.
- READ, D., G. LOEWENSTEIN, AND M. RABIN (1999): “Choice Bracketing,” *Journal of Risk and Uncertainty*, 19, 171–197.
- SHELLING, T. C. (1984): “Self-Command in Practice, in Policy, and in a Theory of Rational Choice,” *American Economic Review*, 74, 1–11.
- SHAFRAN, R., Z. COOPER, AND C. G. FAIRBURN (2002): “Clinical perfectionism: a cognitive-behavioural analysis,” *Behaviour Research and Therapy*, 40, 773 – 791.
- SHU, S. B. (2008): “Future-biased search: the quest for the ideal,” *Journal of Behavioral Decision Making*, 21, 352–377.
- SPRENGER, C. (2010): “An Endowment Effect for Risk: Experimental Tests of Stochastic Reference Points,” Working paper.
- STROTZ, R. (1956): “Myopia and Inconsistency in Dynamic Utility Maximization,” *The Review of Economic Studies*, 23, 165–180.
- SUVOROV, A. AND J. VAN DE VEN (2008): “Goal Setting as a Self-Regulation Mechanism,” Working Papers w0122, Center for Economic and Financial Research (CEFIR).
- THALER, R. H. (1999): “Mental accounting matters,” *Journal of Behavioral Decision Making*, 12, 183–206.
- TRACY, B. (2003): *Goals!: How to Get Everything You Want - Faster Than You Ever Thought Possible*, Berrett-Koehler Publishers, Inc.

A Appendix

A.1 Derivation of value functions v and w

To construct the continuation value function $v(x, \hat{r})$, I first suppose that all future selves inherit goal \hat{r} and employ a threshold \hat{x} such that they wait if $x < \hat{x}$ and stop if $x \geq \hat{x}$. By continuity of the geometric Brownian motion, there is zero probability that the project value x_t can jump discontinuously from the “wait” region ($x < \hat{x}$) to the “stop” region ($x \geq \hat{x}$)

from one moment to the next. Therefore, I can construct $v(x, \hat{r})$ by considering its behavior in each region separately, then joining them using appropriate boundary conditions.

Using the threshold strategy implies that the value of Equation (6) in the stop region ($x \geq \hat{x}$) is simply given by $x - I + \eta(x - I - \hat{r})$. In the wait region, standard results imply that $v(x, \hat{r})$ must obey the following linear differential equation:

$$\rho v(x, \hat{r}) = \mu x \left(\frac{\partial v}{\partial x} \right) + \frac{1}{2} \sigma^2 x^2 \left(\frac{\partial^2 v}{\partial x^2} \right) \quad \text{if } x < \hat{x}. \quad (15)$$

By definition of the geometric Brownian x , $x = 0$ is an absorbing barrier. The continuation value function must also be continuous at the threshold \hat{x} between the waiting and stopping regions. Therefore, the relevant boundary conditions for v are:¹⁶

$$\text{Boundary: } v(0, \hat{r}) = 0, \quad (16)$$

$$\text{Value Matching: } v(\hat{x}, \hat{r}) = \hat{x} - I + \eta(\hat{x} - I - \hat{r}). \quad (17)$$

Because the stopping decision is only made by *current* selves, and never by future selves, no optimality condition applies to $v(x, \hat{r})$ if the agent is present-biased.

Combining Equation (15) with conditions (16) and (17), the continuation value function is

$$v(x, \hat{r}) = \begin{cases} [(\hat{x} - I) + \eta(\hat{x} - I - \hat{r})] \left(\frac{x}{\hat{x}}\right)^{\gamma_1} & \text{if } x < \hat{x} \\ x - I + \eta(x - I - \hat{r}) & \text{if } x \geq \hat{x}, \end{cases} \quad (18)$$

where $\gamma_1 > 1$ is the positive root¹⁷ of the quadratic equation

$$\frac{1}{2} \sigma^2 \gamma_1^2 + \left(\mu - \frac{1}{2} \sigma^2\right) \gamma_1 - \rho = 0, \quad (19)$$

reflecting the fact that from an ex-ante perspective, the agent discounts the future exponentially at the rate ρ .

