An Economic Model of The Planning Fallacy*

Markus K. Brunnermeier†
Princeton University, CEPR and NBER

Filippos Papakonstantinou‡
Princeton University

Jonathan A. Parker§
Northwestern University and NBER

February 2008 (version 7w)

Abstract

People tend to underestimate the work involved in completing tasks and consequently finish tasks later than expected or do an inordinate amount of work right before projects are due. We present a theory in which people procrastinate because the ex-ante utility benefits of anticipating that a task will be easy to complete outweigh the average ex-post costs of poor planning. We show that, given a commitment device, people self-impose deadlines that are binding but require less smoothing of work than that chosen by a person with objective beliefs. We test our theory using extant experimental evidence on differences in expectations and behavior. We find that reported beliefs and behavior generally respond as our theory predicts. For example, monetary incentives for accurate prediction ameliorate the planning fallacy while incentives for rapid completion aggravate it.

Keywords: Planning Fallacy, Optimal Beliefs, Procrastination, Expectations, Optimism, Self-Control, Deadlines
JEL Classification: D10, D80, E21

*For helpful comments, we thank seminar participants at New York University, Princeton University, and the 2006 Whitebox Advisors Conference at Yale University, although we remain responsible for all errors, omissions, and typos. Brunnermeier acknowledges financial support from the National Science Foundation (SES 140-6139) and the Sloan Foundation.

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

‡Department of Economics, Princeton University, Princeton, NJ 08544-1013, e-mail: fpapakon@princeton.edu, http://www.princeton.edu/∼fpapakon

§Kellogg School of Management, Northwestern University, 2001 Sheridan Road, Evanston, IL 60208-2001, e-mail: Jonathan-Parker@Kellogg.Northwestern.edu, www.kellogg.northwestern.edu/faculty/parker/htm

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.

- - Kahneman and Tversky (1979)

1 Introduction

Faced with an unpleasant task, people tend both to underestimate the time necessary to complete the task and to postpone working on the task. Thus, projects often take inordinately long to complete and people struggle to meet, or even miss, deadlines. Kahneman and Tversky (1979) terms this behavior *the planning fallacy*, and such overoptimistic beliefs and behavioral delays seem to be important factors in a wide range of human economic behaviors, from filing taxes or doing referee reports, to planning for retirement or making investments in health.

This paper makes two contributions. First, we develop a theory of the planning fallacy and derive its implications. In our theory, people are biased toward believing that a project will be easy to complete because such a bias increases expected utility more than it decreases utility due to poor smoothing of work over time. Since people understand that they have a tendency to poorly smooth future work over time, they may choose to commit to future actions by imposing deadlines on themselves. The planning fallacy and the desire for deadlines arise solely from optimistic beliefs, as in Kahneman and Tversky's description, rather than through preferences with objective beliefs, as in most of the extant economic research on procrastination and commitment (e.g. Laibson (1997), Gul and Pesendorfer (2004)).

Second, we test our theory using evidence on both behavior and, importantly, beliefs. It is generally the case that observed behavior can be matched both by some utility function for agents with objective beliefs and by some other utility function with some other model of beliefs.[1] One approach to this identification problem between beliefs and preferences over outcomes is to make the assumption of rationality, perhaps based on philosophical arguments.[2] We instead address this identification problem by using data on beliefs. We derive the testable implications of our model for beliefs and test these using variation in both observed behavior and reported beliefs across experimental treatments. The predictions of our model for both beliefs and behaviors across experimental settings are largely consistent with the actual pattern of beliefs and behaviors found in psychologists' experiments on the planning fallacy. This finding supports our model of the planning fallacy and the use of survey data on expectations to discipline research in behavioral economics.

[1] With enough data on the same choice situation and enough assumptions about preferences, one can achieve identification, as for example in Savage (1954).

[2] Two versions of these arguments bear mention. First, rationality is sometimes viewed as providing greater evolutionary fitness. But these arguments are specific to particular environments and utility functions, and are overturned in other environments. The failure of this argument in general follows directly from the original identification problem. A second argument is that people should learn true probabilities over time, which simply pushes back the assumption of rationality to one of rational learning and so is still an assumption rather than a scientific basis for distinguishing among models.

To be more specific about our theory, we study the planning fallacy in a model of a person who faces a task of uncertain total difficulty and who has quadratic disutility of work in each of two periods. Right before the first period, prior beliefs are set, and the person may choose an intermediate deadline. In the first period, the person gets a signal about the total work involved, beliefs are reset, and the person decides how much work to do. In the second period, the person completes the task, however difficult it turns out to be.

Without an intermediate deadline, two features of the model lead a person to exhibit the planning fallacy. First, the person has anticipatory utility. Thus, a person who initially believes that the task will be easy to complete has higher expected utility because he anticipates less work in the future. This first ingredient provides an ex ante anticipatory benefit of overly optimistic beliefs. Second, the person optimizes given his beliefs. Thus, a person with optimistic beliefs does little work in the present and ends up poorly smoothing work over time. This second ingredient implies an ex post cost of optimism on average: optimistic assessments lead to potentially costly delays and/or rushing at the end. Given these two ingredients, it is natural for people to exhibit the planning fallacy because a little optimism has first-order ex ante anticipatory benefits and, by the envelope theorem, only second-order ex post behavioral costs. We define well-being as average liftetime utility and show that optimistic beliefs optimally balance these benefits and costs. This gives us a theory of procrastination that is endogenous and situational. The severity of belief bias and of procrastination is larger the greater the anticipatory benefits of optimism and the less the ex-post costs of mis-planning.

With an intermediate deadline that binds, the person can still get the benefits of believing a task will be easy to complete during the first period, without the costs of poorly smoothing work over time. Thus, a binding deadline can make the agent better off. But in order for the person initially to choose a deadline that binds, he has to understands that he will tend to exhibit the planning fallacy during period 1. Thus, optimal beliefs are time-inconsistent. Beliefs are initially somewhat realistic, so that the person chooses a deadline that binds and smooths work, and later more optimistic, so that the person during period 1 gets the benefits of anticpating little work.

Because we assume that peoples' beliefs optimally balance the anticipatory benefits of more optimistic beliefs and the ex post costs of poor smoothing of work effort over time, the person's beliefs, his choice of deadline, and the extent of procrastination are all endogenous and situational, giving testable predictions across environments.

To be more specific about our testing, we find that the predictions of our model are generally consistent with experimental data and survey data on both beliefs and task-completion times reported in the psychology literature on the planning fallacy. In particular, we derive and test eight predictions of our model. (i) As our model predicts, expected and actual completion times are highly correlated across people within an experiment and across experiments. That is, reported beliefs are informative and not simply noise. (ii) Our model predicts that framing does not cause the planning fallacy. In existing experiments, we find that the fallacy is robust to many, but not all, framing manipulations and attempts to de-bias beliefs. (iii) As our model predicts, where there is no benefit to optimistic anticipation as when the task is not onerous, people's behavior does not exhibit the planning fallacy. (iv) As predicted, when people are given an opportunity to impose a deadline on themselves prior to beginning the task, they tend to do so. (v) People impose deadlines that require less than perfect smoothing of work over time. (vi)

The self-imposed deadlines increase performance, but not as much as exogenous deadlines that impose perfect work-smoothing over time.

Finally, we modify our environment to match two different experimental settings: one in which subjects are paid ex post for rapid task completion and one in which they are paid for accurate prediction of task completion times. (vii) As our model predicts, experiments find that monetary incentives for rapid completion of tasks increase the degree of overoptimism. (viii) Consistent with our model, monetary incentives for accurate prediction increase the accuracy of people's predictions, but, inconsistent with our model, this does not in turn lead to more rapid completion, although this evidence is statistically weak.

Given the large recent literature on procrastination and commitment, it is worth emphasizing that the key novelty in our theory is that beliefs are endogenous and are the central cause of procrastination and the demand for commitment, while in existing economic models they play at most a secondary role. According to the conventional economic view, beliefs are exogenously specified, typically as rational or naive, and procrastination occurs due to utility costs of self-control or benefits to delay. In one branch of the literature, people discount non-exponentially in an otherwise standard time-separable utility maximization problem (following Strotz (1955-1956)); in the other, people maximize utility that has an additional component, 'temptation utility,' that makes it unpleasant to discipline oneself to work today (Gul and Pesendorfer (2001)). In either case, procrastination occurs because the utility function places special importance on the present relative to the future. Beliefs are less important. Whether people completely understand their tendency to procrastinate (as in Laibson (1997) and Gul and Pesendorfer (2004)), are completely oblivious to it (as in Akerlof (1991)), or (exogenously) partially understand their tendencies (as in O'Donoghue and Rabin (1999a)), people still procrastinate.[3] In our model, people fail to smooth effort over time *only* because they endogenously mispredict the difficulty of doing the task in the future. Thus we focus on beliefs in testing.

This paper is also related to economic research that endogenizes beliefs, such as Akerlof and Dickens (1982), Yariv (2002), Bénabou and Tirole (2002) and Bernheim and Thomadsen (2005). While this paper is most closely related to Brunnermeier and Parker (2005), our analysis differs in that we do not require that agents update as Bayesians.

Third, this paper is related to economic research developing a scientific foundation for the modeling of beliefs. Survey data on beliefs have been used to test rationality or evaluate its usefulness in the context of specific models (Manski (2004)).[4] The fact that our model largely fits variation in elicited beliefs provides evidence that such data is not simply noise but has meaningful variation (linked to behavior).

Finally, there are many psychological theories of the planning fallacy which focus on the mental processes that lead people to fail to make correct predictions. These theories are generally consistent with our theory and inconsistent with rational models of procrastination and commitment. Kahneman and Tversky (1979) argue that the fallacy arises because people ignore distributional information available in related past outcomes and instead focus only an a single plausible scenario for completion of the current specific task. Liberman and Trope (2003)

[3]See also O'Donoghue and Rabin (1999b), Fischer (1999), Harris and Laibson (2001), and O'Donoghue and Rabin (2001).

[4]Examples include Hamermesh (1985), Dominitz (1998), Nyarko and Schotter (2002), Souleles (2004), and Mankiw, Reis, and Wolfers (2004).

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 host of potential difficulties and small tasks involved in task completion. Further, a large set of papers on the planning fallacy have investigated whether the planning fallacy is caused by incorrect memory of past events (that people are not aware of and so do not adjust for), such as biased self-attribution. Buehler, Griffin, and Ross (1994) (study 3) for example find that people describe the reasons for their own past failures to meet expected completion dates as more external, more transitory and more specific than they describe the reasons for the failures of others. Finally, and more related to our work, there is the general theory that people are optimistic, and this is helpful in generating motivation, effort and persistence (see for example Armor and Taylor (1998)).

The paper is structured as follows. The next section introduces our model. Section 3 presents our theory of the planning fallacy and relates it to a first set of experimental evidence. Section 4 shows how people mitigate their misplanning using deadlines and again tests the predictions of the model with behavior observed in experiments. Section 5 derives predictions of our model for a broad set of experimental settings and tests these predictions using reported beliefs and behavior. A final section concludes and an appendix provides proofs of all propositions.

2 The model

This section presents a model of the temporal allocation of work. We incorporate an intermediate deadline that commits the person to complete some amount of the work early in the project. A first subsection describes utility, subjective beliefs, and objective uncertainty; a second subsection presents the person's well-being, the objective function for subjective beliefs.

2.1 The environment

We consider a person choosing an intermediate deadline and then how much work, w_t, to allocate to completing a task in each of two periods, $t = 1, 2$.

The total work required to complete the task is random, and is given by $\eta_1 + \eta_2$ where η_t is realized in period t. The person must complete all work by the end of period 2, so he faces the constraint:

$$w_1 + w_2 = \eta_1 + \eta_2. \tag{1}$$

We assume that η_1 and $\frac{\eta_2}{\eta_1}$ are i.i.d random variables with mean one and each with strictly positive support $[0, 2]$.[5] Importantly, this assumption implies that $E[\eta_2|\eta_1] = \eta_1$ so that the realization of η_1 in the first period provides a good guide as to the realization of η_2 in the second. For example, the work involved in preparing for a midterm exam is a good indicator of the work that will be involved in preparing for the final.[6]

[5] We also assume that each has zero probability of equalling zero. All propositions hold with probability one.

[6] The assumption $E[\eta_2|\eta_1] = \eta_1$ also ensures that neither the rational agent nor the optimistic agent chooses $w_1 > \eta_1$ which would lead to $w_2 < 0$ for some realizations of η_2 – a situation in which our interpretation of the mathematical structure becomes strained. Note also that all our results generalize to a setting in which $E[\eta_2|\eta_1] = \kappa\eta_1$ for $\kappa \in [0, 1]$.

The intermediate deadline requires that the person complete at least a fraction $\phi \in [0,1]$ of η_1, the work amount realized in period 1:

$$w_1 \geq \phi\eta_1. \tag{2}$$

Since $E[\eta_2|\eta_1] = \eta_1$, this can equivalently be viewed as requiring that the person complete $\phi/2$ of the total work on average in period 1.

The person holds prior subjective beliefs about the joint distribution of η_1 and η_2 immediately before observing η_1, and holds subjective beliefs about the distribution of η_2 during period 1 after observing η_1. These beliefs may or may not coincide with the objective distributions. We require that prior beliefs are internally consistent, satisfying Bayes' rule, but we do *not* require that period 1 beliefs follow from Bayes' rule applied to prior beliefs. We denote subjective prior expectations $\hat{E}[\cdot]$ .and subjective period 1 expectations after observing η_1 as $\hat{E}_1[\cdot]$ (and we do not require that $\hat{E}[\cdot|\eta_1] = \hat{E}_1[\cdot|\eta_1]$). We denote subjective (co)variances by $\hat{V}ar[\eta_t]$, $\hat{C}ov[\eta_1, \eta_2]$ and $\hat{V}ar_1[\eta_2]$, where the last depends on η_1.

At the start of period 1, prior to starting work and observing η_1, the person chooses the deadline, ϕ, to maximize his expected present discounted value of utility flow, $\hat{E}[V_1]$, where

$$V_1 := u(w_1) + u(w_2),$$

taking into account how w_1 will be chosen as a function of η_1 and that the task will be completed in period 2, equation (1). Utility flow in each period is decreasing and concave in work. We are assuming for simplicity that the future is not discounted.

During period 1, after observing η_1, the person chooses w_1 to maximize the expected present discounted value of the utility flow, $\hat{E}_1[V_1]$, subject to meeting the deadline, equation (2). During period 2, he completes the task.

Utility flow in each period is quadratic:

$$u(w_t) = -\frac{1}{2}w_t^2.$$

Quadratic utility delivers certainty equivalence conditional on beliefs, and implies that only the subjective means and variances of η_t are relevant for behavior and expected utility.[7]

2.2 The objective function for beliefs

We take a classical utilitarian rather than a traditional revealed-preference interpretation of the objective functions. We define $\hat{E}[V_1]$ and $\hat{E}_1[V_1]$ – the expected utilities of current and future work – as the person's *felicities* in the instant he chooses the deadline and during period 1 as he does w_1, respectively. Thus, the expected value functions represent a person's current 'happiness.' That is, as emphasized by Bentham, Hume, Böhm-Bawerk and other early economists, people are made happy in the present – get more felicity – both by what they currently experience and also by what they anticipate happening in the future.[8] Since the happiness of people

[7] Important for tractability, this property reduces the dimensionality of the problem of solving for optimal beliefs about η_t from infinite to 2.

