Optimal Time-inconsistent Beliefs: Misplanning, Procrastination, and Commitment*

Markus K. Brunnermeier[†]
Princeton University
CEPR and NBER

Filippos Papakonstantinou[‡]
Imperial College London

Jonathan A. Parker[§]
MIT Sloan School of Management and NBER

Abstract

We develop a structural theory of beliefs and behavior, that relaxes the assumption of time-consistency in *beliefs*. Our theory is based on the trade-off between optimism, which raises anticipatory utility, and objectivity, which promotes efficient actions. We present it in the context of allocating work on a project over time, develop testable implications to contrast it with models assuming time-inconsistent preferences, and compare its predictions to existing evidence on behavior and beliefs. Our predictions are: (i) optimal beliefs are optimistic and time-inconsistent; (ii) people optimally exhibit the planning fallacy; (iii) incentives for rapid task completion make beliefs more optimistic and worsen work smoothing, while incentives for accurate duration prediction make beliefs less optimistic and improve work smoothing; (iv) without a commitment device, beliefs become less optimistic over time; (v) in the presence of a commitment device, beliefs may become more optimistic over time, and people optimally exhibit preference for commitment.

Keywords: Optimal Beliefs, Optimism, Time-Inconsistent Beliefs, Time-Inconsistent Preferences, Preference for Commitment

JEL Classification: D10, D80, E21

*For helpful comments, we thank Roland Benabou, Alex Michaelides, Wolfgang Pesendorfer, and seminar participants at several universities, although we remain responsible for all errors, omissions, and typos.

[†]Department of Economics, Bendheim Center for Finance, Princeton University, Princeton, NJ 08544-1021, e-mail: markus@princeton.edu, http://scholar.princeton.edu/markus

[‡]Imperial College London, South Kensington Campus, London SW7 2AZ, UK, e-mail: fpapakon@imperial.ac.uk

[§]MIT Sloan School of Management, 100 Main Street, E62, Cambridge, MA 02142-1347, e-mail: japarker@mit.edu

People tend to postpone work previously planned for the present, revising earlier plans despite a complete lack of new relevant information and sometimes even despite costs of doing so. Likewise, people sometimes commit to deadlines, choosing to constrain their future options even despite potential costs from not being able to respond to new information. Akerlof (1991) argues that "a major function of management is to ... monitor accomplishment so as to prevent procrastination," in both project initiation and project termination. And a large number of financial advice books and web sites deal with how individual investors can automate decisions like saving to avoid mistakes from procrastination.[1]

Economic theory has made significant progress in explaining this type of behavior as arising due to preferences, either due to time-inconsistency as in models of quasi-hyperbolic discounting (e.g., Laibson, 1997) or due to temptation as in the revealed-preference approach of Gul and Pesendorfer (2001). However, psychological evidence suggests that it is not only the utility or disutility of certain actions that generates these behaviors, but also misperceptions of the probabilities of future events. According to Kahneman and Tversky (1979):

> The context of planning provides many examples in which the distribution of outcomes in past experience is ignored. Scientists and writers, for example, are notoriously prone to underestimate the time required to complete a project, even when they have considerable experience of past failures to live up to planned schedules.

In this paper we develop a structural theory based on optimistic beliefs about the future and show that it naturally leads to these behaviors. Relative to earlier work, we allow beliefs to be updated in a non-Bayesian fashion, and we demonstrate that such a theory can not only match observed patterns in behavior, but also observed patterns in beliefs. We derive a number of predictions of our theory that are supported by existing experimental evidence from the psychology literature. Our theory predicts that optimal beliefs are optimistic and that people optimally exhibit the planning fallacy (consistent with the findings of Roy et al., 2005), that incentives for rapid task completion make beliefs more optimistic (consistent with Byram, 1997) while incentives for accurate task duration prediction make beliefs less optimistic (consistent with Byram, 1997; Buehler et al., 1997), and that people optimally self-impose deadlines that improve work smoothing less than externally-imposed deadlines (consistent with Ariely and Wertenbroch, 2002). Our theory also predicts that monetary incentives that affect beliefs also affect behavior; this predic-

[1] See, for example, Ashraf et al. (2006) and http://www.apa.org/helpcenter/willpower-finances.aspx.

tion can be used to perform additional experimental tests to differentiate our model of optimal time-inconsistent beliefs from existing models of time-inconsistent preferences.

Our theory is based on the trade-off between optimism, which raises anticipatory utility, and objectivity, which promotes efficient actions. An individual's current utility depends on his expectations about the course of the future, so optimistic beliefs make him happier. But such beliefs come at a cost: the chosen actions are optimal given these optimistic beliefs, but they can be objectively suboptimal. Optimal beliefs balance the anticipatory benefit with the efficiency cost, but since these vary from one period to another, optimal beliefs are time-inconsistent (making our theory distinct from Brunnermeier and Parker, 2005). Since people understand that their plans of action are time-inconsistent, they choose to constrain future actions by utilizing commitment devices.

While our theory is general, we present it in the context of planning. By focusing on planning, we can use existing evidence from the psychology literature on the planning fallacy — i.e., people's tendency to underestimate the time necessary to complete tasks and therefore to make suboptimal plans — and on the use of deadlines, to evaluate our theory's implications with observed behavior and, importantly, beliefs. Further, we can contrast our theory with theories that assume time-inconsistent preferences. The predictions of these models differ, both in terms of beliefs and in terms of behavior. Evidence on beliefs is particularly important for our theory, in which beliefs play a central role and belief biases are endogenously determined by preferences and the economic environment.

More specifically, we consider a three-period model of the behavior and beliefs of an agent who plans when and how much to work in order to complete a task of uncertain total difficulty. The agent first has the option to constrain himself by imposing an intermediate deadline, and then chooses how to allocate his work among the remaining two periods, in order to complete the task. Decisions maximize the expected present-discounted value of utility — including utility from anticipation — given subjective beliefs, while optimal subjective beliefs maximize the *well-being* — the "lifetime" discounted utility averaged over the states of the world. Well-being is evaluated using objective beliefs so it captures the trade-off between distorting beliefs and distorting actions, across realizations of uncertainty.

Without a commitment device, two features of the model form the agent's optimal beliefs and behavior, which are characterized by optimism and the planning fallacy. First, the agent has anticipatory utility, so if he believes the task to be easy, he is happier in the present because he anticipates less work in the future. This ingredient provides an anticipatory benefit of optimistic beliefs. Second, the agent chooses optimal actions given his beliefs, so if he has optimistic beliefs, he does little work in the present and ends up poorly smoothing work over time. This ingredient implies a cost of optimism on average. Given these ingredients, people exhibit optimism, because it has first-order anticipatory benefits and, by the envelope theorem, only second-order behavioral costs. Thus, in our theory, optimism and the planning fallacy are endogenous, and the severity of these biases is larger, the greater the anticipatory benefits of optimism and the smaller the costs of mis-planning.

In the presence of a commitment device, if beliefs become more optimistic over time and the individual understands the resulting tendency to postpone work, he may impose a binding deadline on his future self. Given a deadline, the individual can still receive the anticipatory benefits of believing the task will be easy, without suffering as high costs of poorly smoothing work over time. Thus, beliefs that induce a binding deadline may make the agent better off on average, so may be optimal. As above, the individual's beliefs and his choice of deadline are endogenous and situational, giving different predictions across different environments.

The predictions of our model are generally consistent with experimental and survey data on beliefs and task durations reported in the psychology literature on the planning fallacy. (i) Our model predicts that beliefs are optimistic and people suffer from the planning fallacy. There is a wealth of evidence, starting with Kahneman and Tversky (1979) that supports these predictions. (ii) Our model predicts that, without commitment, beliefs become less optimistic over time. This is consistent with the finding, e.g., by Gilovich et al. (1993), that people become less optimistic about their performance as the time of performance approaches. As predicted by our model, Ariely and Wertenbroch (2002) find that (iii) when people are given an opportunity to impose a deadline on themselves prior to beginning a task, most do so; and that (iv) people self-impose deadlines that are less strict and increase performance less than externally-imposed deadlines. Finally, our model predicts that (v) monetary incentives for rapid task completion increase the planning fallacy, and (vi) monetary incentives for accurate prediction of task duration reduce the planning fallacy. These predictions match the evidence, e.g., by Buehler et al. (1997). In addition to presenting the substantial body of evidence that is consistent with our model's results, we derive novel, but testable, predictions. For example, our theory predicts that if people choose to impose a deadline, they should still hold optimistic beliefs; in fact beliefs should become more optimistic over time. We also derive comparative statics with respect to the model's structural parameters.

Our paper is closely related to the literature on commitment. In one branch of the literature, people discount non-exponentially in an otherwise standard time-separable utility maximization problem (following Strotz, 1955);[2] while in the other, people maximize preferences that have an additional component, temptation utility, that makes it unpleasant to make optimal decisions (Gul and Pesendorfer, 2001).[3] In either case, procrastination occurs because the utility function places special importance on the present relative to the future.[4] In some of these models, people are exogenously assumed to partially (as in O'Donoghue and Rabin, 2001) or completely (as in Laibson, 1997) understand their tendency to procrastinate, and therefore choose to constrain their future choices. In contrast, in our theory beliefs, rather than preferences, are the central cause of these interesting behaviors, and importantly, the propensity to commit is not assumed, but rather arises endogenously as optimal behavior due to incentives. Compared to preference-based models, our theory clearly yields different testable predictions in terms of beliefs; for example, it predicts that beliefs about the total work necessary for task completion are optimistic and are affected by incentives for accuracy of prediction, speed of completion, etc. But, interestingly, our theory also yields some different predictions in terms of behavior. First, since beliefs affect optimal actions, incentives (e.g., the aforementioned monetary incentives for the accuracy of prediction of task duration) that affect optimal beliefs also affect optimal actions; this is not the case in models with time-inconsistent preferences, since beliefs are assumed to be rational. Second, our model predicts that the optimal work, the optimal deadline, and the propensity to choose a deadline vary with the structural model parameters in different ways than in a model with present-biased preferences.

The theoretical contribution of this paper is also related to research that endogenizes beliefs, such as Akerlof and Dickens (1982), Yariv (2002), and Bénabou and Tirole (2002). More closely, Bernheim and Thomadsen (2005) and Koszegi (2006) present models in which people choose to

[2] Related to this branch of the literature, Dasgupta and Maskin (2005) explain preference reversals as instinctual behavior due to natural selection. With standard preferences, optimal behavior in environments with uncertainty — which are most commonly faced by animals — becomes instinctual, hence is exhibited even in environments without uncertainty; the authors argue that such behavior would be indistinguishable from behavior arising due to non-exponential discounting.

[3] Epstein (2006) interprets this temptation utility as the expected utility from incorrect beliefs, and shows that if beliefs are a mix of the correct and incorrect beliefs, they may not be Bayesian, hence people may exhibit a preference for commitment. Both the incorrect beliefs and the precise mix of correct and incorrect beliefs that an agent uses, hence non-Bayesian updating and attitudes towards commitment, are exogenously assumed and embedded in the preference primitives. Instead, in our model, non-Bayesian beliefs and a preference for commitment arise endogenously due to incentives.

[4] The two alternatives' predictions for behavior are similar. Also as Thaler and Benartzi (2004) argue, non-exponential discounting models are more widely used in economic applications, while temptation models belong more narrowly to the choice theory literature. As a result, we focus on the former.

process or collect information not only to make better decisions but also to get utility from anticipation or their self-image. In Koszegi and Rabin (2009), agents derive direct utility from changes in beliefs (i.e., news) and make time-inconsistent plans as long as preferences place special importance on present relative to future news. In contrast, our paper does not focus on the mechanism through which belief distortions occur, rather it concentrates on the optimality of beliefs, which turn out to be time-inconsistent. Brunnermeier and Parker (2005) also focus on the optimality of beliefs, but the model and focus of this paper are different. First, we do not constrain subjective beliefs to be time-consistent; this makes our theory more generally applicable, allows us to better explain the evolution of beliefs over time, and gives rise to the optimality of preference for commitment. Second, agents in our model care about future consumption not only because they gain immediate utility from anticipating it, but also because they care about their future selves through the traditional discount factor. Even though this assumption does not affect our qualitative results regarding the evolution of beliefs and preference for commitment, it has a couple of distinct advantages: i) It enables us to explore a direct comparison between our model of time-inconsistent beliefs and its implications versus the model of time-inconsistent preferences (which necessarily incorporates the traditional discount factor) and its implications, which is a central part of our paper; and ii) it enables us to derive comparative statics with respect to the traditional discount factor, which is a standard and important component of individual preferences.

To see more sharply the distinction between the time-inconsistent preferences model, the optimal-beliefs model of Brunnermeier and Parker (2005), and our theory, some readers may find it useful to consider an application of our model in a context different from the planning context that we use to develop our theory. For example, consider the real options framework of investment under uncertainty, in which an entrepreneur chooses the optimal time to make an investment by balancing the cost of preemption by another entrepreneur with the benefit of delaying the investment in the hope that the (lump-sum) payoff will be larger in the future. Grenadier and Wang (2007) show that, with time-inconsistent preferences, the agent would invest earlier than with time-consistent preferences, due to impatience. On the other hand, with optimal time-consistent beliefs (as in Brunnermeier and Parker, 2005), he would be optimistic about the payoff's future realizations, hence he would invest later than with rational beliefs. Finally, in the theory of optimal time-inconsistent beliefs that we propose, the agent would still invest later than with rational beliefs, but the key distinction is that he would self-impose a *binding* deadline that would force him to invest earlier (though not as early as with rational beliefs); furthermore, the agent would become more optimistic over time, since a binding deadline eliminates the cost of dreaming.

The empirical contribution of this paper is related to economic research that develops a scientific foundation for modeling beliefs. In general, observed behavior can be matched by several combinations of preferences and beliefs. The common approach to this identification problem is to assume rational beliefs, perhaps based on philosophical arguments,[5] and to propose models that match evidence on behavior. We instead address this identification problem by using evidence both on behavior and on beliefs; the implications of our model are largely consistent with the patterns of behavior and beliefs found in experiments on the planning fallacy. This finding supports our model, as well as the use of survey data on beliefs to discipline research in behavioral economics. In a (broadly) similar vein, some papers have used survey data on beliefs to test rationality or to evaluate its usefulness in the context of specific models (Manski, 2004).[6]

Finally, since we present our theory in the context of planning, our paper is related to the psychological theories of the planning fallacy, which focus on the mental processes that lead people to make incorrect predictions. These theories are generally consistent with our theory and inconsistent with existing models in the economics literature. Kahneman and Tversky (1979) argue that the fallacy arises because people ignore distributional information available in related past outcomes and instead focus on a single plausible and optimistic scenario for completion of the current task. Trope and Liberman (2003) apply construal level theory to temporal distance to argue that people view temporally distant events in terms of a few abstract characteristics and so, when forming predictions, overlook the potential difficulties and sub-tasks involved in task completion. Further, several papers on the planning fallacy have investigated whether it is caused by incorrect memory of past events, or due to biased self-attribution (Buehler et al., 1994). Lastly, there is the general theory that people are optimistic, and this is helpful in generating motivation, effort, and persistence (e.g., Armor and Taylor, 1998).

The rest of the paper is structured as follows. Section 1 presents our theory in the context of planning. Section 2 presents our results on optimism, overconfidence, and the planning fallacy

[5]Two versions of such arguments bear mention. First, rationality is sometimes viewed as providing evolutionary fitness. But this is only true for specific environments and utility functions; that it is not generally true follows directly from the identification problem. Second, it is sometimes claimed that people learn true probabilities over time; this simply pushes back the assumption of rationality to one of rational learning, so is still an assumption, not a scientific basis for distinguishing models.

[6]Also see Hamermesh (1985), Dominitz (1998), and Nyarko and Schotter (2002).

without a commitment device. Section 3 shows that people may exhibit preference for commitment, caused entirely by optimal time-inconsistent beliefs. Section 4 presents comparative statics results for optimal beliefs, actions, and deadlines with respect to model parameters. In Section 5, we develop a model with present-biased preferences, which allows us then in Section 6 to compare our model's predictions with those from the model with present-biased preferences, and to discuss the relevant empirical evidence. The appendix provides proofs of all propositions, and an online appendix provides more detailed proofs of some propositions.

1 The model

Here, we present our theory in the context of the task-completion problem. First, we present the environment and then we state the objectives. In short, optimal actions maximize subjective expected utility, including utility from anticipation, and optimal beliefs maximize the objective expected lifetime utility. That is, we deviate from the standard economic model in two directions: utility is derived not only from consumption but also from anticipation (Loewenstein, 1987; Caplin and Leahy, 2001), and beliefs are subjective and maximize a welfare function (Landier, 2000; Brunnermeier and Parker, 2005). Notably, we do *not* require that beliefs are Bayesian. But we retain rationality in the sense that agents are fully sophisticated, i.e., they know their preferences and beliefs in the present and in the future, for all states of the world, and correctly take these into account when making optimal decisions.

Our model is related to that in Brunnermeier and Parker (2005), but differs in the following key respects. First, we do not constrain subjective beliefs to be Bayesian; this makes our theory more generally applicable, allows us to explain the endogenous evolution of beliefs over time, and gives rise to the optimality of preference for commitment. Second, in our model, agents care about future consumption not only because they gain immediate utility from anticipating it, but also because they care about their future selves; this makes our model more standard and enables us to directly compare it with a model with time-inconsistent preferences, as well as to calculate comparative statics with respect to the traditional discount factor.

1.1 The task-completion environment

There are three periods: 0, 1, and 2. The total amount of work required to complete the task is $1+\eta$, where $\eta \geq 0$ is a random variable realized in period 2, with mean $E := \mathbb{E}[\eta] > 0$ and variance

$\Sigma := \mathbb{E}[\eta^2] - (\mathbb{E}[\eta])^2 > 0$, where $\mathbb{E}[\cdot]$ denotes the unconditional objective expectation.[7] At $t = 1$, the agent chooses work $w_1 \leq 1$,[8] and at $t = 2$ he completes the task. So the resource constraint is

$$w_1 + w_2 = 1 + \eta. \tag{1}$$

At $t = 0$, the agent does not work but may have the option to impose an interim deadline in the form of a lower bound ψ on the amount of work w_1 completed in period 1. That is, he may impose

$$w_1 \geq \psi.$$

The following time-line summarizes our setup:

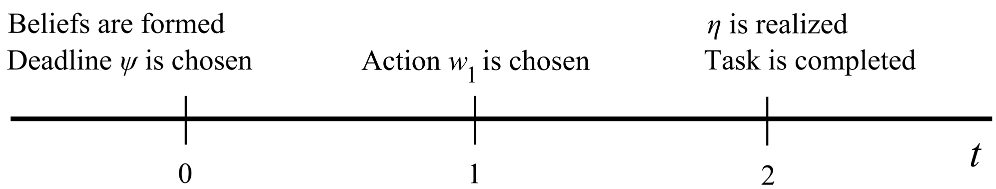


Figure 1: Time-line of belief formation, action choices, and information arrival.