Benchmark: The Standard Case

In the standard optimal-stopping problem, the agent is time-consistent ($\beta = 1$ or $\lambda = 0$) and is not reference dependent ($\eta = 0$). His optimal strategy is to use the fixed stopping threshold x^* : at any time s , he waits if $x_s < x^*$ and stops if $x_s \geq x^*$. Time consistency implies that the agent's preferences are fully described by v , so he chooses x^* to maximize v . Thus, the optimality, or smooth pasting, condition applies to the continuation value function v , so the marginal values of waiting and stopping must be equal at the optimal threshold.

¹⁶The first boundary condition is obtained by noting that given any $\hat{r} \geq 0$, which can be verified in equilibrium, the agent never stops if $x = 0$, since he incurs negative overall utility.

¹⁷The negative root is ruled out by the boundary condition for $x = 0$.

Since $\eta = 0$ here, the smooth pasting condition is given by:

$$\text{Smooth Pasting: } \frac{dv}{dx}(x^*) = 1. \quad (20)$$

Solving Equation (15) subject to (20) and barrier absorption and value matching conditions (16) and (17) when $\eta = 0$ allows us to determine the optimal threshold x^* :

$$x^* = \left(\frac{\gamma_1}{\gamma_1 - 1}\right)I, \quad (21)$$

where $\gamma_1 > 1$ is again described by Equation (9) because the continuation value is derived by exponentially discounting the future at the rate ρ in both cases. The standard agent always waits for a project value that exceeds its direct cost ($x^* > I$). Due to the forgone possibility of higher realizations of x in the future, there exists an opportunity cost of stopping today. In equilibrium, his value of the option to stop, denoted $v^*(x)$, is given by

$$v^*(x) = \begin{cases} (x^* - I)\left(\frac{x}{x^*}\right)^{\gamma_1} & \text{if } x < x^* \\ x - I & \text{if } x \geq x^*. \end{cases} \quad (22)$$

In contrast to the time-consistent case, the present-biased agent maximizes the current value function w , rather than the continuation value v . Proceeding with the derivation of $w(x, \bar{r})$ analogously, I first suppose that all current selves inherit goal \bar{r} and employ a threshold \bar{x} such that they wait if $x < \bar{x}$ and stop if $x \geq \bar{x}$. Again, I can construct $w(x, \bar{r})$ by characterizing it in the “wait” and “stop” regions separately, then joining the two regions through appropriate boundary conditions.

Using the threshold strategy implies that the value of w in the stop region ($x \geq \bar{x}$) is simply given by $x - I + \eta(x - I - \bar{r})$. In the wait region, standard results imply that $w(x, \bar{r})$ must obey the following linear differential equation:

$$\rho w(x, \bar{r}) = \lambda(\beta v(x, \hat{r}) - w(x, \bar{r})) + \mu x \left(\frac{\partial w}{\partial x}\right) + \frac{1}{2} \sigma^2 x^2 \left(\frac{\partial^2 w}{\partial x^2}\right) \quad \text{if } x < \bar{x}. \quad (23)$$

Comparing Equation (23) to Equation (15), the additional term $\lambda(\beta v(x, \hat{r}) - w(x, \bar{r}))$ is the expected value of the change in the current value w that occurs through the stochastic arrival of a transition from the present to the future.

As with v , $x = 0$ is an absorbing barrier and w must clearly be continuous at the threshold \bar{x} between the wait and stop regions. Finally, the smooth pasting condition must apply to w with respect to x , because the optimal threshold is chosen by the current self to maximize his current value function w . At \bar{x} , the marginal value of waiting must equal that of stopping

so that the current self is unwilling to deviate from stopping. Thus, the relevant boundary conditions for w are

$$\text{Boundary: } w(0, \bar{r}) = 0, \quad (24)$$

$$\text{Value Matching: } w(\bar{x}, \bar{r}) = \bar{x} - I + \eta(\bar{x} - I - \bar{r}), \quad (25)$$

$$\text{Smooth Pasting: } \frac{\partial w}{\partial x}(\bar{x}, \bar{r}) = 1 + \eta. \quad (26)$$