[8] See Loewenstein (1987) and the discussion of the Samuelsonian and Jevonian views of utility in Caplin and Leahy (2000).

that plan for and care about the future is affected by their expectations of future events, people who care about the future can get more current felicity not only by having better plans for their actions but also by having more optimistic beliefs about future uncertainties.

People also get felicity in period 2:

$$V_2 := \delta u(w_1) + u(w_2)$$

where $0 \leq \delta \leq 1$. Thus, just as people get felicity from the future (anticipatory utility), they can also get it from the past (memory utility). When $\delta = 0$ the agent gets no utility (or disutility) from past actions. When $\delta = 1$, the agent discounts the past at the same rate that he discounts the future. Felicity in the second period does not affect the agent's actions given beliefs, but does affect what beliefs are optimal, to which we now turn.

Since a person's current felicity depends on his expectations about the course of the future, he is happier in the present with more optimistic beliefs about the future. A person who believes that a task is unrealistically easy to complete exerts less work in period 1 and ends up working more in period 2. But such beliefs come at a cost: they sacrifice the benefits of smoothing the necessary work over time. There is thus a trade-off between optimism, which raises anticipatory utility, and objectivity, which allows better smoothing of work over time.

To capture this trade-off, we define a welfare function which we refer to as a person's *well-being*, $\mathcal{W}$, as his average felicity over states and time. Our well-being function is thus:

$$\begin{aligned} \mathcal{W} \quad &: = \tfrac{1}{2}E\left[\hat{E}_1[V_1] + \hat{E}_2[V_2]\right] \\ &= \tfrac{1}{2}E\left[(1+\delta)u(w_1) + \hat{E}_1[u(w_2)] + u(w_2)\right] \end{aligned} \tag{3}$$

where the second line follows since $\hat{E}_2[V_2] = V_2$.

We refer to the beliefs that maximize well-being subject to the constraints as optimal beliefs and define them as follows.

Definition 1 *Optimal beliefs are the set of probability distributions defined on the support of the objective distributions that maximize well-being*

$$\mathcal{W} := \tfrac{1}{2}E\left[\hat{E}_1[V_1] + V_2\right]$$

given that actions are optimally chosen given beliefs subject to resource constraints.

In a deterministic setting, this collapses to an example of the social welfare function proposed by Caplin and Leahy (2004). We choose this welfare function for the following reasons. First, as argued in Caplin and Leahy (2000), with anticipatory utility, it is natural to consider a welfare function that is the (weighted) sum of felicities, rather than utilities, over time. Second, and perhaps more important, this choice of $\mathcal{W}$ has the advantage that if expectations were objective and $\delta = 1$, well-being would coincide with the person's expected felicity in each period:

$$\tfrac{1}{2}E[E[V_1] + V_2] = E[u(w_1) + u(w_2)] = E[V_1] = E[V_2].$$

In this case, a person's actions that maximize felicity also maximize well-being, so that any distortion of behavior from that chosen by an objective agent would reduce well-being. Alternatively, when $\delta = 0$,

$$\tfrac{1}{2}E\left[\hat{E}_1[V_1] + V_2\right] \propto u(w_1) + \hat{E}_1[u(w_2)] + E[u(w_2)],$$

and well-being simply adds subjective anticipatory utility to the optimization problem that would be faced in period 1 under rational expectations.[9]

3 The planning fallacy

This section shows the optimality of the planning fallacy when a person is not faced with the option of a deadline. Since choices are only made after η_1 is realized, we simplify exposition by assuming η_1 is a constant. We let η_2 be distributed with mean $E_1[\eta_2] = \eta_1$ and variance $Var_1[\eta_2] > 0$ (with subjective counterparts $\hat{E}_1[\eta_2]$ and $\hat{V}ar_1[\eta_2]$).

The first subsection shows that there is a natural human tendency towards optimism in planning situations and that this optimism causes the planning fallacy: a person's average felicity is increased by a small amount of optimism and is maximized by optimistic beliefs. Further, it is optimal for a person to be overconfident. A second subsection presents the basic experimental evidence of the planning fallacy and summarizes the evidence that we use in the balance of the paper to test our theory.

3.1 Optimal beliefs and behavior

Given beliefs, the person chooses w_1 to maximize the expected present discounted value of the utility flow, $\hat{E}_1[V_1]$, where

$$V_1 := u(w_1) + u(w_2),$$

and $w_2 = \eta_1 + \eta_2 - w_1$. Because utility is concave, people want to smooth work across periods. By certainty equivalence, only the subjective mean matters for behavior. Thus, since people do not discount the future and there is no cost to doing the task later, people do half the work they expect in the first period, or more if the intermediate deadline requires.

Proposition 1 *(Optimal work given deadline and beliefs)*
The person chooses $w_1^* = \frac{1}{2}\left(\eta_1 + \hat{E}_1[\eta_2]\right)$.
In particular, a person with objective beliefs (rational expectations) chooses $w_1^{RE} = \eta_1$.

Our first main result is that some small planning fallacy is always an improvement over objective beliefs. This result establishes that with anticipatory utility, rational expectations are suboptimal. A small degree of optimistic mis-planning is in fact better than objective planning.

[9] Two other points on well-being deserve mentioning. First, we use the objective expectations operator to evaluate well-being because we are interested in maximizing the happiness of the person on average, across realizations of uncertainty. The objective expectation captures this since the actual unfolding of uncertainty is determined by objective probabilities rather than those probabilities that the agent believes. Second, our results do not hinge upon our use of the simple average of felicities; significant generalizations are possible but at the cost of significant complexity.

Proposition 2 *(A small amount of the planning fallacy is beneficial)*
A small degree of optimism increases a person's expected well-being and decreases work in the first period: for any η_1*,* $\frac{d\mathcal{W}}{d\hat{E}_1[\eta_2]}|_{\hat{E}_1[\eta_2]=E_1[\eta_2]} < 0$ *and* $\frac{dw_1^*}{d\hat{E}_1[\eta_2]}|_{\hat{E}_1[\eta_2]=E_1[\eta_2]} > 0$*.*

The proof of this proposition comes directly from a comparison of the increase in the person's felicity in the first period, which benefits from more optimistic anticipation, with the decrease in the second period, which on average suffers the costs of worse planning. A small amount of optimism increases felicity in the first period ($\hat{E}_1[V_1]$) both through a lower level of work effort and through a decrease in the expected disutility of the work in the second period. This is a first-order increase in felicity. When $\delta = 1$, average second-period felicity ($E_1[V_2]$) declines because the work is not perfectly smoothed over the two periods. But this is a second-order cost since it is a small deviation from an action that sets a first-order condition equal to zero. Thus well-being rises with a small amount of optimism. When $\delta < 1$, felicity in the second period places less weight on utility flow from the first period. Thus, a small amount of optimism has first-order benefits for felicity in both the first and second period. When $\delta < 1$, the costs are first order, but still smaller than the first-order benefits to some optimism.

Our second main result is that optimal beliefs and behavior are characterized by the planning fallacy. Thus optimism and the planning fallacy are not only locally but also globally optimal. Further, this *situational* theory of the planning fallacy generates testable predictions.

To reiterate, optimal beliefs maximize well-being, equation (3), given optimal behavior given beliefs, Proposition (1) and equation (1). We denote optimal beliefs by $\hat{E}_1^{**}[\eta_2]$ and $\hat{V}ar_1^{**}[\eta_2]$ and the optimal work they induce as w_1^{**} and w_2^{**}.

Proposition 3 *(The planning fallacy is optimal)*
The agent with optimal beliefs exhibits the planning fallacy:
(i) $\hat{E}_1^{**}[\eta_2] = \frac{1-\delta}{3+\delta}\eta_1 < E_1[\eta_2]$*,*
(ii) $w_1^{**} = \frac{2}{3+\delta}\eta_1 < w^{RE}$*.*

People initially underestimate the amount of work that the project will require and so do less than half the total work in the first period, on average. They gain the benefits of doing little work in the first period and expecting little work in the second. They lose some of the benefits of optimally smoothing effort and on average suffer in the second period when they have more work to do than expected. In general, our 'economic' model predicts the planning fallacy is greater, the greater the anticipatory benefits of optimism and the smaller the ex post costs of mis-planning.

According to our theory, people at some subconscious level know the objective distribution, but choose to be optimistic anyway. This description of the planning fallacy is strikingly similar to the original description of Kahneman and Tversky (1979), quoted at the beginning of the paper.[10]

To conclude this section, we show that, according to our theory, not only should mean beliefs be optimistic, but also people should be overconfident about the precision of their predictions.

[10] In this quote, one can interpret the $E_1[\eta_2]$ and $Var_1[\eta_2]$ as objective learned probabilities given past realized sample values rather than parameters of the true distribution.

That is, in a similar sense as for optimism, our model implies that it would be surprising if people did not exhibit some overconfidence in their predictions.

Proposition 4 *(A small amount of overconfidence is beneficial; complete overconfidence is optimal)*
(i) A small decrease in the perceived uncertainty about future work increases a person's well-being: $\frac{d\mathcal{W}}{d\hat{V}ar_1[\eta_2]}|_{\hat{V}ar_1[\eta_2]=Var_1[\eta_2]} < 0$*;*
(ii) A person's well-being is maximized by the belief that he knows what work level will be required: $\hat{V}ar_1^{**}[\eta_2] = 0 < Var_1[\eta_2]$.

Certainty-equivalence implies that overconfidence has no behavioral consequences. On the other hand, since utility is concave, overconfidence does have anticipatory benefits. Thus the optimal perceived uncertainty is the corner solution of certainty. Our result is extreme because utility is quadratic. But some overconfidence is optimal for a wider range of utility functions.[11]

Even in its extremity, this prediction again closely matches the initial psychological interpretation of the planning fallacy by Kahneman and Tversky. They theorize 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)

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 because of the difficulty in eliciting distributional information. The bias in mean explains both main experimental findings: (i) overconfidence about how often the task will be completed by the predicted completion date, and (ii) underestimation of task completion times on average. The bias in variance makes overconfidence more extreme. In any case, our model delivers both biases.

In Appendix A, we show three comparative statics results. The planning fallacy is worse the more important memory utility is, because fond memory of little work in the past lasts. The planning fallacy is worse the lower the intertemporal elasticity of substitution is, because curvature in the utility function increases the costs of misallocation of work over time.[12] Finally, the planning fallacy is worse the less impatient the agent is, because impatience decreases the importance of anticipatory utility.

3.2 Basic experimental evidence on the planning fallacy

This subsection describes two specific experiments showing the existence of the planning fallacy and then summarizes the set of experimental results that are the basis for the testing of the

[11] For any utility function there must be sufficient curvature in utility relative to marginal utility. The more curved utility is, the larger is the direct increase in well-being for any reduction in variance of future outcomes. The less curved marginal utility is, the smaller the behavioral response to the decrease in uncertainty, from the 'precautionary' channel, and therefore the smaller the indirect decrease in well-being is from changed behavior.

[12] To model different intertemporal elasticities of subsitution (IES), we generalize our utility function to $u(w_t) = -aw_t - \frac{1}{2}w_t^2$, so that the IES is given by $\frac{a+w}{w}$.

model in Section 5. The first specific experiment demonstrates that people tend to underestimate the amount of time it will take them to complete a task. The second shows that people whose effort is forcibly smoothed over time perform better than people simply making decisions based on their own beliefs.

First, Buehler, Griffin, and Ross (1994) report the results of the following experiment (study 2). In the experiment, over 100 undergraduate psychology students filled out a questionnaire asking them to describe an academic task and a non-academic task that they intended to complete in the upcoming week, to predict the amount of time that it would take them to complete each task, and to report how confident they were that they would indeed complete each task by the predicted time. Actual completion times were collected using a follow-up questionnaire one week later and telephone interviews at the end of the semester. For academic (non-academic) tasks, the average subject expected to complete his task in 5.8 (5.0) days and actually completed the task in 10.7 (9.2) days. Only 37 (43) percent of subjects completed their academic (non-academic) tasks in the time they predicted, while they reported that they were 74 (70) percent certain that they would do so.

Importantly, to encourage accuracy in prediction, prior to the experiment, half of the subjects were primed by being told that the purpose of the study was to assess the accuracy of people's predictions. But these subjects exhibited just as optimistic predictions as the other half of the subjects on average.[13]

Second, Ariely and Wertenbroch (2002) report the results of the following experiment (study 2) on the temporal allocation of work. In the experiment, 60 MIT students were paid for the quality and timeliness of three proofreading exercises over a three-week period. Half of the subjects were assigned deadlines that required completion of one exercise per week, the other half simply had a terminal deadline at the end of the three weeks for all three exercises. The former group completed their work more evenly over the three weeks, detected more errors, and earned more money in the experiment (even taking into account losses due to missed deadlines). Thus optimistic beliefs go hand in hand with delay that leads to poorer outcomes.[14]

Many experiments have confirmed these central tendencies – optimistic prediction error and subsequent misallocation of effort – and also documented a set of stylized facts about the planning fallacy. These experiments show that the planning fallacy is not a fixed bias, but is situational. Following our treatment of deadlines in Section 4, we use these experiments in Section 5 to test our model. A very brief summary of the findings of these experiments are:

1. Reported beliefs about task completion times are correlated with actual completion times both across subjects within an experiment and across experiments (see Section 5.1).

2. Many, but not all, different framings and manipulations of the environment that are not payoff-relevant do not eliminate misestimation of task completion times or delay of work (see Section 5.2).

[13] If instead of framing, one actually bases rewards on the accuracy of prediction, then, as we show in section 5.5, our model predicts that the planning fallacy should be mitigated. Buehler, Griffin, and MacDonald (1997), study 2 find that payment for accuracy eliminates the planning fallacy.

[14] We note however that Buehler, Griffin, and Ross (1994) in experiment 5 show a finding not consistent with this link. A framing manipulation that generates more accurate prediction actually slightly increases time to completion. See Section 5.5.

3. The misestimation of task completion times is much less severe or absent in short, in-laboratory experiments (see Section 5.3).

4. When people are paid for rapid task completion, the misestimation of task completion times is more severe and work is completed sooner (see Section 5.4).

5. When people are paid for accurate prediction of task completion times, the misestimation of task completion times is less severe and delay of work is similar (see Section 5.5).

We now return to our theoretical model and develop its predictions for beliefs and behavior when people are given the opportunity to impose intermediate deadlines, and show that the experimental evidence is consistent with the predictions of our model.

4 Deadlines

Returning to our general model, we show that our theory predicts that people choose to impose binding intermediate deadlines, that these deadlines improve the temporal allocation of work, and that they do not require perfect intertemporal smoothing of work. The extant experimental evidence matches each of these predictions.