1.2 Optimal Actions

Our starting point is the standard assumption that optimal actions maximize the subjective expected present-discounted value of current and future utility, i.e., at time t the individual chooses actions to maximize

$$V_t := \hat{\mathbb{E}}_t \left[\sum_{\tau \geq t} \beta^{\tau - t} U_\tau \right], \tag{2}$$

where $\hat{\mathbb{E}}_t[\cdot]$ denotes subjective expectation at time t, U_t denotes utility at time t, and $0 < \beta \leq 1$ is the conventional discount factor.

Our point of departure from the standard model of subjective expected utility is that — as proposed by early economists like Bentham, Hume, and Böhm-Bawerk, and as assumed, e.g., in Loewenstein (1987) and Caplin and Leahy (2001) — people's present utility depends both on what they currently consume and on what they anticipate consuming in the future. The implication of incorporating utility from both sources — consumption as well as anticipation — in U is that current utility is affected by subjective expectations of future events, and people can be happier by holding more optimistic beliefs as well as by forming better plans.

In particular, we follow the literature on anticipatory utility (see, e.g., Loewenstein, 1987), and

[7] We normalize the known component of work required to 1, so η is the relative size of the random component.

[8] The constraint $w_1 \leq 1$ ensures that neither a rational agent nor one with optimal beliefs chooses $w_1 > 1$, leading to wasted work for some η, in which case our interpretation of the mathematical structure is strained.

define utility U as the expected present-discounted value of current and future consumption utility:

$$U_t := \hat{\mathbb{E}}_t \left[\sum_{\tau \geq t} \phi^{\tau - t} u(w_\tau) \right], \tag{3}$$

where $u(w) := -\frac{1}{2} w^2$ is the consumption utility the agent directly experiences from doing work w (and consuming the corresponding amount of leisure), and $0 < \phi \leq 1$ is the anticipatory discount factor.[9] For example, utility at $t = 1$ is $U_1 = \hat{\mathbb{E}}_1 [u(w_1) + \phi u(w_2)]$, where $u(w_1)$ is the consumption utility that is considered in standard economic models, and $\phi u(w_2)$ is the anticipatory utility from savoring or dreading future consumption.

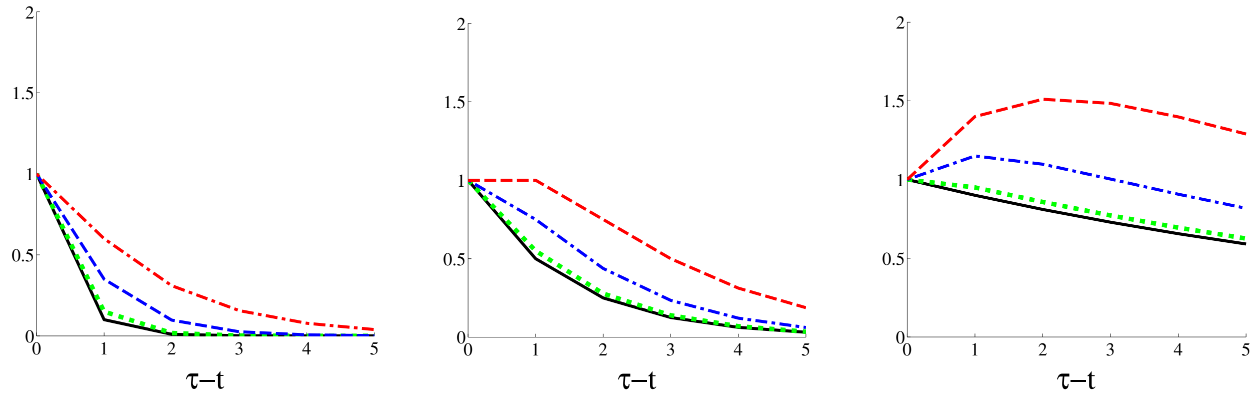
It is important to distinguish between the conventional discount factor β and the anticipatory discount factor ϕ. The former is the discount factor that appears in standard economic models and captures the degree to which the agent cares about his future selves, while the latter appears in models of anticipatory utility and captures the degree to which he gains immediate utility from anticipating future consumption. Indeed, with no anticipatory utility ($\phi = 0$) and objective expectations ($\hat{\mathbb{E}}_t = \mathbb{E}_t$, where $\mathbb{E}_t$ is the objective expectation at time t), our model reduces to the standard consumption model. The inclusion of anticipatory utility with discount factor ϕ in an otherwise standard consumption model generates a non-exponential compound discount factor in the preferences represented by V. Using objective beliefs and Equation 3 in Equation 2, we write

$$V_t = \mathbb{E}_t \left[\sum_{\tau \geq t} h_{\tau - t}(\beta, \phi) \; u(w_\tau) \right],$$

where $h_{\tau-t}(\beta, \phi) := \sum_{k=0}^{\tau-t} \beta^k \phi^{\tau-t-k}$ is the complete homogeneous symmetric polynomial of degree $\tau - t$ in β and ϕ, i.e., the sum of all cross-terms of power $\tau - t$.[10] For example, the compound discount factor applied to consumption utility one period ahead is $\beta + \phi$, and the one applied to consumption utility two periods ahead is $\beta^2 + \beta\phi + \phi^2$. We note that as long as $\beta + \phi \leq 1$, the compound discount factor does not exceed 1 for any period. Figure 2 plots the compound discount factor applied at period t to consumption utility that is experienced $\tau - t$ periods in the future, for various values of the conventional discount factor β and the anticipatory discount factor ϕ.

[9] The assumption that u is quadratic is important for tractability; it implies that only the mean and variance of η matter, so it reduces the dimensionality of the problem of solving for optimal beliefs. But our qualitative results are more general; for example, restricting attention to beliefs that can be completely ordered by first-order stochastic dominance, we can show that they hold for a concave u.

[10] This compound discount factor, $h_{\tau-t}(\beta, \phi)$, is the *devaluation* factor of Loewenstein (1987), with the difference that in that case utility flows are continuous.



(a) $\beta = 0.1$ (b) $\beta = 0.5$ (c) $\beta = 0.9$

Figure 2: Plots, for various values of the conventional discount factor β and the anticipatory discount factor ϕ, of the compound discount factor $h_{\tau-\tau}(\beta, \phi)$ in V_t applied to consumption utility of time τ, i.e., $\tau - t$ periods in the future. In all panels, the solid black lines correspond to $\phi = 0$, the dotted green lines to $\phi = 0.1$, the dash-dotted blue lines to $\phi = 0.25$, and the dashed red lines to $\phi = 0.5$. The compound discount factors are plotted for $\beta = 0.1$ in Panel (a), for $\beta = 0.5$ in Panel (b), and for $\beta = 0.9$ in Panel (c).

In all panels in the figure, the solid black line plots the standard exponential discount factor, which corresponds to the case of no anticipatory utility, i.e., $\phi = 0$. These plots demonstrate that, with $\phi > 0$, the compound discount factor discounts the future less than the standard exponential discount factor, hence introduces a *future* bias in preferences. As a result, the optimal work w_1 from the time-0 perspective is smaller than the optimal w_1 from the time-1 perspective. Crucially, this bias is the opposite of the present bias assumed in the literature on time-inconsistent preferences, and cannot induce at $t = 0$ the choice of a deadline of the form $w_1 \geq \psi$; as a result, it does not affect our message that time-inconsistent beliefs endogenously cause preference for commitment.[11]

1.3 Optimal Beliefs

In our model, agents may hold subjective beliefs that may differ from objective beliefs. Formally, let $(\Omega, \mathcal{F})$ be the measurable space with Ω the sample space and $\mathcal{F}$ a $\sigma-$algebra of Ω representing the possible events, and let $\{\mathcal{F}_t\}_{t\in T}$ with $\mathcal{F}_s \subset \mathcal{F}_t \subset \mathcal{F}$, $\forall s < t \in T$, be a filtration representing information flow, with T the time index set. Objective beliefs are represented by the objective probability measure $\mathbb{P}$ on $\mathcal{F}$, while subjective beliefs at time t are represented by the $\mathcal{F}_t$-measurable conditional probability measure $\hat{\mathbb{P}}_t$ on $\mathcal{F}$. We use $\hat{\mathbb{E}}_t$ to denote expectations given

[11]We note that if we are willing to assume that people also care about the past, then it is possible to introduce anticipatory together with memory utility, without introducing any time-inconsistency in preferences. Indeed, in Section B.1 of the online appendix we extend our model by introducing the possibility that the past also matters, and show that, as expected, all our results continue to hold in the absence of this future bias.

beliefs $\hat{\mathbb{P}}_t$ and information represented by $\mathcal{F}_t$. We assume that the objective conditional probability measure $\mathbb{P}_t := \mathbb{P}(\cdot|\mathcal{F}_t)$ dominates $\hat{\mathbb{P}}_t$ for all t (i.e., $\mathbb{P}_t(A) = 0$ implies $\hat{\mathbb{P}}_t(A) = 0$, for $A \in \mathcal{F}$), which ensures that the agent understands the underlying model at all times and that information that has been revealed is not forgotten.[12] We also assume that the agent knows the full sequence of probability measures $\left\{\hat{\mathbb{P}}_t\right\}_{t \in T}$; this ensures that when the agent at time t makes his optimal decision to maximize $V_t = \hat{\mathbb{E}}_t\left[\sum_{\tau \geq t} \beta^{\tau-t} U_\tau\right]$, he correctly evaluates the $\mathcal{F}_\tau-$measurable utility $U_\tau = \hat{\mathbb{E}}_\tau\left[\sum_{\tau' \geq \tau} \phi^{\tau'-\tau} u(\omega_{\tau'})\right]$ at time $\tau \geq t$ using time-τ subjective beliefs $\hat{\mathbb{P}}_\tau$, and then he calculates the conditional expectation of the discounted sum of current and future utilities using time-t beliefs and information.

Our innovation is that we do not place any restrictions on how $\hat{\mathbb{P}}_s$ and $\hat{\mathbb{P}}_t$ are related for $s < t \in T$. In particular, given some $\mathcal{F}-$measurable random variable X, we do not require that $\hat{\mathbb{E}}_s[X] = \hat{\mathbb{E}}_s\left[\hat{\mathbb{E}}_t[X]\right]$, that is the law of iterated expectations and Bayesian updating may not hold. In the setup of Section 1.1, Ω is the sample space that dictates all the possible values of the random variable η, $\{\mathcal{F}_t\}_{0 \leq t \leq 2}$ is the filtration generated by η whose value is revealed at $t = 2$, the assumption that $\eta \geq 0$ together with the assumption that $\mathbb{P}_t$ dominates $\hat{\mathbb{P}}_t$ implies that $\hat{\mathbb{P}}_t(\eta < 0) = 0$ hence $\hat{\mathbb{E}}_t[\eta] \geq 0$, and our innovation to not place restrictions on the relation between beliefs at different points in time implies that we do not require $\hat{\mathbb{E}}_0[\eta] = \hat{\mathbb{E}}_1[\eta]$, even though no information is revealed between $t = 0$ and $t = 1$. Our framework provides the necessary structure that allows us to relax the standard — but restrictive — assumption of time-consistent beliefs, in a way that is consistent with the economic principle of maximization. As a result, we gain new insights regarding preference for commitment and time-inconsistent beliefs and preferences.

Now, we turn our attention to how the optimal subjective beliefs are determined. As explained in Section 1.2, since an individual's current utility depends on his expectations about the course of the future, more optimistic beliefs about the future make him happier in the present. But such beliefs come at a cost: they sacrifice the benefits of smoothing work over time, since an individual who believes that a task is unrealistically easy to complete exerts less work in period 1 and ends up working more in period 2. So there is a trade-off between optimism, which raises anticipatory utility, and objectivity, which allows better smoothing of work over time. Like Landier (2000) and Brunnermeier and Parker (2005), to capture this trade-off we define a welfare function, which we refer to as the individual's *well-being* $\mathcal{W}$, as the objective expected discounted sum of utility over

[12] Thinking in terms of a tree diagram, at time t the agent knows which node has been reached, and his time-t subjective beliefs assign a probability to each of the nodes that can be reached from the current node.

states and time. Optimal beliefs maximize $\mathcal{W}$, given that the individual makes optimal decisions based on these beliefs.

Definition. *Optimal beliefs are characterized by the sequence* $\left\{\hat{\mathbb{P}}_t\right\}_{t\in T}$ *of subjective beliefs for each* $t \in T$*, that maximizes well-being*

$$\mathcal{W} := \mathbb{E}\left[\sum_{\tau\geq 0}\beta^{\tau}U_{\tau}\right], \tag{4}$$

where actions are optimally chosen given beliefs and subject to resource constraints.

In our setup, quadratic utility implies that only the mean and variance of η matter, so subjective beliefs at time t are determined by the subjective mean and variance; this greatly reduces the complexity of solving for optimal beliefs. In addition, since the only random variable in our model is η, we simplify notation by defining its subjective mean and variance at time t as $\hat{E}_t := \hat{\mathbb{E}}_t[\eta]$ and $\hat{\Sigma}_t := \hat{\mathbb{E}}_t\left[\eta^2\right] - \left(\hat{\mathbb{E}}_t\left[\eta\right]\right)^2$, and use the explicit form $\hat{\mathbb{E}}_t\left[\eta\right]$ only when stating results.

We choose the welfare function in Equation 4 for the following reasons. First, as argued in Caplin and Leahy (2000), with anticipatory utility it is natural to consider a welfare function that involves the (weighted) sum of utility from all sources, rather than just consumption utility, over time. In fact, in a deterministic setting, $\mathcal{W}$ collapses to an example of the social welfare function proposed by Caplin and Leahy (2004). Second, comparing $V_0 = \hat{\mathbb{E}}_0\left[\sum_{\tau\geq 0}\beta^{\tau}U_{\tau}\right]$ with $\mathcal{W} = \mathbb{E}\left[\sum_{\tau\geq 0}\beta^{\tau}U_{\tau}\right]$ we see that this choice of $\mathcal{W}$ has the nice property that V_0 and $\mathcal{W}$ are the same, except that the former is evaluated at time-0 subjective beliefs while the latter is evaluated at objective beliefs. Indeed, we use objective expectations to evaluate well-being, because we are interested in capturing the trade-off between distorting beliefs and distorting actions, across realizations of uncertainty; objective expectations capture this since uncertainty unfolds according to objective rather than subjective probabilities.

While our objective function for beliefs captures the central trade-off of interest parsimoniously, other candidates could also be considered. An alternative that is close to — but computationally more complex than — what we do, is that optimal beliefs at time t maximize the well-being from time-t's perspective, i.e., for each $t \in T$, optimal beliefs at time t maximize $\mathcal{W}_t := \mathbb{E}_t\left[\sum_{\tau\geq t}\beta^{\tau-t}U_{\tau}\right]$. The same trade-off would exist between optimism which raises anticipatory utility, and objectivity which allows better smoothing, so our main results would hold. An alternative that is conceptually further from what we do, is that the objective for beliefs is evaluated at subjective beliefs. In this case, completely optimistic beliefs would be trivially optimal, so the model would not capture the trade-off of interest, unless it incorporated explicit costs of belief distortion.[13]

In terms of interpretation, beliefs may be formed unconsciously or partly consciously. Our model captures an optimal balancing of the benefits of unrealistic optimism against the average costs of the poor behavior that they induce, rather than any particular mechanism through which this is accomplished. An example of one possible mechanism is that people have an inherent optimistic bias which they choose to discipline with attention and more realistic thinking when they see potentially costly outcomes. Alternatively, beliefs could be formed through an automatic process which is the result, e.g., of evolutionary forces.[14,15]

2 Optimism and overconfidence

In this section, we present results for the case without access to a commitment device. First, we show that optimism and overconfidence maximize the well-being, hence they are optimal; as a result, people exhibit the planning fallacy in planning situations. Second, it is optimal for an individual to become less optimistic as the temporal distance to the task decreases.

Given some arbitrary subjective beliefs, the individual chooses work w_1 at time 1 to maximize the subjective expected present-discounted value of current and future utility, $V_1 = \hat{\mathbb{E}}_1 [U_1 + \beta U_2]$, subject to the constraint $w_1 \leq 1$ and to the resource constraint $w_1 + w_2 = 1 + \eta$.

Proposition 1. *(Optimal work given arbitrary beliefs)*
With arbitrary subjective expectations $\hat{\mathbb{E}}_0$, $\hat{\mathbb{E}}_1$, at $t = 1$ the agent chooses action

$$w_1^* \left(\left\{ \hat{\mathbb{E}}_t \right\} \right) := \min \left\{ 1, B_1 \left(1 + \hat{\mathbb{E}}_1 [\eta] \right) \right\}, \tag{5}$$

with $B_1 := \frac{\beta+\phi}{1+\beta+\phi}$.[16] *With rational expectations, $w_1^{\mathrm{RE}} := w_1^*(\mathbb{E}) = \min\{1, B_1(1+\mathbb{E}[\eta])\}$.*

[13] Modeling explicit psychological costs of belief distortion requires a theory of the internal process through which this distortion is achieved. This not only falls more into the scope of the psychology literature, but also complicates the analysis without providing much economic insight. We note, however, that introducing a simplistic yet sensible explicit cost of belief distortion — a quadratic cost in the period in which the distortion occurs — should not affect our qualitative results; see Section B.3 in the online appendix for a formal proof showing that our main results on optimism and the planning fallacy still hold.

[14] Veenhoven (2008) surveys the evidence from more than 30 studies on the effects of happiness on longevity and shows that happy people live on average a decade longer than unhappy people. He also cites evidence that happiness has a much stronger effect on later health than prior health does on later happiness. So holding beliefs that maximize the discounted sum of objective expected utility over time may maximize health, hence may become an instinct due to natural selection.