Because the current self fully anticipates that his future selves will use threshold \hat{x} given goal \hat{r} , we can substitute the continuation value function, given by Equation (18) into the differential Equation (23). Under the assumption that $x \leq \hat{x}$, which is satisfied in a stationary equilibrium, it is the value of v in its wait region that applies to (23). Combining Equation (23) with (18) and conditions (24), (25), and (26) yields the solution to the optimal threshold \bar{x} as a function of current goal \bar{r} and the conjectured future goals \hat{r} and thresholds \hat{x} .

A.2 Proof of Proposition 3

Each self prefers that future selves choose their thresholds in order to maximize the continuation value v rather than the current value w . Therefore, we look for the threshold \bar{x} that maximizes the wait region of v^{SI} . The first order condition $\frac{\partial v^{SI}}{\partial \bar{x}} = 0$ gives us the first-best threshold $\bar{x} = (\frac{\gamma_1}{\gamma_1 - 1})I$, which is identical to the solution for a standard agent x^* . We can also see this feature by inspecting the wait region of v^{SI} , which is the same function of the threshold as v^* . Because utility incurred upon stopping is not directly distorted, the ex ante option value of waiting, given a stopping threshold, is not distorted by η directly.

However, the presence of $\eta > 0$ affects the equilibrium threshold by changing the marginal value of stopping. The current self's optimal threshold is $\bar{x}^{SI} = (\frac{\bar{\gamma}}{\bar{\gamma} - 1 - \eta})I$, so the first-best can only be achieved for η^* such that $\frac{\bar{\gamma}}{\bar{\gamma} - 1 - \eta^*} = \frac{\gamma_1}{\gamma_1 - 1}$. We can verify that $\eta^* < \bar{\gamma} - 1$ by noting that $\frac{\bar{\gamma} - \gamma_1}{\gamma_1} = \frac{\bar{\gamma}}{\gamma_1} - 1$. Since $\gamma_1 > 1$, then this inequality is satisfied for any $\beta \in [0, 1]$.

A.3 Proof of Corollary 1

The first part follows from the above proof of Proposition 3. The second statement follows by comparing the value functions when $\eta > 0$ versus $\eta = 0$. When $\eta = 0$, the agent stops at threshold $(\frac{\bar{\gamma}}{\bar{\gamma} - 1})I$. When $\eta > 0$, the agent stops at threshold $(\frac{\bar{\gamma}}{\bar{\gamma} - 1 - \eta})I$. Thus, goal-setting is

detrimental when

$$\begin{aligned} [(\frac{\bar{\gamma}}{\bar{\gamma}-1-\eta})I - I](\frac{1}{(\frac{\bar{\gamma}}{\bar{\gamma}-1-\eta})I})^{\gamma_1} &< [(\frac{\bar{\gamma}}{\bar{\gamma}-1})I - I](\frac{1}{(\frac{\bar{\gamma}}{\bar{\gamma}-1})I})^{\gamma_1} \\ [(\frac{1+\eta}{\bar{\gamma}-1-\eta})I](\frac{1}{(\frac{\bar{\gamma}}{\bar{\gamma}-1-\eta})I})^{\gamma_1} &< [(\frac{1}{\bar{\gamma}-1})I](\frac{1}{(\frac{\bar{\gamma}}{\bar{\gamma}-1})I})^{\gamma_1} \\ 0 &< (\frac{\bar{\gamma}-1}{\bar{\gamma}-1-\bar{\eta}})^{\gamma_1-1}(\frac{1}{1+\bar{\eta}}) - 1, \end{aligned}$$

and the lower bound $\bar{\eta}$ satisfies the above condition with equality. Define the function $H(\eta)$ such that

$$H(\eta) = (\frac{\bar{\gamma}-1}{\bar{\gamma}-1-\bar{\eta}})^{\gamma_1-1}(\frac{1}{1+\bar{\eta}}) - 1.$$