4.1 Optimal beliefs and behavior

Prior to observing η_1 and choosing w_1, the person chooses ϕ to maximize the expected present discounted value of the utility flow, $\hat{E}[V_1]$, taking into account how w_1 will be chosen as a function of η_1. At this point, the person wants to smooth work effort according to his prior beliefs, so that he would like $w_1(\eta_1) = \frac{1}{2}\left(\eta_1 + \hat{E}[\eta_2|\eta_1]\right)$. If this amount is more than he expects to do absent a deadline, which is the case if $\hat{E}[\eta_2|\eta_1] > \hat{E}_1[\eta_2|\eta_1]$, then he has a time consistency problem stemming from changing beliefs. In this case, if he has the ability to commit to an intermediate deadline, then he will choose a binding intermediate deadline that commits himself to what he believes to be perfect smoothing – that is, to doing in period 1 half of the total work he currently expects.[15]

We now consider optimal beliefs. Both prior and period 1 subjective beliefs are set to maximize well-being, given the choices of deadline and work that these beliefs induce. We begin by considering the situation in which the person chooses his own intermediate deadline to maximize his felicity, and then contrast this to the situation in which an outsider imposes a deadline to maximize the objective expectation of the disutility of work. Finally, we compare both situations to the situation with no deadline studied in Section 3.

First, consider letting the person choose the deadline. After η_1 is realized, if the deadline does not bind, then optimal beliefs and behavior match those of the situation with no deadline. If the deadline binds, then $w_1 = \phi\eta_1$ and optimal beliefs are completely optimistic, $\hat{E}_1[\eta_2] = 0$.

[15] Subsequently, we show that optimal beliefs, $\hat{E}[\eta_2|\eta_1]$ and $\hat{E}_1[\eta_2|\eta_1]$, are linear in η_1. In this case, if $\hat{E}[\eta_2|\eta_1] > \hat{E}_1[\eta_2|\eta_1]$, the optimal deadline is $\phi^* = \hat{E}\left[\frac{\eta_1+\eta_2}{2}|\eta_1\right]/\eta_1$. If $\hat{E}[\eta_2|\eta_1] > \hat{E}_1[\eta_2|\eta_1]$, then the agent chooses a non-binding deadline. If beliefs are objective, the person is indifferent between all deadlines $\phi \in [0, 1]$ since he knows that in period 1 he will choose the optimal amount of work anyway.

Thus, a binding deadline can lead to better smoothing of work effort *and* higher anticipatory utility. Both of these effects increase well-being, thus it is optimal for the agent to impose a binding deadline.[16] Moving back to prior beliefs, in order to induce the person to impose a binding deadline, he must believe that, absent a deadline, he will do insufficient work in period 1. Thus to induce a deadline, the person must hold more realistic beliefs about η_2 before observing η_1 than after. In sum, the person is initially somewhat more realistic and chooses a binding deadline, understanding that without it he would do less work when the time comes. Subsequently, the person is forced by the deadline to better smooth work effort while at the same time he becomes more optimistic about the amount of work required in the future.

Second, consider, an outsider choosing a deadline to maximize the objective expectation of the flow disutility of work, $E[V_1]$. Clearly, the outsider imposes a deadline that requires perfect smoothing of work effort on average. Formally, we have the following proposition.

Proposition 5 *(Self-imposed and externally-imposed deadlines)*
(i) With no deadline ($\phi = 0$), the person is optimistic, is overconfident, and postpones work in period 1;
(ii) With a self-imposed deadline, the person initially believes $\frac{\hat{E}^{**}\left[\hat{E}^{**}[\eta_2|\eta_1]\eta_1\right]}{\hat{E}^{**}[\eta_1\eta_1]} = \frac{3-\delta}{3+\delta}$, imposes a binding deadline, and is more optimistic and postpones less work in period 1 than in case (i);
(iii) With an externally-imposed deadline, the deadline is stricter than in case (ii), the person is equally optimistic in period 1, but the person does not postpone work.

		subjective beliefs *(i) no deadline*		*(ii) self imposed*		*(iii) externally imposed*	*objective beliefs self imposed*
deadline	ϕ	0	$<$	$\frac{3}{3+\delta}$	$\leq$	1	$[0,1]$
period 1	$\hat{E}_1^{**}[\eta_2|\eta_1]$	$\frac{1-\delta}{3+\delta}\eta_1$	$>$	0	$=$	0	η_1
beliefs	$\hat{V}ar_1^{**}[\eta_2|\eta_1]$	0	$=$	0	$=$	0	$Var_1[\eta_2]$
work	w_1^{**}	$\frac{2}{3+\delta}\eta_1$	$<$	$\frac{3}{3+\delta}\eta_1$	$\leq$	η_1	η_1

The central result of part (ii), that it is optimal to choose a deadline that will later bind, stems from an inconsistency in beliefs. The person choosing a deadline thinks that, absent a deadline, in the future he would choose to work too little. This behavior occurs because his prior conditional expectations, $\hat{E}^{**}[\eta_2|\eta_1]$, exceed $\hat{E}_1^{**}[\eta_2|\eta_1] = 0$, his expectations after observing η_1. He may or may not consciously understand that he will procrastinate because he will become more optimistic about the ease of the task. Nevertheless, it is this belief inconsistency, and the agent's awareness of the resulting behavior, that leads to his willingness to overcome his procrastination by setting a binding deadline for himself.

[16] We will show that, since the problem scales in η_1, the deadline either binds or does not for all realizations of η_1.

The novelty of our results relative to extant economic theories of procrastination is that the desire for commitment arises from changing beliefs over time rather than preference inconsistency.[17] Importantly, our result provides a prediction for behavior that differs from that of a preference-based theory in which people are sophisticated. In the latter, a person would choose a deadline that matches the deadline chosen by an objective outside observer while in our theory the person chooses a less strict deadline (provided $\delta > 0$). The next subsection discusses the experimental evidence on this point. Second, in existing models, the desire for commitment is only as strong as the preference inconsistency, while in our setting an arbitrarily small amount of belief inconsistency can lead to a desire for significant commitment. That is, while belief inconsistency is necessary in our setting for the agent to recognize procrastination as a self-control problem, an arbitrarily small amount is sufficient. To see this, note that unconditional beliefs in period 0 are not determined uniquely; they must induce the optimal deadline, which requires only $\frac{\hat{E}^{**}[\eta_2]}{\hat{E}^{**}[\eta_1]} = \frac{3-\delta}{3+\delta}$. Since the only restriction on $\hat{E}[\eta_2]$ is on its ratio to $\hat{E}[\eta_1]$, both can be chosen arbitrarily small, and therefore arbitrarily close to being consistent with $\hat{E}_1^{**}[\eta_2|\eta_1] = 0$.

Turning to part (iii) of Proposition 5, an outsider with rational beliefs chooses a stricter deadline than the agent would commit himself to, because the rational outsider ignores the beneficial effects of belief distortion, and therefore is only interested in smoothing work effort. An outside observer interested in well-being rather than the smoothing of work effort would choose the same deadline as the person himself would.

In the next subsection we present the experimental evidence on the choice of deadlines. These experiments also include information on task performance. While we have interpreted our model as one in which the task is always completed with the same performance, a reasonable interpretation of our model is that one of the costs of poor smoothing is lower task performance (rather than a utility cost alone). That is, the higher the quantity of work done in a period, the lower the quality of the work performed in that period. Similarly, if a large amount of work is left for period 2, a person might actually be late in completing the project and suffer penalties for missing the final 'due date.' Thus we define performance as

$$-\frac{1}{2}\left[w_1^2 + w_2^2\right]$$

so that the smoother the mathematical expectation of the work profile, the higher the average ex post performance.[18]

Since w_1^{**} with no deadline is less than w_1^{**} with a self-imposed deadline which is less than w_1^{**} that optimally smooths work over time, we have the following corollary to Proposition 5.

Corollary 1 *(Task Performance)*
(i) Self-imposed deadlines improve task performance, but do not maximize task performance unless $\delta = 0$;
(ii) Externally-imposed deadlines maximize task performance.

[17] See Strotz (1955-1956) for a general form of preference inconsistency; Phelps and Pollak (1968) and Laibson (1997) for quasi-hyperbolic discounting; Gul and Pesendorfer (2001) for temptation preferences; and Loewenstein, O'Donoghue, and Rabin (2003) for projection bias about future utility.

[18] It is straightforward to analyze task performance in a model in which the agent also cares about his future performance and derives anticipatory utility from it. Indeed, under some assumptions such a model is isomorphic to our model.

4.2 Experimental evidence on intermediate deadlines

First, there is substantial evidence of inconsistency in updating beliefs. For example, Gilovich, Kerr, and Medvec (1993) report that people's average beliefs about exam performance change with temporal distance from the exam, both before and after (prior to receiving a grade).

Second, there is substantial evidence – formal and informal – that people choose deadlines to constrain their future behavior. Informally, there are institutional arrangements such as weight-loss camps, alcohol clinics, and Christmas clubs that people seem to use as commitment devices. Formally, Wertenbroch (1998), Fishbach and Trope (2000), and DellaVigna and Malmendier (2006) all document that people choose to constrain their future behavior.[19]

Most directly related to our model are the experiments of Ariely and Wertenbroch (2002) which study deadlines in the context of academic tasks. The authors conducted two separate studies: one in which subjects had to write three papers during the course of the term for an executive education class, and one in which subjects had to proof-read three texts in the course of three weeks. Subjects were divided into three groups: one that faced externally-imposed, equally-spaced deadlines; one in which subjects could self-impose deadlines; and one that had no deadline option (the last treatment only existed in the proofreading study).

This experimental design closely resembles our theoretical setup. We can interpret η_1 and η_2 as two papers/proofreading exercises (where the experiment has three), and the assumption $E[\eta_2|\eta_1] = \eta_1$ has the interpretation that by completing (at least part of) the first paper, subjects learn the expected difficulty of the second. In our model, a deadline imposes the fraction of the first project completed at the end of period 1, while in the Ariely Wertenbroch experiments it is the date at which one has to complete η_1.

In terms of deadlines, the authors find first that most people given the option to choose a deadline do choose one. Second, the chosen deadlines are on average earlier than the average completion time of the group without deadlines, but later than the exogenously imposed equally-spaced deadlines. This pattern exactly mirrors our results in Proposition 5, $1 \geq \frac{3}{3+\delta} > \frac{2}{3+\delta}$.

In terms of performance, Ariely and Wertenbroch (2002) find that subjects given an externally-imposed (equally-spaced) deadline performed better in the proof-reading assignment (as measured by the course grade in the first study and by the number of errors detected in the second study) than subjects that self-imposed a deadline. The group who had the option to impose a deadline in turn performed better than subjects who had no deadline.[20] Again, this pattern exactly mirrors our results in Proposition 5 and Corollary 1, $\eta_1 \geq \frac{3}{3+\delta}\eta_1 > \frac{2}{3+\delta}\eta_1$.

These experimental results are not due to endogeneity such as the possibility that good

[19] Wertenbroch (1998) reports on an experiment studying the purchasing choices of MBA students across different quantity discounts and goods, as well as on a study of supermarket scanner data on quantity discounts and relative demands. The experiment finds that people tend to purchase goods perceived as more vice-like in smaller quantities despite volume discounts relative to less vice-like goods (e.g. regular vs. low-fat Oreo cookies). Further, more vice-like food items show steeper price declines with volume in the supermarket data. These results suggest that people are inclined to buy smaller quantities of items that they think they might overconsume in the future. Fishbach and Trope (2000) find that participants self-impose penalties for (intentionally) neglecting to undergo minor medical procedures, and that the magnitude of these penalties is correlated with the severity of the procedure. Finally, DellaVigna and Malmendier (2006) document that a sample of gym users would have saved almost fifty percent if they had bought daily passes instead of gym memberships.

[20] Consistent with our mapping between model and experiment, this performance pattern includes penalties for missing deadlines and the final 'due date.'

students select deadlines and also have better performance. First, students were randomly assigned to treatment groups in the proofreading study.[21] Second, in the proofreading study, some students *self-imposed* equally-spaced deadlines. If they were better-than-average proofreaders, then they should have outperformed the random sample of subjects with *externally-imposed* equally-spaced deadlines. But they did not – their performance was similar. However, they did outperform the remaining subjects choosing their own deadlines. This evidence suggests that even within the group with self-imposed deadlines, choosing a tighter deadline leads to better performance.[22]

Grove and Wasserman (2006) also corroborates that externally-imposed deadlines improve performance, based on evidence from a natural experiment that occurred in an undergraduate course at Syracuse University in 1998. Two professors were to each teach one section of the same course with the same problem sets with the same deadlines. One professor factored problem set performance into course grades, while the other did not. One professor started the semester ill, and the other professor started teaching both sections but with this difference in requirements. The ill professor was unable to return, so the same professor taught both classes. Since it was deemed inappropriate to change grading procedures, a natural experiment arose: there were two groups of students participating in a course with the same professor, exams, etc., but facing different consequences for late or incomplete problem sets. Students for whom problem sets were graded – a condition interpretable as having a deadline – did on average one-third of a letter grade better in the course exams than students for whom problem sets were not graded.[23]

As a final caveat, note that though deadlines improve performance on the task to which they apply, it is possible that they decrease performance on other tasks, so that deadlines may or may not be good overall.

We now turn back to testing our theory of the planning fallacy absent deadlines.

5 Tests using experimental evidence on beliefs and behavior

We further test our theory of the planning fallacy (without intermediate deadlines) using five pieces of evidence from extant experiments without deadlines. Employing our model from Section (3), so there is no choice of deadline and η_1 is nonstochastic, we consider: (i) the informativeness of beliefs; (ii) the effects of debiasing; (iii) the differential findings in laboratory vs. field experiments; (iv) the effects of monetary incentives for rapid completion; and (v) the effects of monetary incentives for accurate prediction. This section also demonstrates that reported beliefs are informative because the observed variation is explained by the variation in costs and benefits of optimistic vs. realistic beliefs according to our theory.

[21] In the paper-writing study, there were two sections of the class; one of them was randomly assigned to one treatment and the other section was assigned to the other treatment.

[22] In the context of our model, heterogeneity in the choice of deadline can come from heterogeneity in δ. Those with $\delta = 0$ choose equally-spaced deadlines.

[23] We interpret these results as informative about exogenously imposed deadlines, but it may instead support our predictions about self-imposed deadlines. That is, in contrast to the Ariely and Wertenbroch (2002) study, in this experiment it is possible for students to choose which section to enroll in. Thus, students who desire commitment can choose to enroll in the course with graded problem sets rather than the section without. In either interpretation, the prediction of our model matches the finding.

5.1 The situational nature of beliefs

Absent deadlines, our model predicts that, despite the overoptimism and overconfidence biases, predicted and actual completion times should be closely related.

Proposition 6 *Optimal expected total work, $\hat{E}_1^{**}[\eta_1 + \eta_2]$, is positively correlated with actual total work across tasks with different objectively expected work, $E_1[\eta_1 + \eta_2]$:*

$$\frac{d\hat{E}_1^{**}[\eta_1 + \eta_2]}{dE_1[\eta_1 + \eta_2]} > 0.$$

Experiments find large positive correlations between predicted completion times and actual completion times both across experimental settings and across participants within given experiments.

Across experiments, Roy, Christenfeld, and McKenzie (2005) surveys the results from a large number of experiments dealing with the planning fallacy and shows that predictions relate to actual completion times. Among the eight studies listed in Table 1 of their paper with actual completion times in days (ranging from 1.9 days to 55.5 days), the correlation between the average predicted completion time and the average actual completion time is 0.97. All eight studies find the planning fallacy: the mean prediction is 4.7 days or 26 percent less than the mean actual completion time.[24]

Across people, Buehler, Griffin, and Ross (1994) find the following correlations in their studies. In study 1, the correlation of subjects' predicted completion times and their actual completion times for the completion of their senior thesis is 0.77. This is despite significant bias: the mean predicted time to completion is 33.9 days, 21.7 days shorter than the mean actual completion time. In study 2 (described in the introduction), the correlation is 0.77 for academic tasks and 0.45 for nonacademic tasks. In study 3, dealing with school projects that subjects expected to complete within 2 weeks, the correlation is 0.81.