[15] We also note that, though market selection could lead to some attrition of agents with optimal (and presumably incorrect) beliefs, agents with optimal beliefs would still survive, not only because all agents hold optimal beliefs of varying degrees according to our theory, but also because holding these beliefs may grant advantages, e.g., in terms of overall health through happiness as already mentioned, or in terms of riskier actions hence higher expected returns (see the literatures on the survival of noise traders and on CEO overconfidence for related arguments).

[16] The notation $w_1^* \left(\left\{ \hat{\mathbb{E}}_t \right\} \right)$ makes it clear that this is optimal work given arbitrary expectations $\hat{\mathbb{E}}_t$ for $t \in T$.

To interpret this result, we make the following observations. First, by certainty equivalence, only the mean of η matters for behavior. Second, because utility is concave, the agent wants to smooth work across periods. Finally, in a standard model, the agent smooths work by choosing w_1 to be a fraction of expected total work using objective beliefs; in our model he uses subjective beliefs.

Our first main result is that, from the perspective of the well-being, optimism and the planning fallacy are optimal, while rational expectations are suboptimal. In what follows, the superscript ND indicates optimal beliefs and corresponding decisions in the no-deadline case, and the superscript RE indicates optimal decisions under rational expectations.

Proposition 2. *(Optimism and the planning fallacy are optimal)*
(i) Optimal expectations are characterized by $\hat{\mathbb{E}}_0^{\mathrm{ND}}[\eta]=0$ *and* $\hat{\mathbb{E}}_1^{\mathrm{ND}}[\eta]$ *a piece-wise linear, weakly increasing function of* $\mathbb{E}[\eta]$.[17]
(ii) Optimal beliefs are optimistic, i.e., $\hat{\mathbb{E}}_0^{\mathrm{ND}}[\eta]<\mathbb{E}[\eta]$ *and* $\hat{\mathbb{E}}_1^{\mathrm{ND}}[\eta]<\mathbb{E}[\eta]$.
(iii) Over time, beliefs become less optimistic, i.e., $\hat{\mathbb{E}}_0^{\mathrm{ND}}[\eta] \leq \hat{\mathbb{E}}_1^{\mathrm{ND}}[\eta]$.
(iv) The planning fallacy (under-estimation of task duration) is optimal, i.e.,
$\hat{\mathbb{E}}_0^{\mathrm{ND}}\left[w_1^*+w_2^*\right] < \mathbb{E}\left[w_1^*+w_2^*\right]$ *and* $\hat{\mathbb{E}}_1^{\mathrm{ND}}\left[w_1^*+w_2^*\right] < \mathbb{E}\left[w_1^*+w_2^*\right]$.
(v) The optimal work is $w_1^{\mathrm{ND}} := w_1^*\left(\left\{\hat{\mathbb{E}}_t^{\mathrm{ND}}\right\}\right) = B_1\left(1+\hat{\mathbb{E}}_1^{\mathrm{ND}}[\eta]\right) \leq w_1^{\mathrm{RE}}$.

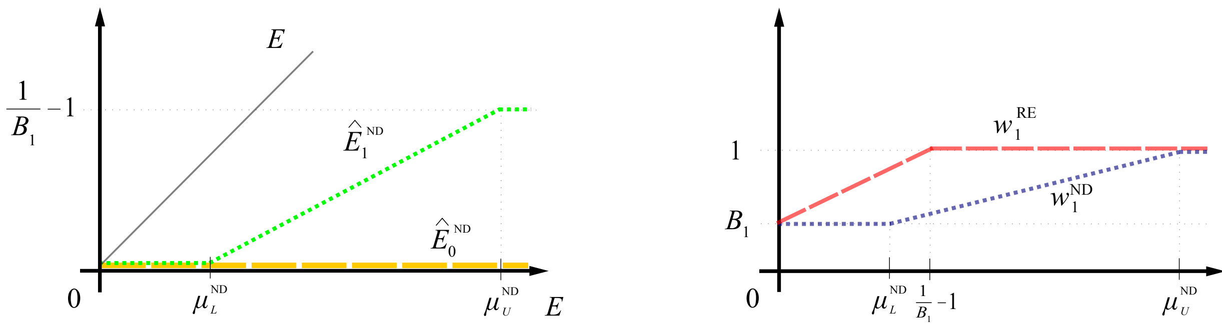
The results in this proposition are illustrated in Figure 3.

So, according to our theory, people at a subconscious level know the objective distribution, but it is optimal to be optimistic, hence they underestimate the amount of work that the project requires and do less work at time 1 than they would have chosen to do if they held objective beliefs. This description of the planning fallacy is similar to its original description (Kahneman and Tversky, 1979; see quote in introduction). To understand why optimism is optimal, we consider its benefits and costs. Optimism has anticipatory benefits for time-0 utility, real and anticipatory benefits for time-1 utility, but an average cost due to suboptimal smoothing for time-2 utility. Optimal beliefs are optimistic, because anticipatory benefits are first-order, while behavioral costs are second-order.[18]

Subsequently, we omit the argument, unless we want to refer to optimal work given some specific beliefs.

[17] We note that the extreme result $\hat{\mathbb{E}}_0^{\mathrm{ND}}[\eta] = 0$ follows from the standard assumption that the agent does not care about the past, and from our simplifying assumption of no explicit costs of belief distortion. The expression for $\hat{\mathbb{E}}_1^{\mathrm{ND}}[\eta]$ is given in the appendix.

[18] We note that our theory is not incompatible with evidence of pessimism related to tail-risk events, e.g., high demand for insurance policies with low deductible and low demand for annuities (often explained as due to people over-weighting the probability of unlikely events). While building our model of optimal beliefs on top of rational preferences does not predict this pessimism, building it on top of, e.g., prospect theory preferences which incorporate probability weighting, would yield agents who are optimistic over most events and pessimistic over tail events.



(a) Optimal Expectations (b) Optimal Action

Figure 3: In Panel (a), the solid black 45° line plots the rational expectation E, the dashed yellow line plots the optimal expectation $\hat{E}_0^{ND}$ at $t = 0$, and the dotted green line plots the optimal expectation $\hat{E}_1^{ND}$ at $t = 1$, all as functions of E, the objective mean of η. In Panel (b), the dashed red line plots the optimal work w_1^{RE} given rational expectations, and the dotted blue line plots the optimal work w_1^{ND} given optimal beliefs absent a commitment device, both as functions of E. The constraint $\hat{E}_0^{ND} \geq 0$ binds everywhere and the constraint $\hat{E}_1^{ND} \geq 0$ binds below μ_L^{ND}, while the constraints $w_1^{RE} \leq 1$ and $w_1^{ND} \leq 1$ bind above $\frac{1}{B_1}$ and above μ_U^{ND}, respectively. All constants are defined in the appendix.

Our model predicts that optimism and the planning fallacy are greater, the greater the anticipatory benefits of optimism and the smaller the ex post costs of mis-planning. While we defer to Section 4 on comparative statics a discussion of the effect of the conventional discount factor β and the anticipatory discount factor ϕ on optimal beliefs, Figure 3 shows that optimal expectations are an increasing piece-wise linear function of the objective expectation, E. Optimal expectations are increasing in E, because the cost of belief distortion due to suboptimal smoothing increases in E; linearity follows from quadratic utility, while the constant segments are due to the constraints $w_1 \leq 1$ and $\hat{E}_t \geq 0$ for all t.

Our theory also predicts that beliefs become less optimistic over time. From Proposition 1, we know that optimal work w_1^* given arbitrary subjective beliefs is a fraction, B_1, of time-1 expected total work. Beliefs at $t = 0$ do not matter at $t = 1$, hence do not determine w_1^*. As a result, while optimistic deviations of time-0 and time-1 subjective beliefs from objective beliefs both yield anticipatory benefits since they both imply a reduction in anticipated total work, the former induces no distortionary costs on work smoothing, while the latter does. Consequently, time-0 optimal beliefs are more optimistic than time-1 optimal beliefs.

Our second main result is that people are overconfident about their predictions.

Proposition 3. *(Overconfidence is optimal)*

(i) From objective beliefs, a small decrease in the perceived uncertainty about future work increases the agent's well-being: $\frac{d\mathcal{W}}{d\hat{\Sigma}_0}|_{\hat{\mathbb{P}}_t=\mathbb{P}_t} < 0$ *and* $\frac{d\mathcal{W}}{d\hat{\Sigma}_1}|_{\hat{\mathbb{P}}_t=\mathbb{P}_t} < 0.$

(ii) The agent's well-being is maximized by the belief that he knows what work level will be required: $\hat{\Sigma}_0^{\text{ND}} = \hat{\Sigma}_1^{\text{ND}} = 0 < \Sigma$.

Certainty equivalence implies that overconfidence has no behavioral consequences, but since utility is concave, it does have anticipatory benefits. Thus, the optimal subjective uncertainty is the corner solution of certainty. Our result is extreme because utility is quadratic and there is no explicit cost to distorting beliefs, but some overconfidence is optimal for a wider range of utility functions.[19] Yet even in its extremity, our prediction again closely matches the initial psychological interpretation of the planning fallacy by Kahneman and Tversky (1979), that people focus on a single plausible scenario for completing the task and ignore uncertainty:

> The planning fallacy is a consequence of the tendency to neglect distributional data and to adopt what may be termed an internal approach to prediction, in which one focuses on the constituents of the specific problem rather than on the distributional outcomes in similar cases. – Kahneman and Tversky (1979)

We note that, while in theory a reduction in the subjective variance is distinct from a reduction in the subjective mean, in practice experimental evidence sometimes blurs this difference. The bias in mean explains both main experimental findings: (i) overconfidence about how often the task will be completed by the predicted time, and (ii) underestimation of task completion time on average. The bias in variance makes overconfidence more extreme. In any case, our model delivers both biases.

3 Preference for commitment

We now examine optimal behavior and beliefs in the presence of a commitment device, in particular a deadline that may be self-imposed. Our theory predicts that people optimally self-impose constraints, which improve the intertemporal smoothing of work but do not implement perfect smoothing. The novelty of our results relative to extant economic theories of preference for commitment is twofold. First, the desire for commitment arises from changing beliefs over time rather than from changing preferences. Second, we do not assume that preferences or beliefs are present-biased, rather we show conditions under which commitment is optimal and beliefs are time-inconsistent

[19] In general, sufficient curvature in utility relative to marginal utility is needed. The more curved utility is, the larger the direct increase in well-being for a reduction in variance. The less curved marginal utility is, the smaller the behavioral response to the decrease in uncertainty, from the precautionary channel, so the smaller the indirect decrease in well-being from changed behavior.

and become more optimistic over time. Thus, our theory also provides predictions for beliefs and behavior that differ from those of a preference-based theory; we defer to Section 6 a detailed comparison of the predictions of the two theories.

We explain the intuition behind our results on self-imposed deadlines, and we present them formally in Proposition 4. Then in Proposition 5, we contrast the optimal self-imposed deadline with the deadline that an external observer with objective beliefs would choose, or equivalently the deadline that maximizes the well-being.

Given some arbitrary subjective beliefs, at $t = 0$ the agent wants to maximize the time-0 subjective expected present-discounted value of current and future utility, $V_0 = \hat{\mathbb{E}}_0\left[U_0 + \beta U_1 + \beta^2 U_2\right]$. From his time-0 perspective, the optimal amount of work at time 1 is

$$w_1^{*,0}\left(\left\{\hat{\mathbb{E}}_t\right\}\right) := \min\left\{1,\, D_0\left(1 + \hat{\mathbb{E}}_0\left[\eta\right]\right) + D_1\left(1 + \hat{\mathbb{E}}_1\left[\eta\right]\right)\right\}, \tag{6}$$

where $D_0 := \frac{\beta^2+\phi^2}{\beta+\phi+\beta^2+\phi^2+\beta\phi}$, $D_1 := \frac{\beta\phi}{\beta+\phi+\beta^2+\phi^2+\beta\phi}$. This amount of work is different from what the agent will in fact choose at time 1, which is $w_1^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = \min\left\{1,\, B_1\left(1 + \hat{\mathbb{E}}_1\left[\eta\right]\right)\right\}$ (see Proposition 1). This is because, from the agent's time-0 perspective, time-0 beliefs as well as time-1 beliefs matter, so $w_1^{*,0}$ is a fraction of time-0 expected total work plus a fraction of time-1 expected total work.

The agent prefers commitment at period 0 and chooses the binding deadline $\psi = w_1^{*,0}$ if he believes that he will subsequently choose to work too little at period 1, i.e., $w_1^* < w_1^{*,0}$. Comparing Equations 5 and 6, this means that the agent chooses a binding deadline only if subjective beliefs satisfy $\hat{E}_0 > \hat{E}_1$, i.e., they become more optimistic over time.[20] The agent may or may not consciously understand that becoming more optimistic is the cause of suboptimal work smoothing. Nevertheless, it is this belief time-inconsistency, and the agent's awareness of the resulting behavior, that leads to his willingness to overcome it by setting a binding deadline for himself.

If a deadline is chosen, optimal beliefs are different than without a commitment device. Time-0 beliefs are more pessimistic because, with a deadline, an optimistic distortion of time-0 beliefs has the same anticipatory benefit as without a deadline, but a higher distortionary cost, since it distorts the optimal deadline more than it distorts the optimal work without a deadline. In contrast, time-1 beliefs are more optimistic because, with a deadline, an optimistic distortion of time-1 beliefs has the same anticipatory benefit as without a deadline, but a lower distortionary cost, since it dis-

[20] In detail, $w_1^{*,0} > w_1^* \Leftrightarrow D_0\left(1+\hat{E}_0\right) > \left(B_1 - D_1\right)\left(1+\hat{E}_1\right)$, and we can show $D_0 < B_1 - D_1$.

torts the optimal deadline less than it distorts the optimal work without a deadline. To see all this, compare Equations 5 and 6, noting that $D_1 < B_1$ and $D_0 > 0$.

Finally, we need to understand for what parameter values (β, ϕ, and E) commitment is optimal. We have already explained that commitment leads to better smoothing of work, and that the associated beliefs lead to lower anticipatory utility in period 0, but higher anticipatory utility in period 1. In addition, the higher the conventional discount factor β is, the higher is the cost of suboptimal smoothing, while the smaller the anticipatory discount factor ϕ is, the more important is the period-1 anticipatory utility relative to the period-0 anticipatory utility in the well-being. Thus, the higher β is relative to ϕ, the higher is the value of commitment hence the more likely is it to be optimal. We show that there is a cutoff $\overline{\beta}(\phi)$, increasing in the anticipatory discount factor ϕ, such that a binding deadline may be optimally chosen if and only if $\beta > \overline{\beta}(\phi)$.[21] Furthermore, a binding deadline is more likely to be optimally chosen (the difference in the well-being $\mathcal{W}$ with and without a deadline is greatest) for intermediate values of the objective mean E. In principle, the benefit from the improved work smoothing that a deadline implements should be minimal when expected total work is small (E is small) and maximal when it is large (E is large). However, due to the constraint $w_1 \leq 1$, there is no such benefit when expected total work is large enough that the optimal amount of work w_1^* without a deadline equals 1, hence the benefit is maximal for intermediate values of the objective mean E.

The following proposition states formally these results and provides the expression for the optimal self-imposed deadline. The superscript D indicates optimal quantities in the case of a self-imposed deadline.

Proposition 4. *(Self-imposed deadline)*
(i) If the degree to which the agent cares about his future selves is small relative to the degree to which he gains immediate utility from anticipating future consumption ($\beta \leq \overline{\beta}(\phi)$) or the objective expectation of work is outside a certain range ($\mathbb{E}[\eta] \notin M(\beta, \phi)$), optimal beliefs are identical to those absent a commitment device and the agent does not impose a binding deadline.[22]

[21] We note that $\overline{\beta}(\phi)$ is close to 2ϕ. Keeping this in mind, one can see in Figure 2 the plots of the compound discount factors for which an optimal deadline will be chosen for some values of E.

[22] In the proof of this proposition in Section A.4 of the appendix, we provide expressions for the optimal beliefs, as well as the cutoff $\overline{\beta}(\phi)$ and the set $M(\beta, \phi)$. We note here that $\overline{\beta}$ is increasing in ϕ and that M is a convex set. M is of the form $\left(\underline{\mu}(\beta, \phi), \infty\right)$, i.e., there is a cutoff value $\underline{\mu}$ above which a binding deadline is optimally chosen, except for values of β and ϕ near the region where M switches from being empty to being non-empty. Also, using a very fine grid of values for β, ϕ, and E, we can show numerically that $M(\beta_1, \phi) \subset M(\beta_2, \phi)$ for $\beta_1 < \beta_2$ and $M(\beta, \phi_1) \supset M(\beta, \phi_2)$ for $\phi_1 < \phi_2$.