We can verify that $\bar{\eta}$ exists and is unique by noting that $H(0) = 0$, $H(\bar{\gamma}-1) \rightarrow \infty$, and $H(\eta)$ strictly decreases for $\eta < \eta^*$ and strictly increases for $\eta > \eta^*$. Thus, we have shown existence and uniqueness of $\bar{\eta}$ for any $\gamma \in (1, \infty)$, as well as the fact that $\eta^* < \bar{\eta} < \bar{\gamma}-1$. Finally, the implicit function theorem yields

$$\frac{\partial \bar{\eta}}{\partial \bar{\gamma}} = -\frac{-(\frac{\bar{\eta}}{1+\bar{\eta}})(\frac{\gamma_1-1}{(\bar{\gamma}-1-\bar{\eta})^2})(\frac{\bar{\gamma}-1}{(\bar{\gamma}-1-\bar{\eta})^{\gamma_1-2}})}{(\frac{1}{1+\bar{\eta}})(\frac{\bar{\gamma}-1}{(\bar{\gamma}-1-\bar{\eta})^{\gamma_1-1}})[- \frac{1}{1+\bar{\eta}} + (\frac{\gamma_1-1}{\bar{\gamma}-1-\bar{\eta}})]} > 0,$$

which is positive since the numerator is always negative and the denominator is positive since $\bar{\eta} > \eta^*$.

A.4 Verification

To verify that the constructed current value function w is optimal for any given $r \geq 0$, note that Equation (7) implies that any solution must satisfy the following two conditions for all $x \in (0, \infty)$ (whether the agent is sophisticated or naive):

$$w(x, r) \geq x - I + \eta(x - I - r) \tag{27}$$

$$0 \geq -\rho w(x, r) + \lambda(\beta v(x, \hat{r}) - w(x, r)) + \mu x (\frac{\partial w}{\partial x}) + \frac{1}{2} \sigma^2 x^2 (\frac{\partial^2 w}{\partial x^2}). \tag{28}$$

Let \bar{x} denote current self's stopping threshold and \hat{x} denote the (perceived) future self's stopping threshold. By construction, $w(x, r) = x - I + \eta(x - I - r)$ when $x \geq \bar{x}$ so equation (27) holds with equality. When $x < \bar{x}$, $w(x, r)$ is of the form $w(x, r) = A_1 x^{\gamma_1} + A_2 x^{\gamma_2}$, where $\gamma_2 \geq \gamma_1 > 1$, $A_1 > 0$, and $A_2 > 0$. Since $w(x, r)$ is convex and increasing, it must lie above

the line $x - I + \eta(x - I - r)$ for all $x < \bar{x}$.

Equation (28) holds with equality when $x < \bar{x}$ by construction. Define the function $J(x)$ as follows:

$$J(x) = -\rho w(x, r) + \lambda(\beta v(x, \hat{r}) - w(x, r)) + \mu x \left(\frac{\partial w}{\partial x} \right) + \frac{1}{2} \sigma^2 x^2 \left(\frac{\partial^2 w}{\partial x^2} \right)$$

When $x \geq \bar{x}$, we have $w(x, r) = x - I + \eta(x - I - r)$. Since we have shown that $\bar{x} \leq \hat{x}$ (with equality only if the individual is sophisticated), then $v(x, r) = x - I + \eta(x - I - r)$ if $x \geq \bar{x}$. Then we have

$$\begin{aligned} J(x) &= -\rho[x - I + \eta(x - I - r)] + \lambda(\beta[x - I + \eta(x - I - r)] - [x - I + \eta(x - I - r)]) + \mu x(1 + \eta) \\ &= (1 + \eta)[\mu - \rho - \lambda(1 - \beta)]x + [\rho + \lambda(1 - \beta)][(1 + \eta)I + \eta r], \end{aligned}$$

which is strictly decreasing in x since $\mu < \rho$. We have previously shown that $\frac{\partial \bar{x}}{\partial \hat{x}} > 0$, and recall that $\bar{x} = \hat{x}^{SE}$ when $\hat{x} = \hat{x}^{SE}$. So it is sufficient to show that $J(\hat{x}^{SE}) < 0$ to satisfy Equation (28).