Strikingly, the informativeness of beliefs is similar to that found in experiments in which manipulations eliminate the planning fallacy. Thus, reported beliefs are not biased due to a lack of subconscious understanding of the objective situation. As in our model, beliefs are biased, but closely related to the true probabilities. In study 4 of Buehler, Griffin, and Ross (1994), subjects were asked to recall and report past experiences with expected and actual completion times for similar assignments in doing a computer project due in 1 or 2 weeks ('recall' manipulation). Interestingly, a random subset of subjects were also asked the following two questions immediately prior to predicting their completion time: 1) when they would complete the task if it was typical of past similar tasks, and 2) to describe a plausible scenario based on past experiences for completion at the typical time ('relevant' manipulation). Subjects receiving the 'relevant' treatment exhibited no planning fallacy: the mean predicted and actual times to completion were the same (both 7.0 days).[25] Yet the correlation of the 'relevant' group's

[24] In the four studies with single-day actual completion times, the correlation is 85 percent, and the mean bias 26 percent. In the four studies with double-day actual completion times, the correlation is 96 percent, and the mean bias 26 percent. We omit one study in which the completion times of others are predicted, and a number of studies for which the only reported statistics are in terms of predicted and actual time before deadline rather than time to completion. The remaining studies are in terms of minutes or hours and are mostly laboratory experiments, which we discuss in Section 5.3.

[25] The random subset of subjects asked just the first question still exhibited the planning fallacy.

predicted and actual completion times (0.75) was similar to those in the other experiments, and only slightly (and statistically insignificantly) higher than the correlation for the 'recall' group (0.66) which exhibited the planning fallacy.

5.2 Framing

Like in canonical economic theory, according to our theory framing and attempts at debiasing should have little effect on observed beliefs and behavior. Our theory is a mapping from the objective environment to subjective beliefs and behavior. And like neoclassical economics, we take objective probabilities as primitives, independent of framing. That said, there is evidence (mostly in psychology and marketing sciences) that framing sometimes can change behavior, particularly in laboratory experiments. One interpretation of the effect of framing is that frames provide information about the objective probability distribution in the experiment. This idea is not inconsistent with our theory if one considers the primitives as post-framing objective probabilities. This said, we maintain the central assumption that objective probabilities are invariant to framing for two reasons: First, in most experiments the tasks either are rehearsed prior to prediction or are common tasks that subjects have experience with (as in school assignments) so that subjects at some level have a good assessment of the true probabilities. Second, we do not have a theory or understanding of when and how frames change people's perhaps unconscious understanding of the truth. Thus, our theory predicts that framing should not eliminate the planning fallacy (provided that framing has little effect on the perceived objective probabilities).

Many framing manipulations intended by experimenters to debias beliefs and change behavior do not. Researchers have found a behaviorally and statistically significant planning fallacy despite the following debiasing techniques: (i) allowing subjects to familiarize themselves with the task at hand (most studies); (ii) asking subjects to describe multiple scenarios for task completion (Newby-Clark, Ross, Buehler, Koehler, and Griffin (2000)); (iii) telling subjects that the "primary purpose of the study was to assess the accuracy of people's predictions" (Buehler, Griffin, and Ross (1994), study 1); (iv) asking subjects to list relevant past experiences, such as the distribution of completion times of similar tasks in the past (Buehler, Griffin, and Ross (1994), study 4 'recall' manipulation; Hinds (1999)); (v) asking subjects to list several reasons for failure and several reasons for successful completion (Sanna and Schwarz (2004)); (vi) asking subjects to list possible surprises that could lead to failure to meet expected completion time (Byram (1997), experiment 1; Hinds (1999)); (vii) asking subjects to list the components of the task (Byram (1997), study 1; Connolly and Dean (1997)); (viii) use of anchoring and adjustment (Byram (1997)); (ix) group prediction of the task completion time of a group (Buehler, Messervey, and Griffin (2005)).

However, the following variations that are not relevant for objective payoffs have been found to largely eliminate the misestimation of task completion times: (i) asking subjects prior to prediction to list past relevant experiences, then asking them when they would complete the studied task if it was typical of these past tasks, and finally asking them to describe a plausible scenario for task completion (Buehler, Griffin, and Ross (1994) study 4 'relevant' manipulation); (ii) asking subjects to make concrete and detailed plans for completing the task (Koole and Spijker (2000) find a decrease in predicted completion times but an even larger decrease in actual completion times).

There are two other relevant experimental findings. First, the planning fallacy seems not to be cultural: it is also present in Japan and Canada (Buehler, Otsubo, Heine, Lehman, and Griffin (1998)). Second, evidence is mixed on whether the planning fallacy is found when subjects predict the completion times of others. Buehler, Griffin, and Ross (1994) (study 5) finds that subjects tend to overpredict the task completion times of others when they observe the others' descriptions of previous experiences (even though these others still underpredict their own task completion dates after listing these experiences). Hinds (1999) finds significant underprediction when predicting the completion times of others who are novices, and that the bias is worse the more experienced is the predictor in the task. Similarly, Byram (1997) (study 3) also finds no mitigation of the planning fallacy.

5.3 Non-onerous tasks

The experimental literature also finds that the planning fallacy is mitigated and even occasionally reversed in short laboratory experiments.

As an example, Burt and Kemp (1994) reports on an experiment in which subjects are asked how long it will take them to do each of five tasks: buy a stamp, find a book in the library, walk from one building to another, complete a one-page form, and sort a deck of cards. The subjects were then timed undertaking the activities, returned to the laboratory, and were asked more questions. Subjects on average overestimated the time it would take for four out of five tasks. As another example, Buehler, Griffin, and MacDonald (1997) (study 2) reports asking subjects how long it will take them to solve an anagram. Subjects (with no other incentives) on average predicted it would take 6.3 minutes when the average completion time ex post was a statistically indistinguishable 6.4 minutes. Byram (1997) (study 5) reports the results of a laboratory experiment in which subjects fold origami. Average predicted time to completion (with no other incentives) was 10.1 minutes, while average actual time to completion was 9.8 minutes. Median predicted time was 7.8 minutes to median actual time 8.8 minutes.

These findings are consistent with our theory if the time spent in these experiments not doing the task is no less onerous than the time spent in these experiments doing the task. It seems unlikely that sitting in a laboratory waiting for other people to solve their anagrams or get back from buying their stamps provides more utility than actually participating in the activities of the experiment oneself.

Formally, consider a situation where disutility comes from the amount of time spent at an experiment in each period, τ_t, and not the time spent working on the activity of the experiment. Suppose further that the time a person spends at an experiment does not depend on his w (or η_2). Then a person's felicity and wellbeing are independent of his beliefs about η_2, and there is no benefit of his holding optimistic beliefs and postponing work. For $u(w_t, \tau_t) = -\frac{1}{2}\tau_t^2$, $\hat{E}_1^{**}[\eta_2] = E_1[\eta_2] = \eta_1$ and $\hat{V}ar_1^{**}[\eta_2] = Var_1[\eta_2]$ maximize wellbeing and $w_1^{**} = w_1^{RE} = \eta_1$.[26]

Thus, in sum, if felicity and well-being do not depend on w, then our model predicts no planning fallacy; if felicity and well-being actually increase in w – some experiments study

[26] Note that since well-being is independent of the smoothing of work, any beliefs about η_2 actually maximize well-being. In this case it seems reasonable to choose that beliefs be rational, but according to our theory, they need not be. Given this indeterminacy, we would expect framing and other utility concerns to have much greater ability to manipulate beliefs in in-laboratory experiments.

activities that are fun rather than onerous – then our model can predict overestimation of task completion times.

An alternative interpretation of the empirical evidence is that the failure of the planning fallacy is due to the short duration of in-laboratory experiments rather than to the fact that they take place in the laboratory. According to this alternative, people underestimate the time involved in tasks that end up being shorter than the average task, as a rational Bayesian might. The evidence however suggests that there is at least more going on than this. Many of the tasks are either common tasks (such as buying a stamp) or actually rehearsed prior to prediction (as in the anagrams experiment in Buehler, Griffin, and MacDonald (1997)).[27]

We provide two pieces of evidence that the lack of anticipatory benefit is why there is no planning fallacy found for in-laboratory experiments. To clearly test our model, we would like a short laboratory experiment in which each person can leave when he completes his task and in which the task is clearly unpleasant. Our first piece of evidence, study 1 in Byram (1997), has these features, but pertains to a slightly longer task than the brief ones just cited.[28] In the study, subjects were asked how long they thought it would take them to build a computer stand. Six months later, a random subsample were asked to predict completion times again, and then to actually build the stand in the laboratory. Prior to each prediction, people were told to read the assembly instructions. Subjects strongly exhibited the planning fallacy. The average predicted completion time was 65.7 minutes six months before (and with only a possibility rather than a certainty of having to build the stand), and was 48.2 minutes right before building the stand. The average actual completion times was 76.1 minutes.

Our second piece of evidence is presented in the next subsection. In a laboratory experiment dealing with a brief task, a control group does not exhibit the planning fallacy. However, a treatment group that receives a payment for rapidly completing the task suffers from the planning fallacy. Thus, even in a short experiment, when there is anticipatory benefit to believing that one will finish quickly, people exhibit the planning fallacy.

5.4 Incentives for speed

Byram (1997) (experiment 5) and Buehler, Griffin, and MacDonald (1997) (study 2) report the results of experiments in which subjects are randomly assigned to a treatment giving them payment for rapid completion of the task.

In the context of our theory, this is a payment that is decreasing in work in the second period, which we model formally as: $P - cw_2$ where $P > 0$ and $c > 0$. Thus, in this treatment, felicity in the first period is given by the current level of work effort and the anticipation of both future work effort and the payment:

$$\hat{E}_1 [V_1] = u (w_1) + \hat{E}_1 [u (w_2) + (P - cw_2)] .$$

Felicity in the second period is given by realized outcomes:

$$V_2 = \delta u (w_1) + u (w_2) + (P - cw_2) .$$

[27] It is possible to combine both theories. Optimal beliefs would lead to an underestimation bias from the baseline Bayesian overestimation of task completion times.

[28] We infer that each person could leave when finished because "each subject was tested individually" when actually building the computer stand.

Well-being continues to be the simple average of both felicities.

The additional penalty for low work levels in the first period implies that it is optimal given objective beliefs to do more work in the first period.

$$w_1^* = (1/2)(\eta_1 + \hat{E}_1[\eta_2] + c)$$

But the additional cost of leaving work to the future implies there is an additional incentive to believe that there will be little work to do in the future. Thus, there is also a tendency to be more optimistic and mis-plan by more. In sum, incentives for rapid completion increase mis-planning and increase the difference in initial effort between an agent with objective beliefs and the agent with optimal expectations.

Proposition 7 *Incentives for speed increase the planning fallacy:* $\frac{d\hat{E}_1^{**}[\eta_2]}{dc} \leq 0$ *and* $\frac{dw_1^{RE}}{dc} \geq \frac{dw_1^{**}}{dc} > 0$.

Experiments find that the plannning fallacy worsens with incentives for speed.

In Byram (1997) (experiment 5), subjects were given folding instructions for origami and asked to make predictions about their median time to completion.[29] Then they were given the materials for the origami figure and then asked to complete it. A randomly selected treatment group was also given explicit incentives for rapid completion prior to making their predictions. Subjects were paid $4 for finishing in the top quartile, $2 for finishing in the next quartile, $1 for finishing in the second quartile, and nothing for finishing in the bottom quartile. The control group was paid $3.

For the control group the median prediction time was 7.8 minutes and the median actual time was 8.8 minutes, and for the treatment group the median prediction time was 5.0 minutes and the median actual time was 7.8 minutes. Thus, the incentive for speed raised the prediction error by 180 percent and decreased the actual time to completion by 11 percent. Average prediction and actual completion times also imply that the incentive worsened the prediction error, but using averages the control group exhibited no planning fallacy and the incentives did not increase average actual completion time.

Buehler, Griffin, and MacDonald (1997) (study 1) reports the results of telephone surveys of household expectations and actual behavior about tax filing. The study finds that people who expected a refund, and therefore had a monetary incentive to finish their returns earlier, expected to finish their returns on average 12 days before they actually did, while people who expected to owe taxes expected to finish their returns on average only 4 days before they actually did. In the end, the average filing times of both groups were almost the same – the difference was a statistically insignificant 3 days.

Since refund status is potentially endogenous, in study 2, Buehler, Griffin, and MacDonald (1997) study a laboratory experiment in which subjects complete anagram puzzles. In findings similar to Byram (1997), a randomly-selected treatment group that is given payment for rapid completion shows greater bias in prediction and more rapid completion of the anagrams.

[29] A best guess described as "half the time will be faster and half the time will be slower."

5.5 Incentives for accuracy of beliefs

Buehler, Griffin, and MacDonald (1997) (study 2) also report the results of an experiment in which subjects are randomly assigned to a treatment giving them payment for accurate prediction of task completion times.

In the context of our theory, this is a payment that is decreasing in the error in predicting the total work involved in a project, which we model formally as: $P - k\left(\hat{E}_1[\eta_2] - \eta_2\right)^2$ where $P > 0$ and $k > 0$. Thus, in this experiment, felicity in the first period is given by the current level of work effort and the anticipation of both future work effort and the payment:

$$\hat{E}_1[V_1] = u(w_1) + \hat{E}_1\left[u(w_2) + \left(P - k\left(\hat{E}_1[\eta_2] - \eta_2\right)^2\right)\right]$$

In the second period felicity is given by realized outcomes:

$$V_2 = \delta u(w_1) + u(w_2) + \left(P - k\left(\hat{E}_1[\eta_2] - \eta_2\right)^2\right).$$

Well-being continues to be the simple average of both felicities.

Since the payment is based only on beliefs (and η_2), there is no effect on behavior given beliefs. But on average the ex post loss associated with incorrect prediction gives an additional cost to overly optimistic beliefs, and thus beliefs are more objective.

Proposition 8 *Incentives for accuracy of prediction decrease the planning fallacy:* $\frac{d\hat{E}_1^{**}[\eta_2]}{dk} \geq 0$, *and* $\frac{dw_1^{**}}{dk} \geq \frac{dw_1^{RE}}{dk} = 0$.

Again, the experiment confirms this prediction for the pattern of beliefs. However, there is no evidence that the tasks are completed sooner.

Buehler, Griffin, and MacDonald (1997) (study 2) report the results of an experiment on prediction and task completion done on undergraduates. Subjects were given anagrams to complete, and were asked to practice by doing two puzzles, each of which was observed to take 5 to 7 minutes. Then all subjects were given an incentive for speed, as described in the previous subsection.[30] Next, a random subsample was given an incentive for accurate prediction in addition: \$2 if the predicted completion time was within 1 minute of the actual time and an additional \$2 if the predicted time was also within 30 seconds of the actual time. Then all subjects were asked to predict their time to complete an anagram puzzle and then to actually complete the anagram.