(ii) If the degree to which the agent cares about his future selves is large relative to the degree to which he gains immediate utility from anticipating future consumption ($\beta > \overline{\beta}(\phi)$) and the objective expectation of work is intermediate ($\mathbb{E}[\eta] \in M(\beta, \phi)$), then:

- *Optimal expectations $\hat{\mathbb{E}}_0^{\mathrm{D}}[\eta]$, $\hat{\mathbb{E}}_1^{\mathrm{D}}[\eta]$ are weakly increasing functions of $\mathbb{E}[\eta]$.*

- *Optimal beliefs are optimistic ($\hat{\mathbb{E}}_0^{\mathrm{D}}[\eta] < \mathbb{E}[\eta]$ and $\hat{\mathbb{E}}_1^{\mathrm{D}}[\eta] < \mathbb{E}[\eta]$).*

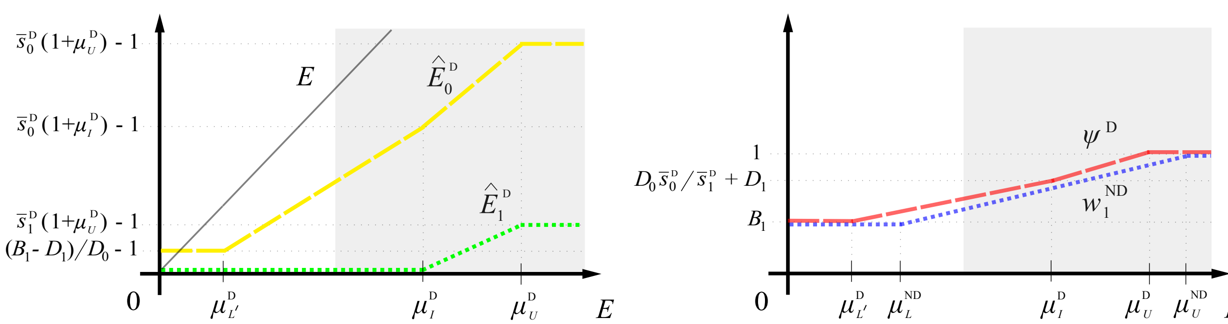
- *Optimal beliefs become more optimistic over time ($\hat{\mathbb{E}}_0^{\mathrm{D}}[\eta] \geq \hat{\mathbb{E}}_1^{\mathrm{D}}[\eta]$).*

- *Time-0 optimal beliefs are more pessimistic ($\hat{\mathbb{E}}_0^{\mathrm{D}}[\eta] \geq \hat{\mathbb{E}}_0^{\mathrm{ND}}[\eta]$) and time-1 optimal beliefs more optimistic ($\hat{\mathbb{E}}_1^{\mathrm{D}}[\eta] \leq \hat{\mathbb{E}}_1^{\mathrm{ND}}[\eta]$) than absent a commitment device.*

- *The optimal deadline $\psi^{\mathrm{D}} := D_0\left(1+\hat{\mathbb{E}}_0^{\mathrm{D}}[\eta]\right) + D_1\left(1+\hat{\mathbb{E}}_1^{\mathrm{D}}[\eta]\right)$ binds ($w_1^*\left(\left\{\hat{\mathbb{E}}_t^{\mathrm{D}}\right\}\right) \leq \psi^{\mathrm{D}}$), but is smaller than w_1^{RE}.*

- *Complete overconfidence is optimal ($\hat{\Sigma}_0^{\mathrm{D}} = \hat{\Sigma}_1^{\mathrm{D}} = 0 < \Sigma$).*

Figure 4 plots the optimal expectations that implement a binding deadline (Panel a) and the optimal deadline versus the optimal work absent a commitment device (Panel b).



(a) Optimal Expectations (b) Optimal Action

Figure 4: In Panel (a), the solid black 45° line plots the rational expectation E, the dashed yellow line plots the optimal expectation $\hat{E}_0^{\mathrm{D}}$ at $t = 0$, and the dotted green line plots the optimal expectation $\hat{E}_1^{\mathrm{D}}$ at $t = 1$, all as functions of E, the objective mean of η. In Panel (b), the dashed red line plots the optimal deadline ψ^{D} and the dotted blue line plots the optimal work w_1^{ND} given optimal beliefs absent a commitment device, both as functions of E. In both panels, the shaded region indicates the values of E for which the binding deadline is optimally chosen. The constraints $\hat{E}_0 \geq 0$ and $\hat{E}_1 \geq 0$ bind below $\mu_{L'}^{\mathrm{D}}$ and μ_I^{D}, respectively, while the constraint $\psi \leq 1$ binds above μ_U^{D}. All constants are defined in the appendix.

Next, we contrast the optimal self-imposed deadline with the deadline that an external observer with objective beliefs would choose, or equivalently the deadline that maximizes the well-being.

That is, we contrast the self-imposed deadline that the agent himself chooses at period 0 to maximize the subjectively expected discounted sum of utility, $\hat{\mathbb{E}}_0\left[\sum_{\tau\geq 0}\beta^\tau U_\tau\right]$, with the externally-imposed deadline that an objective external observer chooses at period 0 to maximize the objectively expected discounted sum of the agent's utility, including (subjective) anticipatory utility, $\mathbb{E}\left[\sum_{\tau\geq 0}\beta^\tau U_\tau\right]$, which is also the well-being. In what follows, the superscript ED indicates optimal quantities in the case of an externally-imposed deadline.

Proposition 5. *(Externally-imposed deadline)*
(i) Optimal beliefs are optimistic ($\hat{\mathbb{E}}_0^{\mathrm{ED}}[\eta] < \mathbb{E}[\eta]$ and $\hat{\mathbb{E}}_1^{\mathrm{ED}}[\eta] < \mathbb{E}[\eta]$).
(ii) Optimal beliefs become more pessimistic over time ($\hat{\mathbb{E}}_0^{\mathrm{ED}}[\eta] \leq \hat{\mathbb{E}}_1^{\mathrm{ED}}[\eta]$).
(iii) Optimal beliefs are more optimistic ($\hat{\mathbb{E}}_0^{\mathrm{ED}}[\eta] \leq \hat{\mathbb{E}}_0^{\mathrm{D}}[\eta]$ and $\hat{\mathbb{E}}_1^{\mathrm{ED}}[\eta] \leq \hat{\mathbb{E}}_1^{\mathrm{D}}[\eta]$) than with a self-imposed deadline.
(iv) The optimal deadline $\psi^{\mathrm{ED}} := \min\left\{1,\, D_0(1+\mathbb{E}[\eta])+D_1\left(1+\hat{\mathbb{E}}_1^{\mathrm{ED}}[\eta]\right)\right\}$ binds ($w_1^*\left(\left\{\hat{\mathbb{E}}_t^{\mathrm{ED}}\right\}\right) \leq \psi^{\mathrm{ED}}$), is stricter than a self-imposed deadline ($\psi^{\mathrm{D}} \leq \psi^{\mathrm{ED}}$), but smaller than w_1^{RE}.

The deadline chosen by an outsider with rational beliefs (equivalently the deadline that maximizes well-being) is stricter than the self-imposed one. The reason is that the rational outsider chooses the deadline using rational rather than the more optimistic subjective beliefs, hence he expects that more total work is required. But this externally-imposed deadline requires less work than the optimal amount of work w_1^{RE} that the agent himself would choose at time 1 if he had rational expectations. This is because, though the rational outsider maximizes the objectively expected discounted sum of the agent's utility, we can see from Equation 3 that utility itself is evaluated using the agent's optimistic subjective beliefs, hence the optimal amount of smoothing is different than if the agent had rational beliefs. In fact, optimal beliefs with an externally-imposed deadline are even more optimistic than with a self-imposed one, since: (i) optimistic time-0 beliefs have no distortionary cost, because they do not affect the externally-imposed deadline hence intertemporal smoothing; and (ii) optimistic time-1 beliefs have a smaller distortionary cost due to suboptimal smoothing, because the externally-imposed deadline is stricter.[23]

[23] In more detail, optimistic time-1 beliefs lead to higher w_2 than is rationally optimal; but with a stricter deadline, w_1 is larger and w_2 is smaller, so since u is concave, suboptimal smoothing is less costly.

4 Comparative statics

In this section, we examine how optimal beliefs and optimal decisions vary with the structural parameters of our model, the conventional discount factor β which captures the degree to which an agent cares about his future selves, and the anticipatory discount factor ϕ which captures the degree to which an agent gains immediate utility from anticipating future consumption. These comparative statics serve to further our understanding of what drives optimal beliefs and behavior in our model, and to contrast our theory with that of present-biased preferences, for which we also derive comparative statics in Section 5.2. This comparison is theoretically interesting, and it could also be empirically interesting, if it could form the basis of tests to evaluate the two theories. However, the use of these comparative statics in empirical tests is challenging, as it involves the estimation of individuals' time preference parameters; while there is a literature on the estimation of the conventional discount factor β, we are not aware of any research on the estimation of the anticipatory discount factor ϕ.

Our results here are that i) optimal beliefs both with and without a commitment device become less (more) optimistic as the conventional discount factor β (the anticipatory discount factor ϕ) increases; ii) the optimal work w_1^{ND} without a commitment device increases (first decreases, then increases) as β (ϕ) increases; and iii) the optimal deadline ψ^{D} increases (decreases) as β (ϕ) increases.[24]

Proposition 6. *(Comparative statics without a commitment device)*
*(i) Optimal time-*1 *beliefs become less optimistic and optimal work* w_1^{ND} *increases as agents care more about their future selves:* $\frac{\mathrm{d}\hat{E}_1^{\mathrm{ND}}}{\mathrm{d}\beta} > 0$ *and* $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\beta} > 0$.
*(ii) Optimal time-*1 *beliefs become more optimistic as agents get more utility from anticipation:* $\frac{\mathrm{d}\hat{E}_1^{\mathrm{ND}}}{\mathrm{d}\phi} < 0$. *Optimal work* w_1^{ND} *first decreases and then increases as* ϕ *increases: for* $\phi \leq \beta$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} < 0$*; for* $\beta < \phi \leq \overline{\phi}(\beta)$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} > 0$ *for small* $\mathbb{E}[\eta]$ *and* $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} < 0$ *for large* $\mathbb{E}[\eta]$*; and for* $\overline{\phi}(\beta) < \phi$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} > 0$.[25]

Time-0 beliefs are completely optimistic, so we have no interesting comparative statics for them.

Proposition 7. *(Comparative statics with a self-imposed deadline)*
If the commitment device is used (i.e., a binding deadline is optimally chosen):

[24] In all comparative statics, we focus on intermediate values of E such that the constraints $w_1 \leq 1$ and $\hat{E}_t \geq 0$ do not bind, i.e., optimal beliefs and work are interior rather than corner solutions.

[25] The definition of $\overline{\phi}(\beta)$ is given in the appendix.

*(i) Optimal time-*0 *and time-*1 *beliefs become less optimistic and the optimal deadline* ψ^{D} *increases as agents care more about their future selves:* $\frac{\mathrm{d}\hat{E}_0^{\mathrm{D}}}{\mathrm{d}\beta} > 0$, $\frac{\mathrm{d}\hat{E}_1^{\mathrm{D}}}{\mathrm{d}\beta} > 0$, *and* $\frac{\mathrm{d}\psi^{\mathrm{D}}}{\mathrm{d}\beta} > 0$.
*(ii) Optimal time-*0 *and time-*1 *beliefs become more optimistic and the optimal deadline* ψ^{D} *decreases as agents get more utility from anticipation:* $\frac{\mathrm{d}\hat{E}_0^{\mathrm{D}}}{\mathrm{d}\phi} < 0$, $\frac{\mathrm{d}\hat{E}_1^{\mathrm{D}}}{\mathrm{d}\phi} < 0$, *and* $\frac{\mathrm{d}\psi^{\mathrm{D}}}{\mathrm{d}\phi} < 0$.

To interpret the results of the preceding propositions, we remember that optimal beliefs are optimistic and are associated with a benefit from anticipatory utility and a cost from suboptimal smoothing. The benefit from holding optimistic beliefs comes from anticipatory utility, which is discounted by ϕ; so the less anticipated utility is discounted (the higher ϕ is), the more optimistic are the optimal beliefs (so $\frac{\mathrm{d}\hat{E}_0}{\mathrm{d}\phi} < 0$ and $\frac{\mathrm{d}\hat{E}_1}{\mathrm{d}\phi} < 0$). On the other hand, most of the cost from suboptimal smoothing comes from lowered utility in period 2, where the disutility of having to work more in period 2 is actually experienced; so the less future utility is discounted (the higher β is), the less optimistic are the optimal beliefs (so $\frac{\mathrm{d}\hat{E}_0}{\mathrm{d}\beta} > 0$ and $\frac{\mathrm{d}\hat{E}_1}{\mathrm{d}\beta} > 0$).

The effect of the conventional and of the anticipatory discount factor on optimal actions is both direct and indirect. We see from Proposition 2 that for the optimal work without a deadline, w_1^{ND}, there is a direct effect through the proportion, B_1, of expected total work that is optimally done in period 1, and an indirect effect through time-1 expectations. For given beliefs, a higher conventional discount factor β means that the agent cares more about his future selves so future utility from all sources is discounted less, and a higher anticipatory discount factor ϕ means that he savors future anticipated consumption more so anticipatory utility is discounted less. So with higher β and ϕ, the agent cares more about consumption utility at time 2, therefore the direct effect is that it is optimal to perform a higher proportion of expected total work at time 1 (w_1^{ND} is higher). Since the optimal amount of work at time 1, w_1^{ND}, is increasing in the expected total work, which in turn is increasing in β and decreasing in ϕ as was explained above, the indirect effect on w_1^{ND} through beliefs is positive for β and negative for ϕ. Clearly then, the overall effect of a higher conventional discount factor β on w_1^{ND} is positive, while the overall effect of a higher anticipatory discount factor ϕ is ambiguous. The indirect effect dominates, hence w_1^{ND} decreases with ϕ, for ϕ small relative to β, while the reverse is true for ϕ large relative to β. The effect of β and ϕ on the optimal deadline, ψ^{D}, is similar to that on optimal work without a deadline, w_1^{ND}, except that for ϕ we have the unambiguous result $\frac{\mathrm{d}\psi^{\mathrm{D}}}{\mathrm{d}\phi} < 0$, because a binding deadline is only chosen for high values of β relative to ϕ ($\beta > \overline{\beta}(\phi) > \phi$), where the negative indirect effect through beliefs dominates.

5 A model with present-biased preferences

In this section, we develop the counterpart of our model in the planning context, under the assumption of rational beliefs and present-biased preferences. We follow Phelps and Pollak (1968) and Laibson (1997) in assuming that this present bias takes the form of quasi-hyperbolic discounting. In Section 5.1 we present the model and the optimal decisions — the work and deadline choice — and in Section 5.2 we present comparative statics of the optimal decisions with respect to the discount factors.

5.1 Optimal actions

For the model with quasi-hyperbolic discounting, we assume beliefs are objective and that the discounted terms in U and V are all multiplied by the present bias parameter $0 \leq \xi \leq 1$, which creates a present-bias effect. So Equations 2 and 3 become:

$$V_t := \mathbb{E}_t \left[U_t + \xi \sum_{\tau > t} \beta^{\tau - t} U_\tau \right] \tag{7}$$

$$U_t := \mathbb{E}_t \left[u\left(w_t\right) + \xi \sum_{\tau > t} \phi^{\tau - t} u\left(w_\tau\right) \right], \tag{8}$$

where, as was explained in Section 1.2, β is the conventional discount factor that captures the degree to which an individual cares about his future selves, and ϕ is the anticipatory discount factor that captures the degree to which he gains immediate utility from anticipating future consumption. While models in the literature on quasi-hyperbolic discounting do not consider anticipatory utility, we incorporate it here to make the comparison with our model of time-inconsistent beliefs more direct. But setting the anticipatory discount factor ϕ to 0 to eliminate anticipatory utility does not affect any of our qualitative results below (except those involving variations in ϕ).

The following proposition determines the optimal decisions in the presence of present-bias.

Proposition 8. *(Optimal decisions with rational beliefs and quasi-hyperbolic discounting)*
(i) The optimal work is $w_1^{\mathrm{H}} := \min\left\{1, \tilde{B}\left(1 + \mathbb{E}\left[\eta\right]\right)\right\}$, *where* $\tilde{B} := \frac{\xi(\beta+\phi)}{1+\xi(\beta+\phi)}$.
(ii) The optimal deadline is $\psi^{\mathrm{H}} := \min\left\{1, \tilde{D}(1+\mathbb{E}\left[\eta\right])\right\}$, *where* $\tilde{D} := \frac{\xi\beta\phi+\beta^2+\phi^2}{\xi\beta\phi+\beta^2+\phi^2+\beta+\phi}$.
(iii) If $\xi < \overline{\xi}\left(\beta, \phi\right) := \frac{\beta^2+\phi^2}{\beta^2+\phi^2+\beta\phi}$, *deadline* ψ^{H} *is chosen* $\forall \mathbb{E}\left[\eta\right]$, *otherwise it is never chosen.*

Proposition 8 on deadline choice with time-inconsistent preferences contrasts Proposition 4 on deadline choice with time-inconsistent beliefs. With time-inconsistent preferences, the agent self-imposes a binding deadline at time 0 either for all or for no values of the objectively expected amount of total work, $1 + E$. That is, irrespective of the amount of work expected, either the present bias due to ξ dominates (when, for a given ξ, β and ϕ are sufficiently similar, as explained below) hence the agent chooses a deadline, or the future bias due to anticipatory utility dominates hence the agent does not choose a deadline. With time-inconsistent beliefs, on the other hand, the agent self-imposes a binding deadline for some but not all values of $1+E$, because optimal beliefs are optimistic and the cost of suboptimal smoothing due to optimistic beliefs increases with $1+E$. In addition, in this case the agent is more likely to choose a binding deadline at time 0, the larger the conventional discount factor β is relative to the anticipatory discount factor ϕ (see Section 3 for the intuition).

The optimal work at time 1 from the time-1 perspective is w_1^{H}, and the optimal deadline — equivalently the optimal work at time 1 from the time-0 perspective — is ψ^{H}. Without anticipatory utility ($\phi = 0$), we obtain the usual result in models of quasi-hyperbolic discounting, that the present bias introduced by ξ biases the time-1 optimal decision about work towards lower work w_1^{H}.[26] As a result, the optimal work at time 1 is smaller from the time-1 perspective than from the time-0 perspective, and a sophisticated agent — i.e., one who knows and correctly takes into account his preferences and beliefs in the present and in the future — optimally chooses a binding deadline at time 0. The agent exhibits this preference for commitment for any amount of present bias $\xi < 1$.

In the presence of anticipatory utility ($0 < \phi \leq 1$), ξ introduces an additional type of present bias, since it tilts utility at time t towards consumption utility and away from anticipatory utility from future consumption. To see this clearly, we combine Equations 7 and 8 to write

$$\begin{aligned}
V_0 &= \mathbb{E}\Big[\underbrace{U_0}_{\xi\phi u(w_1)+\xi\phi^2 u(w_2)} + \xi\beta \underbrace{U_1}_{u(w_1)+\xi\phi u(w_2)} + \xi\beta^2 \underbrace{U_2}_{u(w_2)} \Big] \\
V_1 &= \mathbb{E}\Big[\underbrace{U_1}_{u(w_1)+\xi\phi u(w_2)} + \xi\beta \underbrace{U_2}_{u(w_2)} \Big].
\end{aligned} \tag{9}$$

We see that this additional type of present bias tilts optimal decisions both at time 0 and at time 1 towards lower work w_1, but the effect is stronger for time 1, because U_1 is a bigger component of V_1 than of V_0; so this additional type of present bias works in the same direction as the usual type of present bias. So the reason that, with anticipatory utility, the deadline is only optimally chosen for a strong enough present bias ($\xi < \overline{\xi}(\beta, \phi)$) is that, as we have already explained in

[26] In our setting, the present bias has no effect on the time-0 optimal decision about the deadline, ψ^{H}, because there is no consumption at time 0.