$$\begin{aligned} J(x) &\leq J(\hat{x}^{SE}) \\ &= [(1 + \eta)I + r\eta] \left(\frac{1}{\bar{\gamma} - 1} \right) [\bar{\gamma}\mu - \rho - \lambda(1 - \beta)] \\ &= [(1 + \eta)I + r\eta] \left(\frac{1}{\bar{\gamma} - 1} \right) [\beta(\mu\gamma_1 - \rho) + (1 - \beta)(\mu\gamma_2 - \rho - \lambda)]. \end{aligned}$$

Recall that $\gamma_1 > 1$ satisfies $0 = -\rho + \mu\gamma_1 + \frac{1}{2}\sigma^2\gamma_1(\gamma_1 - 1)$. Then $\mu\gamma_1 - \rho = -\frac{1}{2}\sigma^2\gamma_1(\gamma_1 - 1) < 0$. Likewise, $\gamma_2 > 1$ satisfies $0 = -(\rho + \lambda) + \mu\gamma_2 + \frac{1}{2}\sigma^2\gamma_2(\gamma_2 - 1)$, so $\mu\gamma_2 - \rho - \lambda = -\frac{1}{2}\sigma^2\gamma_2(\gamma_2 - 1) < 0$. Thus, $J(x) \leq J(\hat{x}^{SE}) < 0$. Therefore, the constructed value function w is at least as good as the value function generated by any alternative Markov strategy. This result also applies to the case of partial commitment, which corresponds to $r = pr_{RE} + (1 - p)r_w$. In Hsiaw (2010b), I also verify that this result applies to the case of naivete.

A.5 Deriving r_w

The current value function is given by $w(x, r_{RE}) = \beta A_1 x^{\gamma_1} + A_2 x^{\gamma_2}$, where

$$\begin{aligned} A_1 &= \left(\frac{1}{\hat{x}} \right)^{\gamma_1} [\hat{x} - I - \eta(\hat{x} - I - p\hat{r}_{RE} - (1 - p)\hat{r}_w)] \\ A_2 &= \left(\frac{1}{\bar{x}} \right)^{\gamma_2} [\bar{x} - I - \eta(\bar{x} - I - pr_{RE} - (1 - p)r_w)] - \beta A_1 (\bar{x})^{\gamma_1 - \gamma_2}, \end{aligned}$$

\hat{r}_{RE} denotes the goal that a future self inherits, and \hat{r}_w denotes the goal that a future self sets to maximize his current value function. The agent's optimization problem is to choose r_w such that

$$\max_{r_w} \beta A_1 x^{\gamma_1} + A_2 x^{\gamma_2}$$

Since the goals of future selves are unaffected by r_w and current action \bar{x} , the first term of $w(x, r_{RE})$, which describes the value of behavior after the current self's lifetime ends, is irrelevant to the choice of r_w . The first-order condition is

$$\begin{aligned} \frac{\partial w}{\partial r_w} = & x^{\gamma_2} \left[-\gamma_2 \left(\frac{1}{x}\right)^{\gamma_2+1} [\bar{x} - I - \eta(\bar{x} - I - pr_{RE} - (1-p)r_w)] \left(\frac{\partial \bar{x}}{\partial r_w}\right) \right. \\ & \left. + \left(\frac{1}{x}\right)^{\gamma_2} [(1+\eta) \left(\frac{\partial \bar{x}}{\partial r_w}\right) - \eta(1-p)] - \beta A_1 (\gamma_1 - \gamma_2) (\bar{x})^{\gamma_1 - \gamma_2 - 1} \left(\frac{\partial \bar{x}}{\partial r_w}\right) \right]. \end{aligned}$$