For students without the incentive for accurate prediction, the average predicted time to completion was 4.1 minutes and the average actual time was 5.4 minutes, and for the treatment group with the incentive the average predicted time was 5.8 minutes and the average actual

[30] They also report the results of this experiment on students who are not paid for rapid completion. We have argued that there is little incentive to have the planning fallacy in this situation. In this case, the optimizing agent should predict a long completion time and simply finish slowly to meet his predicted completion time in order to maximize his payment. The experiment finds that indeed these incentives lead to longer prediction times and longer actual time to completion, and even to overestimation of completion times.

time was 5.5 minutes.[31] Thus the incentive for accuracy decreased the bias in the expectation of completion time so as to eliminate the planning fallacy. Our theory also predicts that the incentive should reduce actual time to completion, but there is only a trivial difference in average actual completion time.

6 Concluding discussion

In this paper, we develop an economic model of the planning fallacy based on the distortion of subjective probabilities. As in the original description of the planning fallacy, people tend to postpone work because they hold overoptimistic beliefs about the ease of the task at hand. The strength of our approach is that these belief biases are situational, and so our model makes predictions about when the planning fallacy is mitigated or exacerbated.

But like much recent work in behavioral economics, and unlike most research in mainstream economics, biases in beliefs are central to the understanding of behavior, and 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-sample, and a parsimonious model that better represents actual beliefs and utilities is likely to perform better in such an exercise.

Thus, we replace the discipline of the assumption of rationality with the discipline of data by testing our model using reported beliefs. 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. That is, we subject the predictions of our model to testing using subjective expectations reported in experimental and nonexperimental settings. In the experimental settings, we observe causation from environment to reported beliefs that is consistent with our model and inconsistent with objective probability assessments. Reported expectations respond to incentives in the ways predicted by our theory. In sum, the model is consistent with much of the experimental evidence on both mis-planning and on the use and effects of deadlines.

[31] Values are estimates based on Figure 2, page 243.

References

Akerlof, G. (1991): "Procrastination and Obedience," *American Economic Review (Papers and Proceedings)*, 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.

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.

Buehler, R., D. Griffin, and H. MacDonald (1997): "The role of motivated reasoning in task predictions," *Personality and Social Psychology Bulletin*, 23, 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), 366–381.

Buehler, R., D. Messervey, and D. Griffin (2005): "Collaborative Planning and Prediction: Does Group Discussion Affect Optimistic Biases in Time Estimation," *Organizational Behavior and Human Decision Processes*, 97, 46–63.

Buehler, R., Y. Otsubo, S. Heine, D. Lehman, and D. Griffin (1998): "Culture and optimism. The planning fallacy in Japan and North America," *Wilfrid Laurier University manuscript*.

Burt, C. D. B., and S. Kemp (1994): "Construction of Activity Duration and Time Management Potential," *Applied Cognitive Psychology*, 8, 155–68.

Byram, S. J. (1997): "Cognitive and Motivational Factors Influencing Time Prediction," *Journal of Experimental Psychology: Applied*, 3(3), 216–39.

Caplin, A. J., and J. Leahy (2000): "The Social Discount Rate," NBER Working Paper 7983.

——— (2004): "The Social Discount Rate," *Journal of Political Economy*, 112(6), 1257–1268.

Connolly, T., and D. Dean (1997): "Decomposed versus holistic estimates of effort required for software writing tasks," *Management Science*, 43, 1029–45.

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," *The Review of Economics and Statistics*, 80(3), 374–88.

Fischer, C. (1999): "Read This Paper Even Later: Procrastination with Time-Inconsistent Preferences," *Resources for the Future, Discussion Paper 99-20*.

Fishbach, A., and Y. Trope (2000): "Counteractive Self-control in Overcoming Temptation," *Journal of Personality and Social Psychology*, 79, 493–506.

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.

Grove, W. A., and T. Wasserman (2006): "Incentives and Student Learning: Results from a Natural Experiment," *American Economic Review (Papers and Proceedings)*, 96(2), 447–452.

Gul, F., and W. Pesendorfer (2001): "Temptation and Self-Control," *Econometrica*, 69(6), 1403–1435.

——— (2004): "Self-Control and the Theory of Consumption," *Econometrica*, 72(1), 119–158.

Hamermesh, D. S. (1985): "Expectations, Life Expectancy, and Economic Behavior," *The Quarterly Journal of Economics*, 100(2), 389–408.

Harris, C., and D. Laibson (2001): "Dynamic Choices of Hyperbolic Consumers," *Econometrica*, 69(4), 935–57.

Hinds, P. J. (1999): "The curse of expertise: The effects of expertise and debiasing methods on predictions of novice performance," *Journal of Experimental Psychology: Applied*, 5, 205–21.

Kahneman, D., and A. Tversky (1979): "Intuitive Prediction: Biases and corrective procedures," in *Forecasting*, ed. by S. Makridakis, and S. C. Wheelwright, pp. 313–27. North-Holland, TIMS Studies in Management Science 12.

Koole, S., and M. V. Spijker (2000): "Overcoming the planning fallacy through willpower: Effects of implementation intentions on actual and predicted task completion times," *European Journal of Social Psychology*, 30, 873–88.

Laibson, D. (1997): "Golden Eggs and Hyperbolic Discounting," *Quarterly Journal of Economics*, 112(2), 443–477.

Liberman, N., and Y. Trope (2003): "Temporal Construal," *Psychological Review*, 110, 401–21.

Loewenstein, G. (1987): "Anticipation and the Valuation of Delayed Consumption," *Economic Journal*, 47(387), 666–684.

Loewenstein, G., T. O'Donoghue, and M. Rabin (2003): "Projection Bias In Predicting Future Utility," *Quarterly Journal of Economics*, 118, 1209–1248.

Mankiw, N. G., R. Reis, and J. Wolfers (2004): "Disagreement about Inflation Expectations," in *N.B.E.R. Macroeconomics Annual 2003*, ed. by M. Gertler, and K. Rogoff, vol. 18, pp. 294–48. Cambridge: MIT Press.

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

Newby-Clark, I. R., M. Ross, R. Buehler, D. J. Koehler, and D. Griffin (2000): "People focus on optimistic and disregard pessimistic scenarios while predicting their task completion times," *Journal of Experimental Psychology: Applied*, 6, 171–182.

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 (1999a): "Doing It Now or Later," *American Economic Review*, 89(1), 103–124.

——— (1999b): "Incentives For Procrastinators," *The Quarterly Journal of Economics*, 114(3), 769–816.

——— (2001): "Choice and Procrastination," *Quarterly Journal of Economics*, 116(1), 121–160.

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.

Sanna, L. J., and N. Schwarz (2004): "Integrating Temporal Biases. The Interplay of Focal Thoughts and Accessibility Experiences," *Psychological Science*, 15(7), 474–481.

Savage, L. J. (1954): *The Foundation of Statistics*. Wiley, New York.

Souleles, N. S. (2004): "Expectations, Heterogeneous Forecast Errors, and Consumption: Micro Evidence from the Michigan Consumer Sentiment Surveys," *Journal of Money, Credit, and Banking*, 36(1), 39–72.

Strotz, R. H. (1955-1956): "Myopia and Inconsistency in Dynamic Utility Maximization," *Review of Economic Studies*, 23(3), 165–180.

Wertenbroch, K. (1998): "Consumption Self-control by Rationing Purchase Quantities of Virtue and Vice," *Marketing Science*, 17, 317–337.

Yariv, L. (2002): "I'll See It When I Believe It – A Simple Model of Cognitive Consistency," *Mimeo, UCLA*.

Appendices

A Comparative statics

In this appendix we show three comparative statics results. The more important is memory utility, the worse is the planning fallacy, because fond memory of little work in the past lasts. The lower is the intertemporal elasticity of substitutionm the worse is the planning fallacy, because curvature in the utility function increases the costs of misallocation of work over time. Finally, the less impatient is the agent, the worse is the planning fallacy, because impatience decreases the importance of anticipatory utility.

A.0.1 Memory

The continued enjoyment of past experiences or continued unhappiness from past suffering, is almost as central to the human experience as anticipating or dreading future events. The planning fallacy relates to the strength of enjoyment of past memories because memory utility reduces the costs of over-postponing work. The more a person dwells on the past, the greater are the benefits of optimism and little work in the first period. The utility benefits of this enjoyment last. Further, the strength of memory utility does not affect behavior given beliefs. Thus memory utility increases the benefits to optimistic beliefs.

Following the preceding logic, the more important memory utility, captured by our parameter δ, the more severe the planning fallacy.

Proposition 9 *(The planning fallacy and memory utility)*
The planning fallacy becomes worse the more the agent cares about the past: as δ increases, optimism increases, $\frac{d\hat{E}_1^{**}[\eta_2]}{d\delta} \leq 0$, and work decreases, $\frac{dw_1^{**}}{d\delta} \leq 0$.

A formal proof follows directly from differentiation of the results in Proposition 3.

It is worth noting that even for $\delta = 0$, when memory utility is weakest, people are still optimistic and exhibit the planning fallacy. Thus, memory utility increases the incentives to engage in wishful thinking, but it is not necessary.

A.0.2 Intertemporal substitution

For our quadratic utility function, we can vary the importance of the curvature of the utility function by introducing a linear disutility in effort term as

$$u(w_t) = -aw_t - \frac{1}{2}w_t^2.$$

The intertemporal elasticity of substitution (IES) is

$$IES = \frac{a+w}{w}$$

which is increasing in a at any work level.

From the agent's perspective in the first-period, work is optimally smoothed so that one half the expected work is done in the first period. Further, she believes that she faces no risk. Thus curvature is irrelevant for the actions and felicity of the agent in the first period. Curvature does however affect the realized utility in the second period, and therefore enters well-being and influences beliefs. The less curved utility, the less costly is not smoothing, and the worse the planning fallacy. Thus we have the following result.

Proposition 10 *(The planning fallacy and the intertemporal elasticity of substitution)*
The planning fallacy becomes worse the lower the curvature of the utility function: as a increases, optimism increases, $\frac{d\hat{E}_1^{**}[\eta_2]}{da} \leq 0$*, and work decreases,* $\frac{dw_1^{**}}{da} \leq 0$*.*

A.0.3 Impatience

To examine formally the role of impatience, we add a discount factor to preferences, β, so that the person's felicity in period 1 is

$$V_1 := u(w_1) + \beta u(w_2)$$

where $0 < \beta \leq 1$ (felicity in period 2 is unchanged because there is no future to discount). The property that the behavior that maximizes $E_1[V_1]$ is also the behavior that maximizes $\mathcal{W}$ under objective beliefs, requires both $\delta = \beta^{-1}$ and

$$\mathcal{W} := \tfrac{1}{1+\beta}\left(\hat{E}_1[V_1] + \beta E[V_2]\right) = \tfrac{1}{1+\beta}\left((1+\beta\delta)\,u(w_1) + \beta\hat{E}[u(w_2)] + \beta E[u(w_2)]\right).$$

While we employ this definition of well-being, we assume only that $0 \leq \delta \leq \beta^{-1}$ so that impatience can vary independently from the strength of memory utility and we do not confound their effects. Finally, to ensure an interior optimum, we replace the assumption $E[\eta_2|\eta_1] = \eta_1$ with *{{{I haven't derived the new conditions for these, since I don't know if we really want to keep this proposition. WE DO BUT WITH a=0}}}*

We first note that our previous results on optimism, overconfidence and the planning fallacy continue to hold. That is, with some restrictions on parameters, the agent with optimal beliefs exhibits the planning fallacy and is overconfident, and the planning fallacy becomes worse the more the person cares about the past and the lower the curvature of the utility function. These results are stated in Proposition 12 in Appendix B.11.

Turning to the effect of discounting, greater impatience (smaller β) implies that the benefits of optimism are lower since anticipatory utility is less important in first-period felicity. The person who cares less about the future has less of an incentive to be optimistic about it.

Proposition 11 *(Discounting)*
The more impatient the agent, the greater the planning fallacy: $\frac{d\hat{E}_1^{**}[\eta_2]}{d\beta} \geq 0$ *and* $\frac{dw_1^{**}}{d\beta} > 0$*.*

B Proofs of propositions

B.1 Proof of Proposition 1

Substituting the constraint into the objective, we get

$$\max_{w_1} \hat{E}_1 \left[u\left(w_1\right) + u\left(\eta_1 + \eta_2 - w_1\right) \right]$$

which has first-order condition

$$u'\left(w_1^*\right) = \hat{E}_1 \left[u'\left(\eta_1 + \eta_2 - w_1^*\right) \right].$$

The certainty equivalence property of quadratic utility then implies that

$$w_1^* = \hat{E}_1 \left[\eta_1 + \eta_2 - w_1^* \right]$$

which reorganizes to our result, once the deadline constraint is imposed. The second order condition is satisfied. Note that $0 \leq \hat{E}_1 \left[\eta_2 | \eta_1\right] \leq \eta_1$ and $0 \leq E\left[\eta_2 | \eta_1\right] \leq \eta_1$ guarantee that $w_1^* \in [0, \eta_1]$ which implies $w_2^* \geq 0$ by the constraint.

B.2 Proof of Proposition ??

As can be seen from Proposition 1, assuming that $\hat{E}_1 \left[\eta_2 | \eta_1\right]$ is linear in η_1 implies that the optimal work chosen in period 1 is also linear in η_1. As a result, a deadline ϕ of the form $w_1 \geq \phi \eta_1$ is either binding or non-binding for all values of η_1. In particular, the deadline ϕ is always binding if $\phi \geq \frac{1}{2}\left(1 + \frac{\hat{E}_1[\eta_2|\eta_1]}{\eta_1}\right)$.

The agent chooses the deadline that maximizes $\hat{E}\left[V_1\right]$, which substituting for $w_1 = \phi \eta_1$ becomes

$$-\frac{1}{2}\left\{ \hat{E}\left[\phi^2 \eta_1^2\right] + \hat{E}\left[\left(\eta_1 + \eta_2 - \phi \eta_1\right)^2\right] \right\}.$$

The First Order Condition with respect to ϕ yields, under the assumption that $\hat{E}\left[\eta_2 | \eta_1\right]$ is linear in η_1, that

$$\phi^* = \frac{1}{2}\left(1 + \frac{\hat{E}\left[\eta_2 | \eta_1\right]}{\eta_1}\right).$$

This deadline is binding if $\hat{E}\left[\eta_2 | \eta_1\right] > \hat{E}_1\left[\eta_2 | \eta_1\right]$.