Section 1.2, introducing anticipatory utility in the model introduces a future bias, and the present bias introduced by ξ needs to be strong enough (ξ needs to be small enough) to overcome it. The amount of present bias necessary to induce this preference for commitment depends on the values of the conventional discount factor β and the anticipatory discount factor ϕ. With a strong enough present bias ($\xi \leq \frac{2}{3}$), a binding deadline is chosen for all values of β, ϕ; otherwise, at each level of present bias $\xi > \frac{2}{3}$ it is only chosen if the anticipatory discount factor ϕ is sufficiently small relative to the conventional discount factor β, or vice versa. This is because the future bias introduced by anticipatory utility tilts optimal decisions both at time 0 and at time 1 towards higher work w_1, and the effect is stronger for time 1, but less so if ϕ is small relative to β (or vice versa); as a result, in this case a small amount of present bias due to $\xi < 1$ is sufficient to overcome the distortion in optimal decisions caused by the future bias. Figure 5 shows the set of values for β, ϕ for which a deadline is chosen for a range of values $\xi > \frac{2}{3}$.

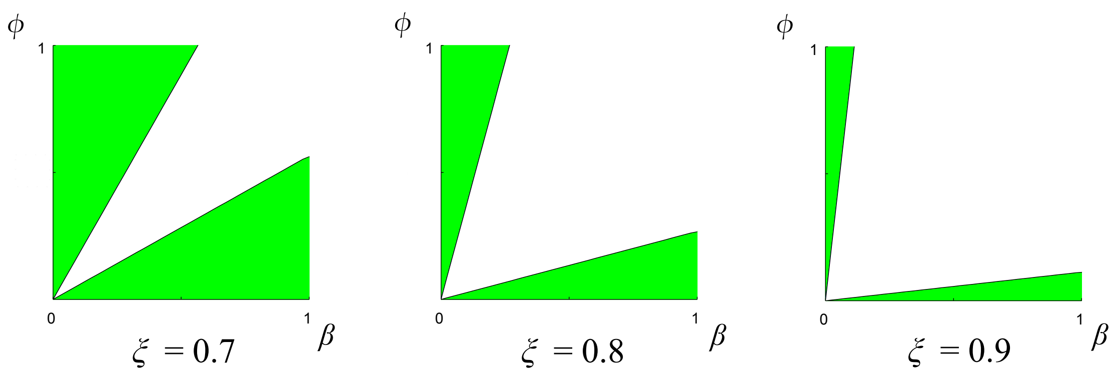


Figure 5: This figure illustrates how, in a model with rational beliefs and quasi-hyperbolic discounting, the set of values of the conventional discount factor β and the anticipatory discount factor ϕ for which a deadline is optimally chosen varies with values of the present-bias parameter, ξ, in the interval $\left(\frac{2}{3}, 1\right)$. In each panel, the two straight lines in black plot the degenerate hyperbola $\xi = \overline{\xi}(\beta, \phi)$ on the $\beta - \phi$ plane; $\xi < \overline{\xi}(\beta, \phi)$, hence the deadline is chosen, in the green-shaded area.

5.2 Comparative statics

Here, we present comparative statics in the model with rational beliefs and quasi-hyperbolic discounting. In short, i) the optimal work w_1^{H} increases both with the conventional discount factor β and with the anticipatory discount factor ϕ, which contrasts our earlier result that the optimal work w_1^{ND} with optimal beliefs increases (first decreases and then increases) as β (ϕ) increases; and ii) the optimal deadline ψ^{H} increases (decreases) with β and ϕ if ξ is large (small), which contrasts our earlier result that the optimal deadline ψ^{D} with optimal beliefs increases (decreases) with β (ϕ).

Proposition 9. *(Comparative statics with rational beliefs and quasi-hyperbolic discounting)*
(i) If present bias is weak enough ($\xi \geq \overline{\xi}(\beta, \phi)$) so a deadline is not chosen at $t = 0$, optimal work w_1^{H} increases with the conventional discount factor β and the anticipatory discount factor ϕ: $\frac{\mathrm{d}w_1^{\mathrm{H}}}{\mathrm{d}\beta} \geq 0$, $\frac{\mathrm{d}w_1^{\mathrm{H}}}{\mathrm{d}\phi} \geq 0$.
(ii) If present bias is strong enough ($\xi < \overline{\xi}(\beta, \phi)$) so a deadline is chosen at $t = 0$, the optimal deadline ψ^{H}:

- *increases (decreases) as β — the degree to which agents care about their future selves — increases, if ξ is large (small): $\frac{\mathrm{d}\psi^{\mathrm{H}}}{\mathrm{d}\beta} > 0 \Leftrightarrow \xi > \frac{\phi^2 - 2\beta\phi - \beta^2}{\phi^2}$,*
- *increases (decreases) as ϕ — the degree to which agents get utility from anticipation — increases, if ξ is large (small): $\frac{\mathrm{d}\psi^{\mathrm{H}}}{\mathrm{d}\phi} > 0 \Leftrightarrow \xi > \frac{\beta^2 - 2\beta\phi - \phi^2}{\beta^2}$.*

To interpret these results, we refer to Equation 9. For weak present bias ($\xi \geq \overline{\xi}$), the agent at time 0 does not choose a deadline, so the agent at time 1 chooses the optimal work to maximize V_1, the subjectively expected discounted sum of utility from time 1 onwards. At time 1, the agent discounts future consumption utility, $u(w_2)$, by factor ϕ because he gains immediate utility from anticipation and by factor β because he cares about his future self. This future consumption utility is increasing in the amount of work at time 1, so the higher the conventional discount factor β is and the higher the anticipatory discount factor ϕ is, the higher the optimal work w_1.

For strong present bias ($\xi < \overline{\xi}$), the agent at time 0 self-imposes the deadline ψ^{H} that maximizes V_0, the subjectively expected discounted sum of utility from time 0 onwards. Without anticipatory utility ($\phi = 0$), the higher the conventional discount factor β is, the higher the optimal deadline. This is simply because, in this case, the present bias introduced by ξ has no effect on the optimal decision at time 0 about the deadline, and since an agent with a higher β cares more about his future selves, he will impose a higher deadline so he will need to work less in the future. With anticipatory utility ($\phi > 0$), the conventional and the anticipatory discount factors β and ϕ have an ambiguous effect on the optimal deadline ψ^{H}.[27]

[27] For a strong enough present bias (small enough ξ), utility at time 1, U_1, is tilted sufficiently towards consumption utility from work at time 1 and away from anticipatory utility from work at time 2, therefore a higher conventional discount factor β has a first-order effect on how much the agent at time 0 cares about consumption utility at time 1 and only a second-order effect on how much he cares about consumption utility at time 2; so if β is also small, increasing it has an overall negative effect on the optimal deadline. If present bias is weaker (ξ is larger) or the conventional

6 Summary of theoretical implications and relevant evidence

With a view towards empirical testing, in this section we summarize the implications of our theory of optimal time-inconsistent beliefs, and contrast them with those from models with present-biased preferences. We relate some of these implications to existing experiments from the psychology literature on the planning fallacy, and argue that our theory is mostly consistent with the existing evidence on beliefs and behavior, while a model with present-biased preferences is only consistent with a fraction of the evidence. We note that, though we cannot always map our theoretical setup precisely to the experimental setup of existing studies, it is usually possible to accommodate differences through re-interpretation or through straightforward extensions of our model; we point these out, wherever relevant, in our subsequent discussion. All the predictions and relevant evidence discussed below are succinctly presented in Table 1.

6.1 Beliefs — implications

First, we consider implications regarding beliefs. In our setup, the amount of work or time necessary to complete the task is $1 + \eta$ (or $w_1 + w_2$). According to Propositions 2, 3, and 4, our model predicts that beliefs are optimistic and overconfident. It also predicts that beliefs become more pessimistic over time if a deadline is not chosen, and more optimistic if a deadline is chosen. According to Proposition 5, our model predicts that beliefs are more optimistic with an externally-imposed than with a self-imposed deadline. From Proposition 6, our model predicts that individuals with higher conventional discount factor β (higher anticipatory discount factor ϕ) hold more pessimistic (more optimistic) beliefs.

Finally, it is straightforward to extend our model to incorporate various incentives, which are sometimes used in experimental settings; we do so for two situations, in Section B.3 of the online appendix. First, we consider an incentive for the speed of task completion, modeled as a payment at time 2 that is decreasing in total work, $1+\eta$. In this case, consumption utility is derived not only from the disutility of work, but also from the subjective expected payment; the latter generates an incentive to believe that the total amount of work is low, hence it is optimal to be even more optimistic and suffer from the planning fallacy even more than without the payment. Second, we

discount factor β is larger, the opposite is true. The effect of the anticipatory discount factor ϕ on the optimal deadline, ψ^{H}, is also ambiguous. For a strong enough present bias (small enough ξ), anticipatory utility from work at time 1 is sufficiently more important than that from work at time 2, hence the less anticipatory utility is discounted (the higher ϕ is) the smaller the optimal deadline ψ^{H}; the opposite is true if present bias is not strong enough.

Table 1: Predictions and Evidence

Predictions and relevant evidence, for models with time-inconsistent beliefs and time-inconsistent preferences. Text in red bold (blue italics) typeface indicates predictions that are different (unless additional conditions hold, e.g., on partial sophistication or parameter values), while text in black regular typeface indicates predictions that are the same.

	Optimal Time-inconsistent Beliefs	*Time-inconsistent Preferences*	*Evidence*
Beliefs about work/time needed for task			
Beliefs are rational/optimistic.	**Optimistic**	**Rational**	Optimistic
Without deadline, beliefs become more/less optimistic as temporal distance to task decreases.	**Less**	**No effect**	Less
With deadline, beliefs become more/less optimistic as temporal distance to task decreases.	**More**	**No effect**	
Beliefs are more optimistic with an externally-imposed deadline than with self-imposed deadline.	**More**	**No difference**	
Beliefs become more/less optimistic with higher β.	**Less**	**No effect**	
Beliefs become more/less optimistic with higher ϕ.	**More**	**No effect**	
Effect of incentive for accuracy of prediction.	**Reduces optimism**	**No effect**	Reduces optimism
Effect of incentive for speed of completion.	**Increases optimism**	**No effect**	Increases optimism
Optimal work (without a deadline)			
Work smoothing is suboptimal.	Yes	Yes	Yes (anecdotal)
w_1 increases/decreases with β.	Increases	Increases	
w_1 increases/decreases with ϕ.	*Decreases if $\phi < \beta$.* *Decreases if $\beta < \phi < \overline{\phi}$ and E$[\eta]$ large.* *Increases if $\overline{\phi} < \phi$.*	*Increases*	
Effect of incentive for accuracy of prediction.	**Reduces w_1**	**No effect**	
Effect of incentive for speed of completion.	**Increases w_1**	**No effect**	
Commitment			
Individuals use commitment devices.	Yes	Yes, unless naïve.	Yes
Propensity for commitment depends on E$[\eta]$.	*Yes*	*No, unless commitment is costly.*	
Propensity for commitment increases/decreases as β increases relative to ϕ.	*Increases*	*Increases if $\beta > \phi$.* *Decreases if $\beta < \phi$.*	
Optimal deadline increases/decreases with β.	*Increases*	*Decreases if ξ is small and β is small relative to ϕ.* *Increases if ξ is large or β is large relative to ϕ.*	
Optimal deadline increases/decreases with ϕ.	*Decreases*	*Decreases if ξ is small.* *Increases if ξ is large.*	
External deadlines improve smoothing and performance more than self-imposed ones.	*Yes*	*No, unless partially-sophisticated.*	Yes

consider an incentive for the accuracy of task duration prediction, modeled as a payment at $t = 2$ that is decreasing in the (absolute) difference between objective and subjective expectations about $1 + \eta$. Again, consumption utility is derived not only from the disutility of work, but also from the payment, which acts as an additional, explicit, cost of belief distortion, hence it is optimal to be less optimistic and suffer from the planning fallacy less than without the payment.[28]

These predictions stand in contrast to the model with present-biased preferences in which beliefs are assumed to be objective. That said, some models with present-biased preferences also assume that beliefs about future preferences are (exogenously) naive or partially sophisticated. And these models can match the evidence that beliefs about task duration are optimistic, that they become more pessimistic over time (without a deadline), and that they become more optimistic given an incentive for the speed of task completion.[29] What these models cannot match is evidence on duration prediction for tasks in which task duration is exogenous, as we describe subsequently. Our model in which beliefs respond (endogenously) to the anticipatory benefits and average ex post costs matches evidence on duration under-prediction, while models with naive or partially sophisticated agents with present-biased preferences do not. Obviously, the predictions of these different approaches will differ in other environments also.

6.2 Beliefs — evidence

There is a wealth of robust experimental evidence demonstrating that people underestimate task completion time. Roy et al. (2005) survey 14 papers on this planning fallacy, and report that 12 papers find that people hold optimistic beliefs about task duration in all tasks studied, 1 paper finds optimistic beliefs in some tasks, and 1 paper does not find optimistic beliefs. For illustration purposes, one of the first studies is by Buehler et al. (1994): individuals were asked to predict the completion time for an academic and a non-academic task they intended to complete in the coming week, and subsequently actual completion times were collected. For academic (non-academic)

[28] We note that all predictions mentioned in this paragraph have been phrased in terms of beliefs about total work/time $1 + \eta$, but they can also be phrased identically in terms of beliefs about work w_2 in the second period, or in terms of the planning fallacy.

[29] For example, consider an individual who believes that a project will take 20 hours of work, and initially plans to complete it in 2 days, working 10 hours a day. If he is naive, he will not anticipate any deviations from this plan, hence will predict that he will complete the task in 2 days. But if he has time-inconsistent preferences, he will in fact procrastinate and work, e.g., 5 hours/day for 4 days. An experimenter would compare the initial prediction with the time from the moment the task started to the moment it ended, and would conclude that the individual held optimistic beliefs, even though he held objective beliefs about the actual task duration. If there was an incentive for speed, the individual would procrastinate less, hence his beliefs would seem less optimistic.

tasks, the average predicted time was 5.8 (5.0) days and the actual completion time was 10.7 (9.2) days, and only 37 (43) percent of subjects completed the tasks in the predicted time, though 74 (70) percent of them reported they were certain they would do so. While this evidence is consistent with our model's prediction of optimistic and overconfident beliefs about actual task duration, as we noted above, it can also be explained using a model with present-biased preferences, since actual task duration is conflated with the time from the moment the task starts to the moment that it ends. But other studies on the planning fallacy allow us to distinguish the two theories more effectively, since they involve tasks with exogenous duration, for which duration under-prediction cannot be explained by procrastination in a model with time-inconsistent preferences. For example, in Konečni and Ebbesen (1976), people waiting in line at a gas station during the 1974 oil crisis were asked to estimate how long they would have to wait; the actual wait time was 29 minutes, but the estimated wait time was only 19 minutes. In this case, under-prediction could be explained by the anticipatory benefits of optimism, but not by naivete regarding future preferences.[30,31]

There is also some experimental evidence that is consistent with our model's prediction that beliefs become less optimistic over time. In particular, Gilovich et al. (1993) show that people's average beliefs about exam performance become less optimistic as temporal distance from the exam decreases.[32] This evidence does not strongly distinguish our model from one with present-biased preferences and partially-sophisticated (or naive) agents, since it involves a task with intervening events, but still it is reassuring. But more generally, there is substantial evidence of non-Bayesian belief updating, which our theory assumes is possible and shows is optimal. A classic study on inconsistent belief updating is the lawyer-engineer experiment by Kahneman and Tversky (1973) that started the large literature on the base rate fallacy. Also see Camerer (1995) for a survey of other experimental evidence on non-Bayesian updating.

Finally, we discuss evidence that is consistent with our model's prediction that beliefs about

[30] In the baseline interpretation of our setup, utility is derived from working, but it could also be derived from waiting (as in this experimental setup), or from any activity that affects one's happiness.

[31] Roy et al. (2005) survey 6 papers on the planning fallacy involving tasks that lasted less than an hour. In 4 out of 6 papers, subjects exhibited the planning fallacy in all tasks, in 1 paper they exhibited the planning fallacy in some tasks, and in 1 paper they did not exhibit the planning fallacy. Taking into account that, as the authors note, most studies involved manipulations aimed at reducing the planning fallacy, this evidence is quite strong in favor of the planning fallacy even for shorter tasks. We note that in these tasks there is no room for suboptimal smoothing, so strictly speaking our model predicts complete optimism as it has no cost. Clearly, this is an extreme result that follows from our focus on incentives, hence our simplifying assumption that there are no explicit, e.g., psychological, costs of belief distortion.

[32] While we do not talk explicitly about task performance in our model setup, it is reasonable to assume that individuals with optimistic beliefs about the total work necessary for the task are also optimistic about their performance on the task, since they perceive it to be shorter/easier.

actual task duration can be manipulated using monetary incentives that generate an anticipatory benefit or cost for optimism. First, Byram (1997) and Buehler et al. (1997) report the results of experiments in which subjects are randomly assigned to a treatment giving them payment for rapid completion of the task. In Byram (1997), subjects in a treatment group were given explicit incentives for rapid completion prior to making their predictions for the duration of an origami folding task: they were paid \$4 / \$2 / \$1 / \$0 for finishing in the first / second / third / fourth quartile, while the control group was paid \$3. For the control group, the median prediction (actual) time was 7.8 (8.8) minutes; for the treatment group, the median prediction (actual) time was 5.0 (7.8) minutes. So the incentive for speed raised the prediction error by 180% and decreased the actual time to completion by 11%. Buehler et al. (1997) study an experiment in which subjects complete anagrams and, like Byram (1997), find that a treatment group that is given incentives for speed exhibits much greater bias in prediction and slightly more rapid completion.

Second, Buehler et al. (1997) report the results of an experiment involving the task of solving anagrams, in which a treatment group is given a payment for accurate prediction of task completion times. A random subsample of subjects was given an incentive for accuracy: \$2 (\$4) if the predicted completion time was within 1 minute (30 seconds) of the actual time. In addition, all subjects were given an incentive for speed, as described previously; presumably this was done to discourage subjects from intentionally making longer predictions and simply slowing down, if necessary, to match them. For subjects without the incentive for accurate prediction, the mean predicted (actual) completion time was 4.1 (5.5) minutes, and for the treatment group with the incentive for accurate prediction, the mean predicted (actual) time was 5.8 (5.5) minutes, hence optimism was eliminated, consistent with our theory's prediction.