In a stationary solution, $\bar{x} = \hat{x}$, $r_{RE} = \hat{r}_{RE}$, and $r_w = \hat{r}_w$. Imposing these conditions yields

$$\bar{x} = \left(\frac{\bar{\gamma}}{\bar{\gamma} - 1}\right)I + (pr_{RE} + (1-p)r_w) \left(\frac{\bar{\gamma}}{\bar{\gamma} - 1}\right) \left(\frac{\eta}{1+\eta}\right),$$

so the first-order condition reduces to

$$\frac{\partial w}{\partial r_w} = \left(\frac{x^{\gamma_2}}{\bar{x}^{\gamma_2+1}}\right) (1-p) \left(\frac{\eta}{1+\eta}\right) \left(\frac{\bar{\gamma}}{\bar{\gamma} - 1}\right) [\bar{x} - I - \eta(\bar{x} - I - pr_{RE} - (1-p)r_w)] [-(\gamma_2 - 1) + \beta(\gamma_2 - \gamma_1)].$$

Since $\beta(\gamma_2 - \gamma_1)$ is monotonically increasing in β and $-(\gamma_2 - 1) + (\gamma_2 - \gamma_1) < 0$, then $\frac{\partial w}{\partial r_w} < 0$. Thus, the optimal choice of r_w is the lowest possible: $r_w = -I$.

A.6 Proof of Proposition 5

Just as in the full commitment case ($p = 1$), each self prefers that future selves choose their thresholds in order to maximize the continuation value v rather than the current value w . Therefore, we look for the threshold \bar{x} that maximizes the wait region of v , which now includes both the net terminal payoff and comparison utility, since the agent anticipates the possibility that he exceeds the unrealistically low goal r_w . When $p \leq 1$, the continuation value function is given by

$$\begin{aligned} v(x, r_{RE} = \bar{x} - I) &= \left(\frac{1}{\bar{\gamma}}\right)^{\gamma_1} \{\bar{\gamma} - I + \eta[\bar{\gamma} - I - p(r_{RE} - I) - (1-p)(r_w - I)]\} \\ &= \left(\frac{1}{\bar{\gamma}}\right)^{\gamma_1} \{[1 + \eta(1-p)]\bar{x} - I\} \end{aligned}$$

The first order condition is:

$$\begin{aligned}\frac{\partial v}{\partial \eta} &= \left(\frac{1}{\bar{\gamma}}\right)^{\gamma_1} \left\{ (1-p)\bar{x} + [1 + \eta(1-p)]\left(\frac{\partial \bar{x}}{\partial \eta}\right) \right\} - \gamma_1 \left(\frac{\partial \bar{x}}{\partial \eta}\right) \left(\frac{1}{\bar{\gamma}}\right)^{\gamma_1+1} \{ [1 + \eta(1-p)]\bar{x} - I \} \\ &= \left(\frac{1}{\bar{\gamma}}\right)^{\gamma_1} \left(\frac{1}{\bar{\gamma} - 1 - \eta + (1-p)\eta\bar{\gamma}} \right)^2 (I) \{ \bar{\gamma} - \gamma_1 + (1-p)\bar{\gamma}(\gamma_1 - 1) + \eta\gamma_1[(1-p)\bar{\gamma} - 1] \}.\end{aligned}$$

If $p < \frac{\bar{\gamma}-1}{\bar{\gamma}}$, then first-order condition is positive for any η , implying that $\eta^{PI} = \infty$. If $p > \frac{\bar{\gamma}-1}{\bar{\gamma}}$, then there exists an interior solution for η^{PI} , given by

$$\eta^{PI} = \frac{\bar{\gamma} - \gamma_1 + (1-p)\bar{\gamma}(\gamma_1 - 1)}{\gamma_1(1 - \bar{\gamma}(1-p))}.$$

We can easily verify that $\bar{x}(\eta^{PI}) = \left(\frac{\gamma_1}{p(\gamma_1-1)}\right)I$ when $p > \frac{\bar{\gamma}-1}{\bar{\gamma}}$, and $\bar{x}(\eta^{PI}) \rightarrow 0$ when $p < \frac{\bar{\gamma}-1}{\bar{\gamma}}$.