B.3 Proof of Proposition 2

Substituting the optimal actions from Proposition 1 into the objective function for beliefs gives the well-being

$$\begin{aligned}
\mathcal{W} &= \tfrac{1}{2} E\left[(1+\delta)\, u\left(w_1\right) + \hat{E}_1\left[u\left(w_2\right)\right] + u\left(w_2\right)\right] \\
&= \tfrac{1}{2} E\left[(1+\delta)\, u\left(\frac{1}{2}\left(\eta_1 + \hat{E}_1\left[\eta_2\right]\right)\right) + \hat{E}_1\left[u\left(\frac{1}{2}\eta_1 + \eta_2 - \frac{1}{2}\hat{E}_1\left[\eta_2\right]\right)\right] + u\left(\frac{1}{2}\eta_1 + \eta_2 - \frac{1}{2}\hat{E}_1\left[\eta_2\right]\right)\right] \\
&= -\tfrac{1}{4} E\left[\frac{1}{4}(3+\delta)\,\eta_1^2 + \eta_2^2 + \eta_1\eta_2 + \hat{E}_1\left[\eta_2^2\right] - \frac{1}{4}(1-\delta)\left(\hat{E}_1\left[\eta_2\right]\right)^2 + \frac{1}{2}(1+\delta)\,\eta_1 \hat{E}_1\left[\eta_2\right] - \eta_2 \hat{E}_1\left[\eta_2\right]\right].
\end{aligned}$$

For any η_1 we have

$$\frac{d\mathcal{W}}{d\hat{E}_1[\eta_2]} = -\tfrac{1}{4}\left[\frac{1}{2}(3+\delta)\hat{E}_1[\eta_2] + \frac{1}{2}(1+\delta)\eta_1 - E_1[\eta_2]\right],$$

which evaluated at $\hat{E}_1[\eta_2] = E_1[\eta_2]$ becomes

$$\frac{d\mathcal{W}}{d\hat{E}_1[\eta_2]}|_{\hat{E}_1[\eta_2]=E_1[\eta_2]} = -\frac{1}{8}(1+\delta)[\eta_1 + E_1[\eta_2]] < 0$$

From Proposition 1 we also have

$$w_1^* = \frac{1}{2}\left(\eta_1 + \hat{E}_1[\eta_2]\right).$$

Clearly, for any η_1 we have

$$\frac{dw_1^*}{d\hat{E}_1[\eta_2]}|_{\hat{E}_1[\eta_2]=E_1[\eta_2]} = \frac{1}{2} > 0.$$

B.4 Proof of Proposition 3

Substituting the optimal actions from Proposition 1 into the objective function for beliefs gives the well-being

$$\mathcal{W} = -\tfrac{1}{4}E\left[\frac{1}{4}(3+\delta)\eta_1^2 + \eta_2^2 + \eta_1\eta_2 + \hat{E}_1\left[\eta_2^2\right] - \frac{1}{4}(1-\delta)\left(\hat{E}_1[\eta_2]\right)^2 + \frac{1}{2}(1+\delta)\eta_1\hat{E}_1[\eta_2] - \eta_2\hat{E}_1[\eta_2]\right]$$

Substituting for $E_1[\eta_2] = \eta_1$, the First Order Condition for $\hat{E}_1^{**}[\eta_2]$ is

$$\hat{E}_1^{**}[\eta_2] = \frac{1-\delta}{3+\delta}\eta_1.$$

Substituting this into Proposition 1, we have

$$w_1^{**} = \tfrac{2}{3+\delta}\eta_1.$$

Clearly, $\hat{E}_1^{**}[\eta_2] = \frac{1-\delta}{3+\delta}\eta_1 < \eta_1 = E_1[\eta_2]$ and $w_1^{**} = \frac{2}{3+\delta}\eta_1 < \eta_1 = w^{RE}$.

B.5 Proof of Proposition 4

(i) By Proposition 1, the effort choice of the agent depends only on beliefs about means, not variances. Substituting the optimal actions from Proposition 1 into the objective function for beliefs gives

$$\mathcal{W} = \tfrac{1}{2}E\left[(1+\delta)\,u\left(\frac{1}{2}\left(\eta_1 + \hat{E}_1\left[\eta_2\right]\right)\right) + \hat{E}_1\left[u\left(\frac{1}{2}\eta_1 + \eta_2 - \frac{1}{2}\hat{E}_1\left[\eta_2\right]\right)\right] + u\left(\frac{1}{2}\eta_1 + \eta_2 - \frac{1}{2}\hat{E}_1\left[\eta_2\right]\right)\right].$$

The subjective variance of η_2 after observing η_1 affects only the middle term of this objective. By the concavity of the utility function, this term, and thus the objective, is decreasing in $\hat{V}ar_1\left[\eta_2\right]$. Therefore, $\frac{d\mathcal{W}}{d\hat{V}ar_1[\eta_2]}|_{\hat{V}ar_1[\eta_2]=Var_1[\eta_2]} < 0$.

(ii) Clearly $\frac{d\mathcal{W}}{d\hat{V}ar_1[\eta_2]} < 0$ for all values of $\hat{V}ar_1\left[\eta_2\right]$, and so $\hat{V}ar_1^{**}\left[\eta_2\right] = 0$.

B.6 Proof of Proposition 5

In this proposition we contrast the predictions of three variations of the model's setup: the case where the person does not have the option to choose a deadline, the case where the person self-imposes a deadline, and the case where the deadline is externally-imposed by an objective observer. In each of these variations, the solution consists of finding the optimal actions w_1 and w_2, the optimal deadline ϕ (set to 0 in the case where choosing a deadline is not an option), and the optimal beliefs, in particular, prior unconditional beliefs $\hat{E}\left[\eta_t\right]$, $\hat{V}ar\left[\eta_t\right]$, for $t = 1, 2$, and $\hat{C}ov\left[\eta_1, \eta_2\right]$, prior conditional beliefs $\hat{E}\left[\eta_2|\eta_1\right]$ and $\hat{V}ar\left[\eta_2|\eta_1\right]$, and posterior conditional beliefs $\hat{E}_1\left[\eta_2\right]$ and $\hat{V}ar_1\left[\eta_2\right]$.

In all three variations of the model's setup, the optimal work given arbitrary deadline and beliefs is given by Proposition 1 as $w_1^* = \max\left\{\phi\eta_1, \frac{1}{2}\left(\eta_1 + \hat{E}_1\left[\eta_2|\eta_1\right]\right)\right\}$ and $w_2^* = \eta_1 + \eta_2 - w_1^*$. In what follows we look separately at the three variations.

(i) In the variation of the model where the person does not have the option to choose a deadline, optimal beliefs are given by Propositions 3 and 4 as $\hat{E}_1^{**}\left[\eta_2\right] = \frac{1-\delta}{3+\delta}\eta_1$ and $\hat{V}ar_1^{**}\left[\eta_2\right] = 0$. The choice of the other beliefs is arbitrary, since they do not affect well-being.

(ii) In the variation where the person self-imposes a deadline, we need to find the optimal deadline given subjective beliefs and the optimal beliefs given this deadline. Note that the optimal deadline given beliefs is not given by Proposition **??**, since that result only holds under the assumption that beliefs are linear in η_1. Fortunately, what interests us is not to find the optimal deadline for arbitrary beliefs, but only to find the optimal deadline for the optimal beliefs, and so we proceed as follows. First, we solve a modified problem, i.e., we find the actions, w_1 and w_2, and beliefs, $\hat{E}_1\left[\eta_2\right]$ and $\hat{V}ar_1\left[\eta_2\right]$, that maximize the well-being, subject only to the primitive constraints on their values and to the resource constraint that requires the task to be completed at the end of the second period; this is a less constrained version of the original problem, since it does not require that behavior is optimal given beliefs. Then we show that the w_1, w_2, $\hat{E}_1\left[\eta_2\right]$, and $\hat{V}ar_1\left[\eta_2\right]$ that solve the modified problem in fact satisfy the additional constraints of the original problem, for appropriate choices of deadline and prior unconditional beliefs. As a result, we have found one of the possible solutions of the original problem. Finally, we show that there are no other solutions to the original problem.

So first we solve the modified problem, i.e., we find $w_1^{\diamondsuit}$, $w_2^{\diamondsuit}$, $\hat{E}_1^{\diamondsuit}[\eta_2]$ and $\hat{V}ar_1^{\diamondsuit}[\eta_2]$ that solve

$$\max_{w_1, w_2, \hat{E}_1[\eta_2], \hat{V}ar_1[\eta_2]} \mathcal{W} \equiv \frac{1}{2} E\left[\hat{E}[V_1] + V_2\right]$$

subject to

$$w_1 + w_2 = \eta_1 + \eta_2,$$

and subject to the primitive constraints on w_1, w_2, $\hat{E}_1[\eta_2]$ and $\hat{V}ar_1[\eta_2]$.

Substituting for V_1 and V_2, and using the constraint $w_1 + w_2 = \eta_1 + \eta_2$, the objective becomes

$$\begin{aligned}
\mathcal{W} &= \tfrac{1}{2} E\left[(1+\delta)\, u(w_1) + \hat{E}_1[u(w_2)] + u(w_2)\right] \\
&= -\tfrac{1}{4} E\Big[(1+\delta)\, w_1^2 + \eta_1^2 + \hat{E}_1[\eta_2^2] + w_1^2 + 2\eta_1 \hat{E}_1[\eta_2] - 2\eta_1 w_1 - 2\hat{E}_1[\eta_2]\, w_1 + \eta_1^2 + \eta_2^2 + w_1^2 + 2\eta_1\eta_2 - 2\eta_1 u \\
&= -\tfrac{1}{4} E\Big[2\eta_1^2 + \eta_2^2 + 2\eta_1\eta_2 + (3+\delta)\, w_1^2 - 4\eta_1 w_1 - 2\eta_2 w_1 + \left(\hat{E}_1[\eta_2]\right)^2 + \hat{V}ar_1[\eta_2] + 2\eta_1 \hat{E}_1[\eta_2] - 2w_1 \hat{E}_1[\eta_2 \\
&= -\tfrac{1}{4} E\Big[2\eta_1^2 + E[\eta_2^2|\eta_1] + 2\eta_1 E[\eta_2|\eta_1] + (3+\delta)\, w_1^2 - 4\eta_1 w_1 - 2E[\eta_2|\eta_1]\, w_1 + \left(\hat{E}_1[\eta_2]\right)^2 + \hat{V}ar_1[\eta_2] + 2\eta \\
&= -\tfrac{1}{4} E\left[5\eta_1^2 + Var[\eta_2|\eta_1] + (3+\delta)\, w_1^2 - 6\eta_1 w_1 + \left(\hat{E}_1[\eta_2]\right)^2 + \hat{V}ar_1[\eta_2] + 2\eta_1 \hat{E}_1[\eta_2] - 2w_1 \hat{E}_1[\eta_2]\right]
\end{aligned}$$

It is obviously optimal to set $\hat{V}ar_1^{\diamondsuit}[\eta_2] = 0$.

Ignoring any constraints, the first-order conditions w.r.t. w_1 and $\hat{E}_1[\eta_2]$ yield

$$\begin{aligned}
w_1^{\diamondsuit} &= \frac{1}{3+\delta}\left(3\eta_1 + \hat{E}_1^{\diamondsuit}[\eta_2]\right) \\
\hat{E}_1^{\diamondsuit}[\eta_2] &= w_1^{\diamondsuit} - \eta_1.
\end{aligned}$$

Imposing the constraint $\hat{E}_1[\eta_2] \geq 0$ {{{Do we also need to impose a constraint $w_1 \leq \eta_1$?}}} the above becomes

$$\begin{aligned}
w_1^{\diamondsuit} &= \frac{3}{3+\delta}\eta_1 \\
\hat{E}_1^{\diamondsuit}[\eta_2] &= 0.
\end{aligned}$$

Now we show that the $w_1^{\diamondsuit}$, $\hat{E}_1^{\diamondsuit}[\eta_2]$, and $\hat{V}ar_1^{\diamondsuit}[\eta_2]$ we have found, together with appropriate choices for deadline ϕ and beliefs $\hat{E}[\eta_t]$ and $\hat{V}ar[\eta_t]$ for $t = 1, 2$, and $\hat{C}ov[\eta_1, \eta_2]$ solve the original problem. That is, we show that there exist $\phi^{\dagger\dagger}$ and $\hat{E}^{\dagger\dagger}[\eta_t]$ and $\hat{V}ar^{\dagger\dagger}[\eta_t]$ for $t = 1, 2$, and $\hat{C}ov^{\dagger\dagger}[\eta_1, \eta_2]$ such that: i) Together with $\hat{E}_1^{\dagger\dagger}[\eta_2] = \hat{E}_1^{\diamondsuit}[\eta_2]$ and $\hat{V}ar_1^{\dagger\dagger}[\eta_2] = \hat{V}ar_1^{\diamondsuit}[\eta_2]$ beliefs are optimal given optimal deadline choice and optimal action choice, ii) Deadline $\phi^{\dagger\dagger}$ is the optimal deadline given these beliefs and optimal action choice, and iii) Actions $w_1^{\dagger\dagger} = w_1^{\diamondsuit}$ and $w_2^{\dagger\dagger} = \eta_1 + \eta_2 - w_1^{\dagger\dagger}$ are the optimal actions given deadline $\phi^{\dagger\dagger}$ and the aforementioned beliefs.

First, substituting $\hat{\mu}_{2|1}^{\dagger\dagger} = \hat{\mu}_{2|1}^{\diamondsuit} = 0$ in (**??**) and (**??**) we have

$$\begin{cases} w_1^{\dagger\dagger} = \frac{1}{2}\eta_1 \\ w_2^{\dagger\dagger} = \frac{1}{2}\eta_1 + \eta_2 \end{cases}$$

if the deadline does not bind, and

$$\begin{cases} w_1^{\dagger\dagger} = \phi\eta_1 \\ w_2^{\dagger\dagger} = (1-\phi)\,\eta_1 + \eta_2 \end{cases}$$

if the deadline binds.

Next, to determine the optimal deadline ϕ given beliefs $\hat{E}^{\dagger\dagger}$, we find the ϕ that solves

$$\begin{aligned} \max_{\phi} V_0 &\equiv \hat{E}^{\dagger\dagger}\left[u\left(w_1^{\dagger\dagger}\right)\right] + \hat{E}^{\dagger\dagger}\left[u\left(w_2^{\dagger\dagger}\right)\right] \\ \text{s.t. } 0 &\leq \phi \leq 1. \end{aligned} \tag{B.1}$$

First we consider the possibility that $\left(\hat{\mu}_1^{\dagger\dagger}\right)^2 + \left(\hat{\sigma}_1^{\dagger\dagger}\right)^2 > 0$. From the expression for $w_1^{\dagger\dagger}$ above we see that a deadline $0 \leq \phi < \frac{1}{2}$ binds for no values of η_1, while a deadline $\frac{1}{2} \leq \phi \leq 1$ binds for all values of η_1. The agent is clearly indifferent between all never-binding deadlines, so in order to find the optimal deadline we simply need to find the optimal always-binding deadline and compare it with a never-binding deadline, say $\phi = 0$.