6.3 Behavior — implications

Now, we consider implications in terms of behavior. According to Proposition 1, our model predicts that work smoothing is suboptimal, and according to Proposition 4 it predicts that individuals may use commitment devices. Such behavior is, of course, also predicted by models with present-biased preferences. According to Proposition 5, our model predicts that externally-imposed deadlines are stricter, hence improve work smoothing more than self-imposed deadlines; this is also predicted by models with present-biased preferences, although only under the additional assumption of partial sophistication. A prediction that is unique to our model is that, since beliefs affect optimal actions, incentives that affect optimal beliefs also affect optimal actions that depend on these beliefs; this is not the case in models with time-inconsistent preferences where beliefs are assumed to be rational. For example, we have already discussed that an incentive for the speed of task completion makes beliefs more optimistic and an incentive for accurate predictions makes beliefs more pessimistic. Therefore, our model predicts that incentives for speed make the optimal work w_1 in the first period smaller, and incentives for accuracy make w_1 larger. Finally, as we can see from Propositions 6, 7, and 9, our model predicts that the optimal amount of work w_1 without a deadline, the optimal deadline ψ, and the propensity to choose a deadline vary with the structural model parameters in different ways than in a model with present-biased preferences. Admittedly, using these predictions for testing is not straightforward, as it involves the estimation of individuals' time preference parameters.

6.4 Behavior — evidence

There is substantial anecdotal evidence that people's plans are suboptimal, and informal as well as formal evidence that people choose to constrain their future behavior. Informally, people use institutional arrangements like weight-loss camps, alcohol clinics, and Christmas clubs as commitment devices. Formally, Trope and Fishbach (2000), DellaVigna and Malmendier (2006), Thaler and Benartzi (2004) and Ashraf et al. (2006) document preference for commitment.[33] Bryan et al. (2010) review the growing evidence on the widespread use of commitment devices.

Most directly related to our model are the experiments on deadlines in Ariely and Wertenbroch (2002). The authors conducted two separate studies: one in which subjects had to write three academic papers during the course of the term, and one in which subjects had to proofread three texts in the course of three weeks. Subjects were randomly assigned into groups: one that faced externally-imposed, equally-spaced deadlines; one in which subjects could self-impose deadlines; and one that had no deadline option (this treatment only existed in the proofreading study). This experimental design resembles but is not identical to our theoretical setup. First, in our model uncertainty is resolved only at one point in time, while in the experiments it is possibly resolved

[33] Trope and Fishbach (2000) find that participants self-impose penalties for neglecting to undergo minor medical procedures. DellaVigna and Malmendier (2006) document that a sample of gym users would have saved almost 70% if they had bought daily passes instead of gym memberships. In a field experiment, Thaler and Benartzi (2004) introduce a commitment device that enables people to commit to a retirement savings plan, and find that 65% to 80% of people choose to participate. Also in a field experiment on savings behavior, Ashraf et al. (2006) find that 28% of the subjects choose to deposit money in savings accounts with withdrawal restrictions.

over time, since the amount of work necessary to complete the first sub-task may be informative about the amount necessary to complete the other two. Second, in our model a deadline imposes the minimum amount of work to be completed in period 1, while in the experiments the deadline specifies the date by which one has to complete one of the papers, so the deadline requires the completion of a proportion rather than an absolute amount of work. Finally, in our model the deadline cannot be violated, while in the experiments performance penalties were imposed on subjects who missed their deadlines. We believe that these differences in the environment do not crucially affect our model's implications, hence the possibility to verify them using this experiment.[34]

Regarding deadlines, the authors found that most people with the option to choose a deadline did so. For the paper assignments, the mean deadlines chosen were 42/26/10 days before the end of the semester for the first/second/third paper, and 73% of deadlines were before the last week of class. In addition, the mean completion times of the group without deadlines were significantly later than the mean self-imposed deadlines, which were significantly later than the externally-imposed equally-spaced deadlines. This pattern mirrors our results in Propositions 4 and 5. In terms of performance, Ariely and Wertenbroch (2002) found that subjects given an externally-imposed equally-spaced deadline performed (as measured by the course grade in the first study and by the number of errors detected in the second study) better than subjects who self-imposed a deadline. Again, this pattern mirrors our results in Propositions 4 and 5, to the extent that it is reasonable to assume that smoother work profiles produce better performance.[35]

Finally, we revisit the experiments on the effects of monetary incentives on beliefs and behavior presented in Byram (1997) and in Buehler et al. (1997). While these studies verify our theory's prediction that monetary incentives affect beliefs, unfortunately the experimental design they utilize does not allow us to test our theory's additional prediction that these monetary incentives also affect *behavior* that depends on these beliefs. We first explain why this is the case, and then we propose a refined design that could provide a powerful test of our theory based on this additional prediction.

[34] Indeed, in an earlier version of the paper, our model setup allowed for two shocks, η_1 and η_2, with the first shock being informative about the second, and the deadline was of the form $\psi\eta_1$, i.e., it required a proportion rather than an absolute amount of work to be completed in the first period. Our results were qualitatively identical to our current results. In our current model, we have essentially set $\eta_1 = 1$ (hence, the deadline is ψ, i.e., an absolute amount); this has enabled us to simplify the exposition and to more clearly state our results and the intuition behind them.

[35] A concern with this result might be that improved performance on the task at hand might have come at the cost of reduced performance on other tasks that also required time as a resource. Ariely and Wertenbroch (2002) provide evidence against this: they found that on a final project which was unrelated to the papers, students with externally-imposed deadlines in fact performed better (mean score of 86) than students with self-imposed or no deadlines (mean score of 77).

In Buehler et al. (1997), subjects in the control group were given a payment for the rapid completion of anagram tasks, while subjects in the treatment group were also given a payment for the accurate prediction of task completion times. According to our theory, the payment for accurate task duration prediction generates anticipatory utility that is decreasing in the (absolute) difference between the objective and subjective expected task duration, which makes optimal beliefs less optimistic; as already discussed in Section 6.2, this prediction is consistent with the experimental results. Additionally, according to our theory, beliefs about task duration affect the optimal split of work between the two periods, w_1 and w_2, but not the total amount of work, $w_1 + w_2$, hence the task duration; this is consistent with the experimental result that the mean actual task duration for both the control and the treatment group was an identical 5.5 minutes. That is, in this experiment, the monetary incentive affects beliefs but not observed behavior, because observed behavior — the actual task duration — does not depend on beliefs.

In Byram (1997), subjects in a treatment group were given a payment for the rapid completion of an origami folding task. According to our theory, this payment generates anticipatory utility that is decreasing in the subjective expected task duration, which makes optimal beliefs more optimistic; as already discussed in Section 6.2, this prediction is consistent with the experimental results. As before, according to our theory, beliefs about task duration affect the optimal split of work between the two periods, but they do not affect the total amount of work, hence the task duration; this is consistent with the experimental result that the median actual task duration for the control group was 8.8 minutes while for the treatment group with the monetary incentive it was a statistically insignificantly smaller 7.8 minutes.[36]

Now, we propose a refined experimental design that could test our theory's prediction that monetary incentives that affect beliefs also affect behavior that depends on these beliefs. For a clean test of our prediction, we need a study in which i) observed behavior depends on beliefs, and ii) monetary incentives affect observed behavior only indirectly, through beliefs. The aforementioned study on the effect of a payment for accurate task duration prediction presented by Buehler et al. (1997) satisfies the second condition, since the payment for rapid completion given to all subjects ensures that they will not strategically vary the rate of work, but it does not satisfy the first condi-

[36]We note that the payment for rapid completion may affect task duration directly by affecting the rate or intensity of work. This could explain why, in a similar experimental design in a study in which subjects complete anagrams, Buehler et al. (1997) find that the mean actual task duration was a statistically significant 21% smaller (5.5 minutes versus 7.0 minutes) for the treatment group with the payment for rapid completion. In our model, we make the simplifying assumption that the agent can control the amount of work, but not the rate or intensity of work, therefore our model does not predict this direct effect that the payment of rapid completion may have on task duration.

tion, since the observed behavior — actual completion time — does not depend on beliefs. Instead, subjects could be asked to choose for how long they would like to work before a break, and then to complete the task after the break; this design closely resembles our model setup, in which our theory predicts that an incentive for accurate duration prediction generates an incentive to hold less optimistic beliefs, hence increases the optimal amount of work in the first period (before the break). On the other hand, the aforementioned studies on the effect of a payment for rapid task completion presented by Buehler et al. (1997) and by Byram (1997) satisfy neither of the stated conditions for a clean test of our prediction. To satisfy the condition that observed behavior depends on beliefs, we could modify the experimental design as suggested above for the case of a payment for accurate duration prediction, while to satisfy the condition that monetary incentives affect observed behavior only through beliefs, we would need to choose a task for which subjects cannot vary the rate of work; a task whose duration is exogenous, e.g., playing a piano piece at the prescribed tempo, would satisfy this requirement.[37]

7 Concluding discussion

In this paper, we develop a structural theory of optimal beliefs that relaxes the assumption of time-consistency in beliefs. We do not impose time-inconsistency, rather we show that time-inconsistent beliefs are optimal and are an endogenous cause of the preference for commitment. We present our theory in the context of planning and show that, as in the original description of the planning fallacy by Kahneman and Tversky (1979), people tend to postpone work because they hold overoptimistic beliefs about the ease of the task. The strength of our approach is that these belief biases are situational, and so our model makes predictions about when optimism and the planning fallacy are mitigated or exacerbated.

In our model, as in much recent work in behavioral economics, biases in beliefs are central to the understanding of behavior, so our theory can be criticized as a step away from the discipline of rationality that mainstream economics imposes on itself. This discipline is used to select among models that can all explain observed choice behavior, and rationality as the preferred assumption has its appeal in many contexts. But the appeal of structural models is that they are useful out-of-

[37]For example, a suitable study could be as follows. Musicians would be given a list of familiar musical pieces which they would need to perform at a prescribed tempo (beats per minute). They would first predict how long it would take them to perform all the pieces, then they would choose for how long they would want to perform before taking a break, and finally, they would play the pieces. Different groups of subjects would be given different monetary incentives for speed of task completion and accuracy of task duration prediction.

sample, and a parsimonious model that better represents actual beliefs and preferences is likely to perform better in such an exercise.

Thus, we replace the discipline of the rationality assumption with the discipline of data by verifying that our model's predictions on beliefs match existing experimental evidence. In doing so, we provide an example of how experimental methods and reported expectations can be used to test and evaluate theoretical models that fall under the broad heading of behavioral economics. In particular we observe causation from environment and incentives to reported beliefs, that is consistent with our model and inconsistent with objective probability assessments. In sum, the model is consistent with much existing experimental evidence on mis-planning and on the use and effects of deadlines. The next step would be to further test our model, by designing experiments with our model's specific implications in mind. Another possible step would be to apply our theory in different contexts (e.g., the real options framework of investment under uncertainty) and study how its implications differ from those from a model with time-inconsistent preferences.

References

Akerlof, G. 1991. "Procrastination and obedience." *American Economic Review P&P*, 81(1): 1–19.

Akerlof, G, and W. T Dickens. 1982. "The economic consequences of cognitive dissonance." *American Economic Review*, 72(3): 307–319.

Ariely, D, and K Wertenbroch. 2002. "Procrastination, deadlines, and performance: Self-control by precommitment." *Psychological Science*, 13(3): 219–24.

Armor, D. A, and S. E Taylor. 1998. "Situated optimism: Specific outcome expectancies and self-regulation." *Advances in Experimental Social Psychology*, 30: 309–79.

Ashraf, N, D Karlan, and W Yin. 2006. "Tying odysseus to the mast: Evidence from a commitment savings product in the Philipinnes." *Quarterly Journal of Economics*, 121(2): 635–672.

Bénabou, R, and J Tirole. 2002. "Self-confidence and personal motivation." *Quarterly Journal of Economics*, 117(3): 871–915.

Bernheim, B. D, and R Thomadsen. 2005. "Memory and anticipation." *Economic Journal*, 117(3): 271–304.

Brunnermeier, M. K, and J. A Parker. 2005. "Optimal expectations." *American Economic Review*, 95(4): 1092–1118.

Bryan, G, D Karlan, and S Nelson. 2010. "Commitment devices." *Annual Review of Economics*, 2: 671–698.

Buehler, R, D Griffin, and H MacDonald. 1997. "The role of motivated reasoning in task predictions." *Personality and Social Psychology Bulletin*, 23(3): 238–47.

Buehler, R, D Griffin, and M Ross. 1994. "Exploring the planning fallacy: Why people underestimate their task completion times." *Journal of Personality and Social Psychology*, 67(3): 366–81.

Byram, S. J. 1997. "Cognitive and motivational factors influencing time prediction." *Journal of Experimental Psychology: Applied*, 3(3): 216–39.

Camerer, C. 1995. "Individual decision making." In *Handbook of experimental economics*, ed. John Kagel and Alvin Roth, 587–703. Princeton: Princeton University Press.

Caplin, A. J, and J Leahy. 2000. "The social discount rate." NBER Working Paper.

Caplin, A. J, and J Leahy. 2001. "Psychological expected utility theory and anticipatory feelings." *Quarterly Journal of Economics*, 106(1): 55–80.

Caplin, A. J, and J Leahy. 2004. "The social discount rate." *Journal of Political Economy*, 112(6): 1257–68.

Dasgupta, P, and E Maskin. 2005. "Uncertainty and hyperbolic discounting." *American Economic Review*, 95(4): 1290–1299.

DellaVigna, S, and U Malmendier. 2006. "Paying not to go to the gym." *American Economic Review*, 96(3): 694–719.

Dominitz, J. 1998. "Earnings expectations, revisions, and realizations." *Review of Economics and Statistics*, 80(3): 374–88.

Epstein, L. G. 2006. "An axiomatic model of non-Bayesian updating." *Review of Economic Studies*, 73(2): 413–436.

Gilovich, T, M Kerr, and V. H Medvec. 1993. "Effect of temporal perspective on subjective confidence." *Journal of Personality and Social Psychology*, 64(4): 552–560.

Grenadier, S. R, and N Wang. 2007. "Investment under uncertainty and time-inconsistent preferences." *Journal of Financial Economics*, 84(1): 2–39.

Gul, F, and W Pesendorfer. 2001. "Temptation and self-Control." *Econometrica*, 69(6): 1403–35.

Hamermesh, D. S. 1985. "Expectations, life expectancy, and economic behavior." *Quarterly Journal of Economics*, 100(2): 389–408.

Kahneman, D, and A Tversky. 1973. "On the psychology of prediction." *Psychological Review*, 80(4): 237–251.

Kahneman, D, and A Tversky. 1979. "Intuitive prediction: Biases and corrective procedures." *TIMS Studies in Management Science 12*, 313–27.

Konečni, V. J, and E. B Ebbesen. 1976. "Distortions of estimates of numerousness and waiting time." *Journal of Social Psychology*, 100(1): 45–50.

Koszegi, B. 2006. "Ego utility, overconfidence, and task choice." *Journal of the European Economic Association*, 4(4): 673–707.

Koszegi, B, and M Rabin. 2009. "Reference-dependent consumption plans." *American Economic Review*, 99(3): 909–936.

Laibson, D. 1997. "Golden eggs and hyperbolic discounting." *Quarterly Journal of Economics*, 112(2): 443–477.

Landier, A. 2000. "Wishful thinking: A model of optimal reality denial." *Mimeo, MIT.*

Loewenstein, G. 1987. "Anticipation and the valuation of delayed consumption." *Economic Journal*, 47(387): 666–684.

Manski, C. F. 2004. "Measuring expectations." *Econometrica*, 72(5): 1329–76.

Nyarko, Y, and A Schotter. 2002. "An experimental study of belief learning using elicited beliefs." *Econometrica*, 70(3): 971–1005.

O'Donoghue, T, and M Rabin. 2001. "Choice and procrastination." *Quarterly Journal of Economics*, 116(1): 121–60.

Phelps, E. S, and R. A Pollak. 1968. "On second-best national saving and game-equilibrium growth." *Review of Economic Studies*, 35(2): 185–199.

Roy, M. M, N. J. S Christenfeld, and C. R. M McKenzie. 2005. "Underestimating the duration of future events: Memory incorrectly used or memory bias?" *Psychological Bulletin*, 131(5): 738–56.

Strotz, R. H. 1955. "Myopia and inconsistency in dynamic utility maximization." *Review of Economic Studies*, 23(3): 165–180.

Thaler, R. H, and S Benartzi. 2004. "Save more tomorrow: Using behavioral economics to increase employee saving." *Journal of Political Economy*, 112(1-2): S164–S187.

Trope, Y, and A Fishbach. 2000. "Counteractive self-control in overcoming temptation." *Journal of Personality and Social Psychology*, 79(4): 493–506.

Trope, Y, and N Liberman. 2003. "Temporal construal." *Psychological Review*, 110(3): 403–21.

Veenhoven, R. 2008. "Healthy happiness: Effects of happiness on physical health and consequences for preventive health care." *Journal of Happiness Studies*, 9: 449–469.

Yariv, L. 2002. "I'll see it when I believe it – A simple model of cognitive consistency." *Mimeo*.

Appendix

A Proofs of propositions

A.1 Proof of Proposition 1

The agent chooses w_1 at $t = 1$ to maximize V_1. Using Equations 1, 3, V_1 becomes

$$\hat{\mathbb{E}}_1 \left[u(w_1) + (\phi + \beta) u(1 + \eta - w_1) \right],$$

which is concave in w_1. Using $u(w) = -\frac{1}{2}w^2$, the F.O.C. yields $w_1^\dagger\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = B_1\left(1+\hat{E}_1\right)$, where $B_1 := \frac{\beta+\phi}{1+\beta+\phi}$, and imposing $w_1 \leq 1$ yields our result.

A.2 Proof of Proposition 2

(i - iii) First, we argue that optimal beliefs satisfy $B_1\left(1 + \hat{E}_1\right) \leq 1$: If not, from Equation 5, the optimal work at $t = 1$ would be 1, and optimal beliefs could become more optimistic, yielding anticipatory benefits without altering behavior, so without cost. Thus, we substitute $w_1 = B_1\left(1 + \hat{E}_1\right)$ and $w_2 = 1 + \eta - w_1$ into $\mathcal{W}$. Second, optimal time-0 beliefs are completely optimistic, because $\mathcal{W}$ is decreasing in $\hat{E}_0$ while w_1^* does not depend on $\hat{E}_0$, so optimistic time-0 beliefs have anticipatory benefits but no distortionary costs. Thus, we also substitute $\hat{E}_0 = 0$ into $\mathcal{W}$. Now we need to find the optimal $\hat{E}_1$.