To find the optimal always-binding deadline we write the objective from problem (B.1) as

$$\hat{E}^{\dagger\dagger}\left[u\left(\phi\eta_1\right)\right] + \hat{E}^{\dagger\dagger}\left[u\left(\left(1-\phi\right)\eta_1 + \eta_2\right)\right].$$

Ignoring the inequality constraint on ϕ, the first-order condition for this restricted problem (hence its solution is denoted by $\phi^{\dagger\dagger,R}$) with respect to ϕ is

$$\hat{E}^{\dagger\dagger}\left[\eta_1 u'\left(\phi^{\dagger\dagger,R}\eta_1\right)\right] = \hat{E}^{\dagger\dagger}\left[\eta_1 u'\left(\left(1-\phi^{\dagger\dagger,R}\right)\eta_1 + \eta_2\right)\right].$$

Imposing the constraint $\frac{1}{2} \leq \phi \leq 1$, the above becomes

$$\phi^{\dagger\dagger,R} = \min\left\{\max\left\{\frac{1}{2}, \frac{1}{2}\left(1 + \frac{\hat{\mu}_1^{\dagger\dagger}\hat{\mu}_2^{\dagger\dagger} + \hat{\rho}^{\dagger\dagger}\hat{\sigma}_1^{\dagger\dagger}\hat{\sigma}_2^{\dagger\dagger}}{\left(\hat{\mu}_1^{\dagger\dagger}\right)^2 + \left(\hat{\sigma}_1^{\dagger\dagger}\right)^2}\right)\right\}, 1\right\}.$$

If (and only if) beliefs are such that

$$\frac{\hat{\mu}_1^{\dagger\dagger}\hat{\mu}_2^{\dagger\dagger} + \hat{\rho}^{\dagger\dagger}\hat{\sigma}_1^{\dagger\dagger}\hat{\sigma}_2^{\dagger\dagger}}{\left(\hat{\mu}_1^{\dagger\dagger}\right)^2 + \left(\hat{\sigma}_1^{\dagger\dagger}\right)^2} = \frac{3-\delta}{3+\delta},$$

then we have

$$\phi^{\dagger\dagger,R} = \frac{3}{3+\delta},$$

and consequently the desired outcome that

$$w_1^{\dagger\dagger,R} = \frac{3}{3+\delta}\eta_1 = w_1^{\diamondsuit}.$$

To find $\phi^{\dagger\dagger}$ for these beliefs, we still need to compare the value of V_0 for $\phi^{\dagger\dagger,R}$ and for $\phi = 0$, i.e., we compare

$$V_0\left(\phi^{\dagger\dagger,R}\right) = \hat{E}^{\dagger\dagger}\left[u\left(\phi^{\dagger\dagger,R}\eta_1\right) + u\left(\left(1-\phi^{\dagger\dagger,R}\right)\eta_1 + \eta_2\right)\right]$$

with

$$V_0\left(0\right) = \hat{E}^{\dagger\dagger}\left[u\left(\frac{1}{2}\eta_1\right) + u\left(\frac{1}{2}\eta_1 + \eta_2\right)\right].$$

Substituting for $\phi^{\dagger\dagger,R}$ and observing that the former is always greater than the latter, we conclude that $\phi^{\dagger\dagger} = \phi^{\dagger\dagger,R} = \frac{3}{3+\delta}$, and so $w_1^{\dagger\dagger} = w_1^{\dagger\dagger,R} = w_1^{\diamondsuit}$.

Thus, we have shown that beliefs $\hat{E}^{\dagger\dagger}$ consisting of $\hat{\mu}_{2|1}^{\dagger\dagger} = 0$ and $\hat{\sigma}_{2|1}^{\dagger\dagger} = 0$, and $\hat{\mu}_t^{\dagger\dagger}$ and $\hat{\sigma}_t^{\dagger\dagger}$ for $t = 1, 2$ such that $\frac{\hat{\mu}_1^{\dagger\dagger}\hat{\mu}_2^{\dagger\dagger} + \hat{\rho}^{\dagger\dagger}\hat{\sigma}_1^{\dagger\dagger}\hat{\sigma}_2^{\dagger\dagger}}{\left(\hat{\mu}_1^{\dagger\dagger}\right)^2 + \left(\hat{\sigma}_1^{\dagger\dagger}\right)^2} = \frac{3-\delta}{3+\delta}$, together with the deadline $\phi^{\dagger\dagger} = \frac{3}{3+\delta}$ and the optimal action choice $w_1^{\dagger\dagger} = \frac{3}{3+\delta}\eta_1$ form a solution of the original problem.

Next we consider the possibility that $\left(\hat{\mu}_1^{\dagger\dagger}\right)^2 + \left(\hat{\sigma}_1^{\dagger\dagger}\right)^2 = 0$. Then under $\hat{E}^{\dagger\dagger}$ the agent believes with certainty that $\eta_1 = 0$ and so that he will choose $w_1 = 0$ and $w_2 = \eta_2$, and therefore he is indifferent between all deadline choices, including the deadline $\frac{3}{3+\delta}$ which would result in the agent being forced to work amount $w_1 = \frac{3}{3+\delta}\eta_1 = w_1^{\diamondsuit}$. Thus, beliefs $\hat{E}^{\dagger\dagger}$ consisting of $\hat{\mu}_t^{\dagger\dagger} = 0$ and $\hat{\sigma}_t^{\dagger\dagger} = 0$ for $t = 1, 2$ and $t =$"2|1", together with the deadline $\phi^{\dagger\dagger} = \frac{3}{3+\delta}$ and the optimal action choice $w_1^{\dagger\dagger} = \frac{3}{3+\delta}\eta_1$ form another solution of the original problem.

Finally, we show that other than the two solutions we have found above, there are no other solutions to the original problem. Since the modified problem we solved above is a less constrained version of the original problem, and since the solutions to both problems are the same and hence yield the same well-being, it must be that all solutions of the original problem are also solutions of the modified problem. But the modified problem is concave in w_1, $\hat{\mu}_{2|1}$, and $\hat{\sigma}_{2|1}$, and therefore all solutions of both problems must have $w_1 = \frac{3}{3+\delta}\eta_1$, $\hat{\mu}_{2|1} = 0$ and $\hat{\sigma}_{2|1} = 0$. Therefore, we can conclude from (**??**) and (**??**) that all solutions to the original problem must have $\phi = \frac{3}{3+\delta}$. And we have seen above that for $\hat{\mu}_{2|1} = \hat{\sigma}_{2|1} = 0$ this is only possible when $\frac{\hat{\mu}_1\hat{\mu}_2 + \hat{\rho}\hat{\sigma}_1\hat{\sigma}_2}{\hat{\mu}_1^2 + \hat{\sigma}_1^2} = \frac{3-\delta}{3+\delta}$ or $\hat{\mu}_1 = \hat{\sigma}_1 = 0$. As a result, we conclude that there are no other solutions to the original problem.

(iii) In the variation where an objective observer imposes a deadline, we need to find the optimal deadline given objective beliefs and the optimal beliefs given this deadline. In particular, we find deadline $\phi^{\triangle\triangle}$ and beliefs $\hat{E}_1^{\triangle\triangle}\left[\eta_2\right]$ and $\hat{V}ar_1^{\triangle\triangle}\left[\eta_2\right]$ that solve, respectively,

$$\begin{aligned} \max_{\phi} E\left[V_1\right] &\equiv E\left[u\left(w_1^{\triangle\triangle}\right) + u\left(w_2^{\triangle\triangle}\right)\right] \\ \text{s.t. } 0 &\leq \phi \leq 1, \end{aligned}$$

where $w_1^{\triangle\triangle}$ and $w_2^{\triangle\triangle}$ are optimal actions given by Proposition 1 evaluated using beliefs $\hat{E}_1^{\triangle\triangle}[\eta_2]$, and

$$\max_{\hat{E}_1[\eta_2], \hat{V}ar_1[\eta_2]} \mathcal{W} \equiv \frac{1}{2} E\left[V_1 + V_2\right],$$

subject to optimal action and deadline choice given beliefs and subject to the primitive constraints on beliefs.

In finding the optimal deadline and beliefs, it is useful to first observe that

$$E\left[u\left(\eta_1\right) + u\left(\eta_2\right)\right] - E\left[u\left(w_1\right) + u\left(w_2\right)\right] \geq 0,$$

for any w_1 and w_2, since using the resource constraint (1) the above becomes

$$\begin{aligned}
& E\left[u\left(\eta_1\right) + u\left(\eta_2\right)\right] - E\left[u\left(w_1\right) + u\left(\eta_1 + \eta_2 - w_1\right)\right] \\
= & -\frac{1}{2} E\left[\eta_1^2 + \eta_2^2 - \left\{\left(w_1\right)^2 + \left(\eta_1^2 + \eta_2^2 + \left(w_1\right)^2 + 2\eta_1\eta_2 - 2\eta_1 w_1 - 2\eta_2 w_1\right)\right\}\right] \\
= & \; E\left[w_1^2 + \eta_1\eta_2 - \eta_1 w_1 - \eta_2 w_1\right] \\
= & \; E\left[\left(\eta_1 - w_1\right)\left(\eta_2 - w_1\right)\right] \\
= & \; E\left[E\left[\left(\eta_1 - w_1\right)\left(\eta_2 - w_1\right)|\eta_1\right]\right] \\
= & \; E\left[\left(\eta_1 - w_1\right) E\left[\left(\eta_2 - w_1\right)|\eta_1\right]\right] \\
= & \; E\left[\left(\eta_1 - w_1\right)^2\right] \\
\geq & \; 0.
\end{aligned}$$

From this we conclude that either $\hat{E}_1^{\triangle\triangle}[\eta_2] < \eta_1$ and so $\phi^{\triangle\triangle} = 1$, or $\hat{E}_1^{\triangle\triangle}[\eta_2] \geq \eta_1$ and so the objective observer is indifferent between all deadlines. We show that the former is true. Indeed, we show that setting $\hat{E}_1[\eta_2] = \frac{1-\delta}{3+\delta}\eta_1$ (which happens to be the optimal belief absent a deadline) yields higher well-being than setting $\hat{E}_1[\eta_2] = \eta_1$, regardless of the value of $\phi^{\triangle\triangle}$, and since we know that setting $\hat{E}_1[\eta_2] > \eta_1$ yields the same well-being as setting $\hat{E}_1[\eta_2] = \eta_1$, we conclude that $\hat{E}_1^{\triangle\triangle}[\eta_2] < \eta_1$ and so it must be that $\phi^{\triangle\triangle} = 1$. Since beliefs maximize well-being

$$\begin{aligned}
\mathcal{W} &= \tfrac{1}{2} E\left[(1+\delta)\, u\left(w_1\right) + \hat{E}_1\left[u\left(w_2\right)\right] + u\left(w_2\right)\right] \\
&= \tfrac{1}{2} E\left[(1+\delta)\, u\left(\frac{1}{2}\left(\eta_1 + \hat{E}_1[\eta_2]\right)\right) + \hat{E}_1\left[u\left(\frac{1}{2}\eta_1 + \eta_2 - \frac{1}{2}\hat{E}_1[\eta_2]\right)\right] + u\left(\frac{1}{2}\eta_1 + \eta_2 - \frac{1}{2}\hat{E}_1[\eta_2]\right)\right] \\
&= \tfrac{1}{4} E\left[-\left\{\frac{1}{4}(3+\delta)\eta_1^2 + \eta_2^2 + \eta_1\eta_2 + \hat{E}_1\left[\eta_2^2\right] - \frac{1}{4}(1-\delta)\left(\hat{E}_1[\eta_2]\right)^2 + \frac{1}{2}(1+\delta)\eta_1\hat{E}_1[\eta_2] - \eta_2\hat{E}_1[\eta_2]\right\}\right] \\
&= \tfrac{1}{4} E\left[-\left\{\frac{1}{4}(3+\delta)\eta_1^2 + E\left[\eta_2^2|\eta_1\right] + \eta_1 E\left[\eta_2|\eta_1\right] + \hat{E}_1\left[\eta_2^2\right] - \frac{1}{4}(1-\delta)\left(\hat{E}_1[\eta_2]\right)^2 + \frac{1}{2}(1+\delta)\eta_1\hat{E}_1[\eta_2] - E\right.\right. \\
&= \tfrac{1}{4} E\left[-\left\{\frac{1}{4}(3+\delta)\eta_1^2 + E\left[\eta_2^2|\eta_1\right] + \eta_1^2 + \hat{E}_1\left[\eta_2^2\right] - \frac{1}{4}(1-\delta)\left(\hat{E}_1[\eta_2]\right)^2 + \frac{1}{2}(1+\delta)\eta_1\hat{E}_1[\eta_2] - \eta_1\hat{E}_1[\eta_2]\right\}\right.
\end{aligned}$$

we compare the expression inside the expectation first evaluated at $\hat{E}_1[\eta_2] = \frac{1-\delta}{3+\delta}\eta_1$ and then evaluated at $\hat{E}_1[\eta_2] = \eta_1$, both using the the same value for $\hat{V}ar_1[\eta_2]$. Their difference is

$$\begin{aligned}
& -\left\{\frac{1}{4}(3+\delta)\eta_1^2 + E\left[\eta_2^2|\eta_1\right] + \eta_1^2 + \left(\frac{1-\delta}{3+\delta}\eta_1\right)^2 + \hat{V}ar_1[\eta_2] - \frac{1}{4}(1-\delta)\left(\frac{1-\delta}{3+\delta}\eta_1\right)^2 + \frac{1}{2}(1+\delta)\frac{1-\delta}{3+\delta}\eta_1^2 - \right. \\
& \left\{\frac{1}{4}(3+\delta)\eta_1^2 + E\left[\eta_2^2|\eta_1\right] + \eta_1^2 + \eta_1^2 + \hat{V}ar_1[\eta_2] - \frac{1}{4}(1-\delta)\eta_1^2 + \frac{1}{2}(1+\delta)\eta_1^2 - \eta_1^2\right\} \\
= \; & \frac{1}{4}\left\{1 - \left(\frac{1-\delta}{3+\delta}\right)^2 - \frac{1}{4}(1-\delta) + \frac{1}{4}(1-\delta)\left(\frac{1-\delta}{3+\delta}\right)^2 + \frac{1}{2}(1+\delta) - \frac{1}{2}(1+\delta)\frac{1-\delta}{3+\delta} - 1 + \frac{1-\delta}{3+\delta}\right\}\eta_1^2 \\
= \; & \frac{1}{16}\left\{\frac{(1-\delta)^2}{3+\delta} + (1+3\delta)\right\}\eta_1^2 \\
\geq \; & 0,
\end{aligned}$$

and so it must be that $\hat{E}_1[\eta_2] < \eta_1$, and hence, $\phi^{\triangle\triangle} = 1$ is the optimal deadline.

To find the optimal beliefs, we substitute for $w_1^{\triangle\triangle} = \eta_1$ in the well-being, which becomes

$$\begin{aligned}
\mathcal{W} &= -\tfrac{1}{4}E\left[(1+\delta)\eta_1^2 + \hat{E}_1\left[\eta_2^2\right] + \eta_2^2\right] \\
&= -\tfrac{1}{4}E\left[(1+\delta)\eta_1^2 + \left(\hat{E}_1[\eta_2]\right)^2 + \hat{V}ar_1[\eta_2] + E\left[\eta_2^2|\eta_1\right]\right],
\end{aligned}$$

and which is clearly decreasing in both $\hat{E}_1[\eta_2]$ and $\hat{V}ar_1[\eta_2]$, and so imposing the constraints $\hat{E}_1[\eta_2] \geq 0$ and $\hat{V}ar_1[\eta_2] \geq 0$, we conclude that $\hat{E}_1^{\triangle\triangle}[\eta_2] = 0$ and $\hat{V}ar_1^{\triangle\triangle}[\eta_2] = 0$.

The choice of the other beliefs is arbitrary, since they do not affect well-being.