Define the following constants:

$$\begin{aligned} F &:= \phi(1+\phi) + (1+\phi)\beta + \beta^2 \qquad & M_{B_1} &:= FB_1^2 + \beta\phi(1 - 2B_1) \\ G_B &:= -\phi^2 B_1, \end{aligned}$$

where $F > 0$ and $G_B < 0$ are obvious, and it is easy to show that $M_{B_1} > 0$. Then:

$$\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\hat{E}_1} = -G_B - M_{B_1}\left(1 + \hat{E}_1\right) + \beta^2 B_1 (1 + E).$$

Setting the derivative to 0, and letting $\underline{s}_1^{\text{ND}} := \beta^2 \frac{B_1}{M_{B_1}}$ and $\underline{c}_1^{\text{ND}} := \underline{s}_1^{\text{ND}} - \frac{G_B}{M_{B_1}} - 1$:

$$\hat{E}_1^\dagger = \underline{s}_1^{\text{ND}} E + \underline{c}_1^{\text{ND}}.$$

Imposing the constraints $\hat{E}_0 \geq 0$, $\hat{E}_1 \geq 0$, and $B_1\left(1 + \hat{E}_1\right) \leq 1$, we have

$$\begin{cases} \hat{\mathbb{E}}_1^{\text{ND}}[\eta] = 0 & \text{if} \quad \mathbb{E}[\eta] \leq \mu_L^{\text{ND}} \\ \hat{\mathbb{E}}_1^{\text{ND}}[\eta] = \underline{s}_1^{\text{ND}}\mathbb{E}[\eta] + \underline{c}_1^{\text{ND}} & \text{if} \quad \mu_L^{\text{ND}} < \mathbb{E}[\eta] \leq \mu_U^{\text{ND}} \\ \hat{\mathbb{E}}_1^{\text{ND}}[\eta] = \frac{1}{B_1} - 1 & \text{if} \quad \mu_U^{\text{ND}} < \mathbb{E}[\eta] \end{cases}, \tag{A.1}$$

where the critical values for $\mathbb{E}[\eta]$ are $\mu_L^{\mathrm{ND}} := \frac{G_B + M_{B_1}}{\beta^2 B_1} - 1$ and $\mu_U^{\mathrm{ND}} := \frac{G_B + \frac{M_{B_1}}{B_1}}{\beta^2 B_1} - 1$.

For optimism, we know $\hat{E}_0^{\mathrm{ND}} = 0 < E$, so we just need to show $\hat{E}_1^{\mathrm{ND}} < E$:

- For $E \leq \mu_L^{\mathrm{ND}}$, we have $\hat{E}_1^{\mathrm{ND}} = 0 < E$.
- For $E > \mu_U^{\mathrm{ND}}$, we have $\hat{E}_1^{\mathrm{ND}} = \frac{1}{B_1} - 1$, so $\hat{E}_1^{\mathrm{ND}} - E < \frac{1}{B_1} - 1 - \mu_U^{\mathrm{ND}} = -\phi \frac{\beta(1+\beta+\phi)+(\beta+\phi)^2}{\beta^2(\beta+\phi)^2} < 0$.
- For $\mu_L^{\mathrm{ND}} < E \leq \mu_U^{\mathrm{ND}}$, as a function of E, $\hat{E}_1^{\mathrm{ND}}$ is a straight line segment whose endpoints lie below the line E, so $\hat{E}_1^{\mathrm{ND}} < E$.

(iv) The planning fallacy is that $\hat{\mathbb{E}}_t^{\mathrm{ND}}\left[w_1^* + w_2^*\right] < \mathbb{E}\left[w_1^* + w_2^*\right]$, for $t = 0$ and $t = 1$. Using Equation 1, this simply becomes $\hat{\mathbb{E}}_t[\eta] < \mathbb{E}_t[\eta]$, which we have shown above to be true, since optimal beliefs are optimistic.

(v) Substituting, respectively, optimal beliefs from Equation A.1 and objective beliefs into Equation 5, we have $w_1^{\mathrm{ND}} = \min\left\{1, B_1\left(1 + \hat{E}_1^{\mathrm{ND}}\right)\right\}$ and $w_1^{\mathrm{RE}} = \min\{1, B_1(1+E)\}$. We showed above that $\hat{E}_1^{\mathrm{ND}} < E$, so $w_1^{\mathrm{ND}} \leq w_1^{\mathrm{RE}}$.

A.3 Proof of Proposition 3

i) If $E \leq \frac{1}{B_1} - 1$, we have $w_1^*(\mathbb{E}) = B_1(1+E) \leq 1$, so we can substitute $w_1 = w_1^*(\mathbb{E})$ and $w_2 = 1 + \eta - w_1$ in $\mathcal{W}$. Differentiating it w.r.t. $\hat{\Sigma}_0$ and $\hat{\Sigma}_1$, we have $\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\hat{\Sigma}_0} = -\frac{1}{2}\phi^2$ and $\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\hat{\Sigma}_1} = -\frac{1}{2}\beta\phi$, both negative. If $E > \frac{1}{B_1} - 1$, use $\hat{E}_1 = E$ in Equation 5, so we can use $w_1^*(\mathbb{E}) = 1$ and $w_2 = \eta$ in $\mathcal{W}$. Differentiating it, we have $\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\hat{\Sigma}_0} = -\frac{1}{2}\phi^2$ and $\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\hat{\Sigma}_1} = -\frac{1}{2}\beta\phi$, both negative.

ii) We require that variances are non-negative, so (i) implies optimal variances are 0.

A.4 Proof of Proposition 4

A.4.1 Step 1 – Optimal work given arbitrary deadline and arbitrary beliefs

Combining a deadline of the form $w_1 \geq \psi$ with the result from Proposition 1, the optimal work is $w_1^*\left(\left\{\hat{\mathbb{E}}_t\right\}, \psi\right) = \min\left\{1, \max\left\{\psi, B_1\left(1 + \hat{E}_1\right)\right\}\right\}$.

A.4.2 Step 2 – Optimal deadline given optimal work and arbitrary beliefs

A deadline $\psi \notin \left[w_1^*, 1\right]$ is ignored at $t = 1$, so at $t = 0$ the agent chooses $\psi \in \left[w_1^*, 1\right]$ to maximize V_0. So at $t = 1$ the agent chooses $w_1^*\left(\left\{\hat{\mathbb{E}}_t\right\}, \psi\right) = \psi$. Substituting in V_0 and differentiating yields

$\frac{\mathrm{d}V_0}{\mathrm{d}\psi} \propto -\left\{\psi - D_0\left(1+\hat{E}_0\right) - D_1\left(1+\hat{E}_1\right)\right\}$, where $D_0 := \frac{\beta^2+\phi^2}{\beta^2+\phi^2+\beta+\phi+\beta\phi}$, $D_1 := \frac{\beta\phi}{\beta^2+\phi^2+\beta+\phi+\beta\phi}$. Imposing $\psi \in \left[w_1^*, 1\right]$:

$$\begin{cases} \psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = B_1\left(1+\hat{E}_1\right) & \text{if } D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) \leq B_1\left(1+\hat{E}_1\right) \leq 1 \\ \psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) & \text{if } B_1\left(1+\hat{E}_1\right) \leq D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) \leq 1 \\ \psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = 1 & \text{otherwise.} \end{cases}$$

We will show below that optimal beliefs are optimistic; combined with $D_0 + D_1 < B_1$, this trivially proves the optimal deadline is smaller than $w_1^{\mathrm{RE}} = B_1\,(1+E)$.

A.4.3 Step 3 – Optimal beliefs given optimal work and optimal deadline

The optimality of complete overconfidence trivially follows from the assumption of quadratic utility, so we turn our attention to optimal expectations. They must satisfy $B_1\left(1+\hat{E}_1\right) \leq 1$, $D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) \leq 1$. If not, from our expressions for $w_1^*\left(\left\{\hat{\mathbb{E}}_t\right\}, \psi\right)$, $\psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right)$, we see that optimal work at $t = 1$ would be 1, and optimal beliefs could become more optimistic, yielding anticipatory benefits without altering behavior, so without cost. So we need only consider two cases.

First, let $\psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = B_1\left(1+\hat{E}_1\right)$; working as in Section A.2, we find the same optimal beliefs, which indeed satisfy $D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) \leq B_1\left(1+\hat{E}_1\right) \leq 1$.

Second, let $\psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right) = D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right)$; next, we find optimal beliefs and check when $B_1\left(1+\hat{E}_1\right) \leq D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) \leq 1$ is satisfied.

Substituting for $w_1^*\left(\left\{\hat{\mathbb{E}}_t\right\}, \psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right)\right)$ in $\mathcal{W}$ and ignoring the constraints, the F.O.C. w.r.t. $\hat{E}_0$, $\hat{E}_1$ yield $\hat{E}_0^\dagger = \bar{s}_0^{\mathrm{D}}\,(1+E) - 1$ and $\hat{E}_1^\dagger = \bar{s}_1^{\mathrm{D}}\,(1+E) - 1$, where $\bar{s}_0^{\mathrm{D}} := \frac{\phi^2 D_1^2 + \beta\phi D_0 (1-D_1)}{M_{D_0} M_{D_1} - G_D^2}\beta^2$ and $\bar{s}_1^{\mathrm{D}} := \frac{\phi^2 D_1 (1-D_0) + \beta\phi D_0^2}{M_{D_0} M_{D_1} - G_D^2}\beta^2$, with

$$\begin{aligned} G_D &:= F D_0 D_1 - \phi^2 D_1 - \beta\phi D_0 \qquad & M_{D_0} &:= F D_0^2 + \phi^2\,(1 - 2D_0) \\ & & M_{D_1} &:= F D_1^2 + \beta\phi\,(1 - 2D_1)\,. \end{aligned}$$

Algebra shows $G_D < 0$, $M_{D_0} > 0$, $M_{D_1} > 0$, and $M_{D_0} M_{D_1} > G_D^2$, and also $\hat{E}_0^\dagger \geq \hat{E}_1^\dagger$.

Imposing $\hat{E}_0 \geq 0$, $\hat{E}_1 \geq 0$, $D_0\left(1+\hat{E}_0\right) + D_1\left(1+\hat{E}_1\right) \leq 1$, but temporarily ignoring $D_0\left(1+\hat{E}_0\right) \geq (B_1 - D_1)\left(1+\hat{E}_1\right)$, the possible optimal beliefs are (see Section B.2.1 in the online appendix for details):

$$\begin{cases} \hat{\mathbb{E}}_0^{\dagger\dagger}[\eta]=0 & = & \hat{\mathbb{E}}_1^{\dagger\dagger}[\eta]=0 & \text{if} & & \mathbb{E}[\eta] & \leq \mu_L^{\mathrm{D}} \\ \hat{\mathbb{E}}_0^{\dagger\dagger}[\eta]=\underline{s}_0^{\mathrm{D}}\mathbb{E}[\eta]+\underline{c}_0^{\mathrm{D}} & > & \hat{\mathbb{E}}_1^{\dagger\dagger}[\eta]=0 & \text{if} & \mu_L^{\mathrm{D}} < & \mathbb{E}[\eta] & \leq \mu_I^{\mathrm{D}} \\ \hat{\mathbb{E}}_0^{\dagger\dagger}[\eta]=\bar{s}_0^{\mathrm{D}}(1+\mathbb{E}[\eta])-1 & > & \hat{\mathbb{E}}_1^{\dagger\dagger}[\eta]=\bar{s}_1^{\mathrm{D}}(1+\mathbb{E}[\eta])-1 & \text{if} & \mu_I^{\mathrm{D}} < & \mathbb{E}[\eta] & \leq \mu_U^{\mathrm{D}} \\ \hat{\mathbb{E}}_0^{\dagger\dagger}[\eta]=\frac{\bar{s}_0^{\mathrm{D}}}{D_0\bar{s}_0^{\mathrm{D}}+D_1\bar{s}_1^{\mathrm{D}}}-1 & > & \hat{\mathbb{E}}_1^{\dagger\dagger}[\eta]=\frac{\bar{s}_1^{\mathrm{D}}}{D_0\bar{s}_0^{\mathrm{D}}+D_1\bar{s}_1^{\mathrm{D}}}-1 & \text{if} & \mu_U^{\mathrm{D}} < & \mathbb{E}[\eta] & , \end{cases} \tag{A.2}$$

with $\mu_L^{\mathrm{D}}:=\frac{-\underline{c}_0^{\mathrm{D}}}{\underline{s}_0^{\mathrm{D}}}$, $\mu_I^{\mathrm{D}}:=\frac{1}{\bar{s}_1^{\mathrm{D}}}-1$, $\mu_U^{\mathrm{D}}:=\frac{1}{D_0\bar{s}_0^{\mathrm{D}}+D_1\bar{s}_1^{\mathrm{D}}}-1$, $\underline{s}_0^{\mathrm{D}}:=\frac{\beta^2 D_0}{M_{D_0}}$, $\underline{c}_0^{\mathrm{D}}:=\underline{s}_0^{\mathrm{D}}-\frac{G_D}{M_{D_0}}-1$.

Now we impose $D_0\left(1+\hat{E}_0\right)\geq(B_1-D_1)\left(1+\hat{E}_1\right)$. The constraint clearly binds for $E<\mu_L^{\mathrm{D}}$, and it either stops binding at $E=\mu_{L'}^{\mathrm{D}}\in\left[\mu_L^{\mathrm{D}},\mu_I^{\mathrm{D}}\right]$ or it never stops binding. This is because i) in $\left(\mu_L^{\mathrm{D}},\mu_I^{\mathrm{D}}\right]$, $\hat{E}_0^{\dagger\dagger}$ is increasing and $\hat{E}_1^{\dagger\dagger}$ is constant, so the constraint relaxes as E increases, and ii) for $E>\mu_I^{\mathrm{D}}$ the constraint does not depend on E. So using Equation A.2, we find that the constraint always binds if and only if $(B_1-D_1)\bar{s}_1^{\mathrm{D}}\leq D_0\bar{s}_0^{\mathrm{D}}$, which is equivalent to $\beta\geq\overline{\beta}_L(\phi):=\frac{2}{3}\sqrt[3]{\sqrt{\frac{31}{108}}+\frac{1}{2}}\phi\approx 0.68\phi$. Next, we examine these two cases.

$\underline{\boldsymbol{\beta \geq \overline{\beta}_L(\phi)}}$

In this case, the constraint $D_0\left(1+\hat{E}_0\right)\geq(B_1-D_1)\left(1+\hat{E}_1\right)$ binds only up to E equal to $\mu_{L'}^{\mathrm{D}}:=\frac{1}{\underline{s}_0^{\mathrm{D}}}\left(\frac{B_1-D_1-D_0}{D_0}-\underline{c}_0^{\mathrm{D}}\right)$; to determine this we use $\hat{E}_1^{\dagger\dagger}=0$ for $E\leq\mu_I^{\mathrm{D}}$ in the constraint. So $\hat{E}_0^{\mathrm{D}}=\frac{B_1-D_1-D_0}{D_0}$, $\hat{E}_1^{\mathrm{D}}=0$ for $E\leq\mu_{L'}^{\mathrm{D}}$, while for $E>\mu_{L'}^{\mathrm{D}}$ optimal beliefs are as in Equation A.2 (simply replace $\dagger\dagger$ with D). Finally, we can show that

$$\begin{cases} \mu_L^{\mathrm{ND}}<\mu_{L'}^{\mathrm{D}}<\mu_I^{\mathrm{D}}<\mu_U^{\mathrm{ND}}<\mu_U^{\mathrm{D}} & \text{if } \overline{\beta}_L(\phi)<\beta\leq\overline{\beta}(\phi) \\ \mu_{L'}^{\mathrm{D}}<\mu_L^{\mathrm{ND}}<\mu_I^{\mathrm{D}}<\mu_U^{\mathrm{ND}}<\mu_U^{\mathrm{D}} & \text{if } \overline{\beta}(\phi)<\beta\leq\overline{\beta}_U(\phi) \\ \mu_{L'}^{\mathrm{D}}<\mu_L^{\mathrm{ND}}<\mu_I^{\mathrm{D}}<\mu_U^{\mathrm{D}}<\mu_U^{\mathrm{ND}} & \text{if } \overline{\beta}_U(\phi)<\beta, \end{cases}$$

where $\overline{\beta}(\phi)$, $\overline{\beta}_U(\phi)$ are increasing in ϕ (see Section B.2.1 in the online appendix for details).

To check whether a binding deadline is optimally imposed or not, we need to compare the well-being $\mathcal{W}$ with and without a deadline, $\mathcal{W}^{\mathrm{D}}$ and $\mathcal{W}^{\mathrm{ND}}$, respectively. First, we observe that optimal beliefs are piece-wise linear in E and $\mathcal{W}^{\mathrm{D}}-\mathcal{W}^{\mathrm{ND}}$ is a quadratic in beliefs, so $\mathcal{W}^{\mathrm{D}}-\mathcal{W}^{\mathrm{ND}}$ is a differentiable piecewise quadratic in E. Using this fact, together with some algebra, we can show (see Section B.2.1 in the online appendix for details) that $\mathcal{W}^{\mathrm{D}}\geq\mathcal{W}^{\mathrm{ND}}$ if and only if $\beta\geq\overline{\beta}(\phi)$ and $E\in M(\beta,\phi)$, where $M(\beta,\phi)$ is a convex set. If a binding deadline is not optimal, then optimal beliefs are as in Equation A.1.

$\underline{\boldsymbol{\beta < \overline{\beta}_L (\phi)}}$

Here, the constraint $(B_1 - D_1)\left(1+\hat{E}_1\right) \leq D_0\left(1+\hat{E}_0\right)$ binds for all E. So optimal beliefs implementing the binding deadline must always be proportional, so they are both constants or both proportional to E. Given that we have the constraints $\hat{E}_0 \geq 0$, $\hat{E}_1 \geq 0$ and $w_1 \leq 1$, we conclude that there are values of E, $\mu^{\mathrm{D}}_{I'}$ and $\mu^{\mathrm{D}}_{U'}$ to be defined below, that partition the E space in regions: for $E \leq \mu^{\mathrm{D}}_{I'}$ optimal beliefs do not depend on E because $\hat{E}_0 \geq 0$ and $\hat{E}_1 \geq 0$ bind, for $\mu^{\mathrm{D}}_{I'} < E \leq \mu^{\mathrm{D}}_{U'}$ optimal beliefs are proportional to E, and for $\mu^{\mathrm{D}}_{U'} < E$ optimal beliefs do not depend on E because $w_1 \leq 1$ binds. Working as above, we find

$$\begin{cases} \hat{\mathbb{E}}^{\dagger\dagger}_0[\eta] = \frac{B_1-D_1-D_0}{D_0} & > \quad \hat{\mathbb{E}}^{\dagger\dagger}_1[\eta] = 0 & \text{if} & & \mathbb{E}[\eta] & \leq \mu^{\mathrm{D}}_{I'} \\ \hat{\mathbb{E}}^{\dagger\dagger}_0[\eta] = \overline{s}^{\mathrm{D}'}_0(1+\mathbb{E}[\eta]) - 1 & > \quad \hat{\mathbb{E}}^{\dagger\dagger}_1[\eta] = \overline{s}^{\mathrm{D}'}_1(1+\mathbb{E}[\eta]) - 1 & \text{if} & \mu^{\mathrm{D}}_{I'} < & \mathbb{E}[\eta] & \leq \mu^{\mathrm{D}}_{U'} \\ \hat{\mathbb{E}}^{\dagger\dagger}_0[\eta] = \frac{1}{B_1}\frac{B_1-D_1}{D_0} - 1 & > \quad \hat{\mathbb{E}}^{\dagger\dagger}_1[\eta] = \frac{1}{B_1} - 1 & \text{if} & \mu^{\mathrm{D}}_{U'} < & \mathbb{E}[\eta] & , \end{cases}$$

with $\overline{s}^{\mathrm{D}'}_0 := \frac{B_1\beta^2}{2G_D + \frac{B_1-D_1}{D_0}M_{D_0} + \frac{D_0 M_{D_1}}{B_1-D_1}}$, $\overline{s}^{\mathrm{D}'}_1 := \frac{D_0}{B_1}\frac{B_1-D_1+D_0}{B_1-D_1}\overline{s}^{\mathrm{D}'}_0$, $\mu^{\mathrm{D}}_{I'} := \frac{1}{\overline{s}^{\mathrm{D}'}_1} - 1$, $\mu^{\mathrm{D}}_{U'} := \frac{1}{D_0\overline{s}^{\mathrm{D}'}_0 + D_1\overline{s}^{\mathrm{D}'}_1} - 1$.

Finally, $\mathcal{W}^{\mathrm{D}} < \mathcal{W}^{\mathrm{ND}}$ everywhere (see Section B.2.1 in the online appendix for details), so a deadline is never optimal and optimal beliefs are as in Equation A.1.

A.5 Proof of Proposition 5

Finding the optimal beliefs We work as in Sections A.4.1 and A.4.2 to find $w^*_1\left(\left\{\hat{\mathbb{E}}_t\right\}, \psi\right)$ and $\psi^{*,\mathrm{ED}}\left(\left\{\hat{\mathbb{E}}_t\right\}\right)$, the optimal externally-imposed deadline given beliefs. $\psi^{*,\mathrm{ED}}\left(\left\{\hat{\mathbb{E}}_t\right\}\right)$ is like $\psi^*\left(\left\{\hat{\mathbb{E}}_t\right\}\right)$ in Section A.4.2, except E replaces $\hat{E}_0$ everywhere. We are interested in optimal beliefs that implement a binding deadline, so as in Section A.4, we differentiate $\mathcal{W}$ w.r.t. $\hat{E}_0$, $\hat{E}_1$ to find i) $\frac{\mathrm{d}\mathcal{W}}{\mathrm{d}\hat{E}_0} < 0$, so with $\hat{E}_0 \geq 0$ we have $\hat{E}^{\mathrm{ED}}_0 = 0$, and ii) "interior" optimal $\hat{E}_1$ is $\hat{E}^{\dagger}_1 = \frac{\beta^2(1+E)+\phi^2}{M_{D_1}}D_1 - 1$.

Also: i) Substituting $\hat{E}^{\mathrm{ED}}_0$, $\hat{E}^{\dagger}_1$ in $B_1\left(1+\hat{E}_1\right) \leq D_0(1+E) + D_1\left(1+\hat{E}_1\right)$, we see that it does not bind for these beliefs. ii) Setting $\hat{E}^{\dagger}_1 = \frac{\beta^2(1+E)+\phi^2}{M_{D_1}}D_1 - 1 = 0$, we see that $\hat{E}_1 \geq 0$ binds for $E < \mu^{\mathrm{ED}}_I := \frac{1}{\beta^2}\left(\frac{M_{D_1}}{D_1} - \phi^2\right) - 1 = \frac{\beta+\phi}{\beta^2} > 0$. iii) For $\hat{E}^{\mathrm{ED}}_0 = \hat{E}^{\mathrm{ED}}_1 = 0$, $B_1\left(1+\hat{E}_1\right) \leq D_0(1+E) + D_1\left(1+\hat{E}_1\right)$ binds at $E = \mu^{\mathrm{ED}}_L := \frac{B_1-D_1-D_0}{D_0} = \frac{\beta\phi}{(\beta^2+\phi^2)(1+\beta+\phi)} \leq \mu^{\mathrm{ED}}_I$; but it cannot be satisfied for $E < \mu^{\mathrm{ED}}_L$ neither by reducing $\hat{E}_0$, $\hat{E}_1$ (which would be impossible) nor by raising them (which would not be helpful), so no beliefs implement the externally-imposed binding deadline for $E \leq \mu^{\mathrm{ED}}_L$. iv) Substituting $\hat{E}^{\mathrm{ED}}_0$, $\hat{E}^{\dagger}_1$ in $D_0(1+E) + D_1\left(1+\hat{E}_1\right) \leq 1$, we see that it binds for $E >$

$\mu_U^{\mathrm{ED}} := \frac{M_{D_1} - \phi^2 D_1^2}{D_0 M_{D_1} + \beta^2 D_1^2} - 1$; using $E = \mu_U^{\mathrm{ED}}$ in $\hat{E}_1^{\dagger}$, we get $\hat{E}_1^{\mathrm{ED}} = \frac{\beta^2 + \phi^2 D_0}{D_0 M_{D_1} + \beta^2 D_1^2} D_1 - 1$ for $E > \mu_U^{\mathrm{ED}}$.

We have shown $\hat{E}_0^{\mathrm{ED}} = 0 \leq \hat{E}_1^{\mathrm{ED}}$, i.e., beliefs become more pessimistic over time. So to prove optimism, we just need to show $\hat{E}_1^{\mathrm{ED}} \leq E$. For $E \leq \mu_I^{\mathrm{ED}}$, we have $\hat{E}_1^{\mathrm{ED}} = 0 < E$. For $\mu_U^{\mathrm{ED}} < E$, we have $\hat{E}_1^{\mathrm{ED}} = \frac{\beta^2 + \phi^2 D_0}{D_0 M_{D_1} + \beta^2 D_1^2} D_1 - 1 < \mu_U^{\mathrm{ED}}$. For $\mu_I^{\mathrm{ED}} < E \leq \mu_U^{\mathrm{ED}}$, as a function of E, $\hat{E}_1^{\mathrm{ED}}$ is a straight line segment whose endpoints lie below the line E, so $\hat{E}_1^{\mathrm{ED}} < E$.

Having determined optimal beliefs $\hat{\mathbb{E}}_t^{\mathrm{ED}}$, we define $\psi^{\mathrm{ED}} := \psi^{*,\mathrm{ED}}\left(\left\{\hat{\mathbb{E}}_t^{\mathrm{ED}}\right\}\right)$.

Outsider's deadline is stricter than the agent's deadline To show this, we need to show $\psi^{\mathrm{ED}} \geq \psi^{\mathrm{D}}$, i.e., $D_0\left(E - \hat{E}_0^{\mathrm{D}}\right) \geq D_1\left(\hat{E}_1^{\mathrm{D}} - \hat{E}_1^{\mathrm{ED}}\right)$. Straightforward algebra shows this is true for interior beliefs (so also for beliefs above the interior). Now we show it is true for all remaining beliefs for which a binding self-imposed deadline is optimal. From Section A.4, we know this is the case only if $E > \mu_L^{\mathrm{D}}$, so we simply need to check the case $\hat{E}_1^{\mathrm{D}} = 0$. But we already know from Section A.4 that $\hat{E}_0^{\mathrm{D}} \leq E$, hence $D_0\left(E - \hat{E}_0^{\mathrm{D}}\right) \geq D_1\left(\hat{E}_1^{\mathrm{D}} - \hat{E}_1^{\mathrm{ED}}\right)$.

Outsider's deadline is smaller than $w_1^{\mathbf{RE}}$ We have shown that beliefs are optimistic; combined with $D_0 + D_1 < B_1$, this trivially proves $\psi^{\mathrm{ED}} < w_1^{\mathrm{RE}}$.

A.6 Proof of Proposition 6

The unconstrained optimal beliefs without a deadline are $\hat{E}_0^{\mathrm{ND}} = 0$ (so $\frac{\mathrm{d}\hat{E}_0^{\mathrm{ND}}}{\mathrm{d}\beta} = \frac{\mathrm{d}\hat{E}_0^{\mathrm{ND}}}{\mathrm{d}\phi} = 0$), and $\hat{E}_1^{\mathrm{ND}} = \underline{s}_1^{\mathrm{ND}} E + \underline{c}_1^{\mathrm{ND}}$, while the optimal work is $w_1^{\mathrm{ND}} = B_1\left(1 + \hat{E}_1^{\mathrm{ND}}\right)$. Straightforward but tedious algebra shows that $\frac{\mathrm{d}\hat{E}_1^{\mathrm{ND}}}{\mathrm{d}\beta} > 0$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\beta} > 0$, and $\frac{\mathrm{d}\hat{E}_1^{\mathrm{ND}}}{\mathrm{d}\phi} < 0$.

To determine the effect of ϕ on w_1^{ND}, we first determine its effect on the endpoints of its non-horizontal segment. The left endpoint moves rightward by $\frac{\mathrm{d}\mu_L^{\mathrm{ND}}}{\mathrm{d}\phi}$ and upward by $\frac{\mathrm{d}B_1}{\mathrm{d}\phi}$, so we calculate $B_1 \underline{s}_1^{\mathrm{ND}} \frac{\mathrm{d}\mu_L^{\mathrm{ND}}}{\mathrm{d}\phi} - \frac{\mathrm{d}B_1}{\mathrm{d}\phi} \propto \beta - \phi$, which means that as ϕ increases, the left endpoint is above the original non-horizontal segment if and only if $\beta > \phi$. The right endpoint moves rightward by $\frac{\mathrm{d}\mu_U^{\mathrm{ND}}}{\mathrm{d}\phi} \propto (\beta + \phi)^3 + \beta(\beta - \phi)$, which is positive for $\phi \leq \overline{\phi}(\beta)$ and negative for $\phi > \overline{\phi}(\beta)$. Thus, for $\phi \leq \beta$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} < 0$; for $\beta < \phi \leq \overline{\phi}(\beta)$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} > 0$ for small E and $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} < 0$ for large E; and for $\overline{\phi}(\beta) < \phi$, $\frac{\mathrm{d}w_1^{\mathrm{ND}}}{\mathrm{d}\phi} > 0$.

A.7 Proof of Proposition 7

In this case, the unconstrained optimal beliefs are (see Equation A.2) $\hat{E}_0^{\mathrm{D}} = \overline{s}_0^{\mathrm{D}}(1 + E) - 1$, $\hat{E}_1^{\mathrm{D}} = \overline{s}_1^{\mathrm{D}}(1 + E) - 1$. The optimal deadline is $\psi^{\mathrm{D}} = D_0\left(1 + \hat{E}_0^{\mathrm{D}}\right) + D_1\left(1 + \hat{E}_1^{\mathrm{D}}\right)$. Straightforward but tedious algebra shows that $\frac{\mathrm{d}\hat{E}_0^{\mathrm{D}}}{\mathrm{d}\beta} \propto \frac{\mathrm{d}\hat{E}_1^{\mathrm{D}}}{\mathrm{d}\beta} \propto 1 + E > 0$ and $\frac{\mathrm{d}\psi^{\mathrm{D}}}{\mathrm{d}\beta} > 0$, and also $\frac{\mathrm{d}\hat{E}_0^{\mathrm{D}}}{\mathrm{d}\phi} \propto \frac{\mathrm{d}\hat{E}_1^{\mathrm{D}}}{\mathrm{d}\phi} \propto -(1 + E) < 0$ and $\frac{\mathrm{d}\psi^{\mathrm{D}}}{\mathrm{d}\phi} < 0$.

A.8 Proof of Proposition 8

Optimal work. Optimal work at $t = 1$ maximizes $V_1 = \mathbb{E}[U_1 + \xi\beta U_2]$. Use U_t, $w_2 = 1 + \eta - w_1$, to write the objective as $\mathbb{E}[u(w_1) + \xi(\beta + \phi)u(1 + \eta - w_1)]$, which is concave in w_1. Ignoring $w_1 \leq 1$ and substituting for $u(\cdot)$, F.O.C. w.r.t. w_1 yields $w_1^{\dagger} = \tilde{B}(1 + E)$, for $\tilde{B} := \frac{\xi(\beta+\phi)}{1+\xi(\beta+\phi)}$. Imposing $w_1 \leq 1$ yields w_1^{H}.

Optimal work given arbitrary deadline. With a deadline $w_1 \geq \psi$, the optimal work is $w_1^{\mathrm{H}}(\psi) = \min\left\{1, \max\left\{\psi, w_1^{\mathrm{H}}\right\}\right\} = \min\left\{1, \max\left\{\psi, \tilde{B}(1 + E)\right\}\right\}$.

Optimal deadline given optimal work. As in previous proofs, we can restrict our attention to $\psi \in \left[w_1^{\mathrm{H}}, 1\right]$, so $w_1^{\mathrm{H}}(\psi) = \psi$. So the optimal deadline at $t = 0$ maximizes V_0 subject to $\psi \in \left[w_1^{\mathrm{H}}, 1\right]$. Manipulating this, we get

$$\begin{aligned} V_0 &\propto -\frac{1}{2}\left\{\psi^2 + \tilde{D}\left[(1 + E)^2 + \Sigma - 2(1 + E)\psi\right]\right\} \\ \frac{\mathrm{d}V_0}{\mathrm{d}\psi} &\propto -\left\{\psi - \tilde{D}(1 + E)\right\}, \end{aligned}$$

so imposing $\psi \in \left[w_1^{\mathrm{H}}, 1\right]$ we get $\psi^{\mathrm{H}} = \tilde{B}(1+E)$ if $\tilde{D} \leq \tilde{B} \leq \frac{1}{1+E}$, $\psi^{\mathrm{H}} = \tilde{D}(1+E)$ if $\tilde{B} \leq \tilde{D} \leq \frac{1}{1+E}$, and $\psi^{\mathrm{H}} = 1$ otherwise, where $\tilde{D} := \frac{\beta^2+\xi\beta\phi+\phi^2}{\beta^2+\beta+\xi\beta\phi+\phi+\phi^2}$.

Note that $\tilde{D} - \tilde{B} = \frac{(1-\xi)\left[(\beta-\phi)^2 + \frac{2-3\xi}{1-\xi}\beta\phi\right]}{\left(\beta^2+\beta+\xi^2\beta\phi+\xi\phi+\xi\phi^2\right)(1+\xi(\beta+\phi))}$. For $\xi < 1$, $\tilde{D} - \tilde{B} = 0$ is a conic section with discriminant $\left(\frac{\xi}{1-\xi}\right)^2 - 4$. For $\xi < \frac{2}{3}$ we have $\tilde{D} - \tilde{B} > 0$; while for $\xi = \frac{2}{3}$, we have $\tilde{D} - \tilde{B} = 0$ for $\beta = \phi$ and $\tilde{D} - \tilde{B} > 0$ otherwise. So the deadline is *not* chosen only for $\xi > \frac{2}{3}$ and $\tilde{D} - \tilde{B} < 0$, where $\tilde{D} - \tilde{B} = 0$ is a degenerate hyperbola, which can be written as $\xi = \frac{\beta^2+\phi^2}{\beta^2+\phi^2+\beta\phi}$.

A.9 Proof of Proposition 9

If $\tilde{D} \leq \tilde{B} \leq \frac{1}{1+E}$, a deadline is not chosen, so $w_1^{\mathrm{H}} = \tilde{B}(1 + E)$ and

$$\frac{\mathrm{d}w_1^{\mathrm{H}}}{\mathrm{d}\beta} = \frac{\xi}{(1+\xi\beta+\xi\phi)^2}(1 + E) > 0 \qquad \frac{\mathrm{d}w_1^{\mathrm{H}}}{\mathrm{d}\phi} = \frac{\xi}{(1+\xi\beta+\xi\phi)^2}(1 + E) > 0.$$

If $\tilde{B} \leq \tilde{D} \leq \frac{1}{1+E}$, the deadline binds, so $\psi^{\mathrm{H}} = \tilde{D}(1+E) = \frac{\beta^2+\xi\beta\phi+\phi^2}{\beta^2+\beta+\xi\beta\phi+\phi+\phi^2}(1+E)$ and

$$\frac{\mathrm{d}\psi^{\mathrm{H}}}{\mathrm{d}\beta} = \frac{\beta^2+2\beta\phi+\xi\phi^2-\phi^2}{\left(\beta^2+\beta+\xi\beta\phi+\phi+\phi^2\right)^2}(1 + E) \qquad \frac{\mathrm{d}\psi^{\mathrm{H}}}{\mathrm{d}\phi} = \frac{-\beta^2+\xi\beta^2+2\beta\phi+\phi^2}{\left(\beta^2+\beta+\xi\beta\phi+\phi+\phi^2\right)^2}(1 + E).$$

Thus, $\frac{\mathrm{d}\psi^{\mathrm{H}}}{\mathrm{d}\beta} > 0 \Leftrightarrow \xi > \frac{\phi^2-2\beta\phi-\beta^2}{\phi^2}$ and $\frac{\mathrm{d}\psi^{\mathrm{H}}}{\mathrm{d}\phi} > 0 \Leftrightarrow \xi > \frac{\beta^2-2\beta\phi-\phi^2}{\beta^2}$.