B.7 Proof of Proposition 7

i) Substituting for V_1, V_2, and w_1^*, the well-being is

$$\begin{aligned}
\mathcal{W} &= \tfrac{1}{2}E\left[(1+\delta)u(w_1) + \hat{E}_1[u(w_2) + (P - cw_2)] + u(w_2) + (P - cw_2)\right] \\
&= -\tfrac{1}{4}E\left[\begin{array}{c}
(1+\delta)\tfrac{1}{4}\left(\eta_1^2 + \left(\hat{E}_1[\eta_2]\right)^2 + 2\eta_1\hat{E}_1[\eta_2] + c^2 + 2\eta_1 c + 2c\hat{E}_1[\eta_2]\right) + \\
\tfrac{1}{4}\hat{E}_1\left[\eta_1^2 + 4\eta_2^2 + 4\eta_1\eta_2 + \left(\hat{E}_1[\eta_2]\right)^2 + c^2 + 2c\hat{E}_1[\eta_2] - 2\eta_1\hat{E}_1[\eta_2] - 4\eta_2\hat{E}_1[\eta_2] - 2c\eta_1 - 4c\eta_2\right] + \\
\tfrac{1}{4}\left\{\eta_1^2 + 4\eta_2^2 + 4\eta_1\eta_2 + \left(\hat{E}_1[\eta_2]\right)^2 + c^2 + 2c\hat{E}_1[\eta_2] - 2\eta_1\hat{E}_1[\eta_2] - 4\eta_2\hat{E}_1[\eta_2] - 2c\eta_1 - 4c\eta_2\right\} + \\
c\eta_1 + c\eta_2 - 2P - c^2
\end{array}\right] \\
&= -\tfrac{1}{4}E\left[\begin{array}{c}
-2P - \tfrac{1}{4}(1-\delta)c^2 + \tfrac{1}{4}(3+\delta)\eta_1^2 + \eta_2^2 + \eta_1\eta_2 + \tfrac{1}{2}(1+\delta)c\eta_1 + \\
\hat{E}_1\left[\eta_2^2\right] - \tfrac{1}{4}(1-\delta)\left(\hat{E}_1[\eta_2]\right)^2 + \tfrac{1}{2}(1+\delta)\eta_1\hat{E}_1[\eta_2] + \tfrac{1}{2}(1+\delta)c\hat{E}_1[\eta_2] - \eta_2\hat{E}_1[\eta_2]
\end{array}\right].
\end{aligned}$$

For any η_1 we have

$$\frac{d\mathcal{W}}{d\hat{E}_1[\eta_2]} = -\tfrac{1}{4}\left[\frac{1}{2}(3+\delta)\hat{E}_1[\eta_2] + \frac{1}{2}(1+\delta)\eta_1 + \frac{1}{2}(1+\delta)c - E_1[\eta_2]\right],$$

and so the optimal beliefs are

$$\hat{E}_1^{**}[\eta_2] = \frac{1-\delta}{3+\delta}\eta_1 - \frac{1+\delta}{3+\delta}c.$$

Thus, for $\hat{E}_1^{**}[\eta_2] > 0$ we have

$$\frac{dE_1^{**}[\eta_2]}{dc} = -\tfrac{1+\delta}{3+\delta} < 0.$$

In addition, we observe that $\hat{E}_1^{**}[\eta_2] = 0$ for $\eta_1 \leq \frac{1+\delta}{1-\delta}c$, which is clearly increasing in c, so the greater c is, the more the values of η_1 for which $\hat{E}_1^{**}[\eta_2] = 0$.

ii) Substituting $\hat{E}_1[\eta_2] = E_1^{**}[\eta_2]$ in w_1^* we have

$$w_1^{**} = \frac{1}{2}\max\left\{\eta_1 + c, \frac{2}{3+\delta}(2\eta_1 + c)\right\},$$

so clearly we have $\frac{dw_1^{**}}{dc} > 0$. In particular, for $\eta_1 \leq \frac{1+\delta}{1-\delta}c$ we have $\frac{dw_1^{**}}{dc} = \frac{1}{2}$, while otherwise we have $\frac{dw_1^{**}}{dc} = \frac{1}{3+\delta}$.

Substituting $\hat{E}_1[\eta_2] = E_1[\eta_2]$ in w_1^* we have $w_1^{RE} = \frac{1}{2}(2\eta_1 + c)$, and so $\frac{dw_1^{RE}}{dc} = \frac{1}{2}$. Thus, $\frac{dw_1^{RE}}{dc} \geq \frac{dw_1^{**}}{dc} > 0$.

B.8 Proof of Proposition 8

Substituting for V_1, V_2, and w_1^*, the well-being is

$$\begin{aligned}\mathcal{W} &= \tfrac{1}{2}E\left[(1+\delta)u(w_1) + \hat{E}_1\left[u(w_2) + \left(P - k\left(\hat{E}_1[\eta_2] - \eta_2\right)^2\right)\right] + u(w_2) + \left(P - k\left(\hat{E}_1[\eta_2] - \eta_2\right)^2\right)\right] \\ &= -\tfrac{1}{4}E\left[\begin{array}{c}\tfrac{1}{4}(3+\delta)\eta_1^2 + \eta_2^2 + \eta_1\eta_2 + \hat{E}_1\left[\eta_2^2\right] - \tfrac{1}{4}(1-\delta)\left(\hat{E}_1[\eta_2]\right)^2 + \tfrac{1}{2}(1+\delta)\eta_1\hat{E}_1[\eta_2] - \eta_2\hat{E}_1[\eta_2] \\ -2P + k\eta_2^2 + k\hat{E}_1\left[\eta_2^2\right] - 2k\eta_2\hat{E}_1[\eta_2]\end{array}\right]\end{aligned}$$

For any η_1 we have

$$\frac{d\mathcal{W}}{d\hat{E}_1[\eta_2]} = -\tfrac{1}{4}\left[\frac{1}{2}(3+\delta+4k)\hat{E}_1[\eta_2] + \frac{1}{2}(1+\delta)\eta_1 - (1+2k)E_1[\eta_2]\right]$$

and so the optimal beliefs are

$$\hat{E}_1^{**}[\eta_2] = \frac{1-\delta+4k}{3+\delta+4k}\eta_1.$$

Thus, for $\hat{E}_1^{**}[\eta_2] > 0$ we have

$$\frac{dE_1^{**}[\eta_2]}{dk} = \frac{8(1+\delta)}{(3+\delta+4k)^2}\eta_1 > 0.$$

In addition, we observe that $\hat{E}_1^{**}[\eta_2] = 0$ for $k \leq \frac{1-\delta}{4}$, so the smaller k is, the more likely it is that $\hat{E}_1^{**}[\eta_2] = 0$.

ii) Substituting $\hat{E}_1[\eta_2] = E_1^{**}[\eta_2]$ in w_1^* we have

$$w_1^{**} = \frac{1}{2}\max\left\{\eta_1, \frac{4(1+2k)}{3+\delta+4k}\eta_1\right\},$$

so clearly we have $\frac{dw_1^{**}}{dk} \geq 0$.

B.9 Proof of Proposition 9

Follows immediately from differentiating $\hat{E}_1^{**}[\eta_2]$ and w_1^{**} in Proposition 3.

B.10 Proof of Proposition 10

{{{To be done.}}}

Differentiating Proposition 3 (i) and using $\frac{dw_1^{**}}{da} = \frac{dw_1^{**}}{d\hat{\mu}^{**}}\frac{d\hat{\mu}^{**}}{da}$ and Proposition 1, we have

$$\begin{aligned} \frac{d\hat{\mu}^{**}}{da} &= -2\frac{1+\delta}{3+\delta} < 0 \\ &\text{and} \\ \frac{dw^{**}}{da} &= -\frac{1+\delta}{3+\delta} < 0. \end{aligned}$$

for $\mu > \frac{1+\delta}{2}(1+2a)$. When beliefs are constrained, $\frac{d\hat{\mu}^{**}}{da} = \frac{dw^{**}}{da} = 0$.

B.11 The planning fallacy and overconfidence with discounting

{{{To be done.}}}

Proposition 12 *(Extension of results for discounting)*
(i) The agent with optimal beliefs exhibits the planning fallacy, $\hat{\mu}^{**} < \mu$ and $w_1^{**} < w^{RE}$, and is overconfident, $\hat{\sigma}^{**} = 0 < \sigma$.
(ii) The planning fallacy becomes worse the more the agent cares about the past: $\frac{d\hat{\mu}^{**}}{d\delta} \leq 0$ and $\frac{dw_1^{**}}{d\delta} \leq 0$, with strict inequalities for $\hat{\mu}^{**} > 0$.
(iii) The planning fallacy becomes worse the lower the curvature of the utility function: a) $\frac{d\hat{\mu}^{**}}{da} \leq 0$, with strict inequality for $\hat{\mu}^{**} > 0$.; and b) $\frac{dw_1^{**}}{da} < 0$
(iv) $\hat{\mu}^{**} > 0$ when $\mu > \frac{1+\delta\beta^2}{(1+\beta)\beta}(1+2a)$.

Proof:

(i) First we solve for the optimal action given arbitrary beliefs $\hat{\mu}$. Substituting the budget constraint, the agent's objective is

$$\max_{w_1} \hat{E}\left[u(w_1) + \beta u(1 - w_1 + \eta)\right].$$

with first-order condition

$$u'\left(w_1^*\right) = \beta \hat{E}\left[u'\left(1 - w_1^* + \eta\right)\right]$$

so that by the certainty equivalence property of quadratic utility

$$w_1^* = \beta \hat{E}\left[1 - w_1^* + \eta\right]$$

which reorganizes to

$$w_1^* = \frac{\beta}{1+\beta}\left(1 - \frac{1-\beta}{\beta}a + \hat{\mu}\right).$$

Imposing the budget constraint, the problem of finding optimal beliefs, given that the agent chooses the optimal action for those beliefs, is

$$\max_{\hat{\mu},\hat{\sigma}} \frac{1}{2} E\left\{u\left(w_1^*\right) + \beta \hat{E}\left[u\left(1 - w_1^* + \eta\right)\right] + \beta\left[\delta u\left(w_1^*\right) + u\left(1 - w_1^* + \eta\right)\right]\right\}.$$

This objective can be simplified to

$$\left(-a w_1^* - \frac{1}{2} w_1^*\right) + \beta\left\{-a\left(1 - w_1^* + \hat{\mu}\right) - \frac{1}{2}\left[\hat{\sigma}^2 + \left(1 - w_1^* + \hat{\mu}\right)^2\right]\right\} +$$
$$+\beta\delta\left(-a w_1^* - \frac{1}{2} w_1^*\right) + \beta\left\{-a\left(1 - w_1^* + \mu\right) - \frac{1}{2}\left[\sigma^2 + \left(1 - w_1^* + \mu\right)^2\right]\right\}.$$

This is clearly maximized by optimally setting the subjective variance equal to 0, i.e. $\hat{\sigma}^{**} = 0$. The first order condition with respect to $\hat{\mu}$ yields (after some messy algebra)

$$\hat{\mu}^{**} = \frac{(1+\beta)\beta^2\mu - \left(1+\delta\beta^2\right)\beta - 2\beta\left(1+\delta\beta^2\right)a}{\beta\left(1+\delta\beta^2\right) + \beta^2(1+\beta)}.$$

One can see that $\hat{\mu}^{**} \leq \mu < \overline{\eta}$ as long as $\mu > -1 - 2a$, which is trivially satisfied since we have assumed $\mu \geq 0$. For small μ, (in particular for $\mu \leq \frac{1+\delta\beta^2}{(1+\beta)\beta}(1+2a)$) $\hat{\mu}^{**}$ is constrained by $\hat{\mu} \geq 0$, so we have

$$\hat{\mu}^{**} = \max\left\{0, \frac{(1+\beta)\beta^2\mu - \left(1+\delta\beta^2\right)\beta - 2\beta\left(1+\delta\beta^2\right)a}{\beta\left(1+\delta\beta^2\right) + \beta^2(1+\beta)}\right\}. \quad \text{(B.2)}$$

Substituting this result into the optimal work choice we found earlier, we have

$$w_1^{**} = \max\left\{\frac{\beta}{1+\beta}\left(1 - \frac{1-\beta}{\beta}a\right), \frac{\beta^3\left(1 - \frac{1+\beta+\delta\beta^2-\beta^2}{\beta^2}a + \mu\right)}{\beta\left(1+\delta\beta^2\right) + \beta^2(1+\beta)}\right\}$$

To see that $w_1^{**} < w_1^{RE}$, observe that $1 + \delta\beta^2 > 0$ implies that $\frac{\beta^3}{\beta\left(1+\delta\beta^2\right)+\beta^2(1+\beta)} < \frac{\beta}{1+\beta}$ as well as that $\frac{1+\beta+\delta\beta^2-\beta^2}{\beta^2} > \frac{1-\beta}{\beta}$.

(ii) For $\hat{\mu}^{**} = 0$ we have $\frac{d\hat{\mu}^{**}}{d\delta} = 0$ and $\frac{dw_1^{**}}{d\delta} = 0$, while otherwise we have

$$\frac{d\hat{\mu}^{**}}{d\delta} \propto -\beta^5(1+\beta)(\mu + 1 + 2a)$$

and

$$\frac{dw_1^{**}}{d\delta} = \frac{\beta}{1+\beta}\frac{d\hat{\mu}^{**}}{d\delta}$$

so clearly we have $\frac{d\hat{\mu}^{**}}{d\delta} < 0$ and $\frac{dw_1^{**}}{d\delta} < 0$.

(iii) For $\hat{\mu}^{**} = 0$ we have $\frac{d\hat{\mu}^{**}}{da} = 0$ and $\frac{dw_1^{**}}{da} = -\frac{1-\beta}{1+\beta}$, while otherwise we have

$$\frac{d\hat{\mu}^{**}}{da} = -\frac{2\left(1+\delta\beta^2\right)}{1+\delta\beta^2+\beta\left(1+\beta\right)}$$

and

$$\frac{dw_1^{**}}{da} \propto -\left(1+\beta+\delta\beta^2-\beta^2\right).$$

Obviously, $\frac{d\hat{\mu}^{**}}{da} \leq 0$ with strict inequality for $\hat{\mu}^{**} > 0$. For $\hat{\mu}^{**} = 0$, $\frac{dw_1^{**}}{da} < 0$ is equivalent to $\beta < 1$. This is clearly also a sufficient condition for $\frac{dw_1^{**}}{da} < 0$ when $\hat{\mu}^{**} > 0$, since in that case we have $\beta > \beta^2$.

(iv) From the expression for $\hat{\mu}^{**}$ we can directly see that $\hat{\mu}^{**} > 0$ if $\mu > \frac{1+\delta\beta^2}{(1+\beta)\beta}(1+2a)$.

B.12 Proof of Proposition 11

{{{To be done.}}}

Differentiating equation (B.2) from the proof of Proposition 12 with respect to β, for $\hat{\mu}^{**} > 0$, the sign of interest is proportional to

$$\frac{d\hat{\mu}^{**}}{d\beta} \propto \beta^2\left(1-\delta\beta^2+2\beta\right)\left(\mu+1+2a\right),$$

hence, showing that $\frac{d\hat{\mu}^{**}}{d\beta} > 0$ is equivalent to showing that

$$1-\delta\beta^2+2\beta > 0.$$

This inequality is satisfied for $\beta \in \left[\frac{1-\sqrt{1+\delta}}{\delta}, \frac{1+\sqrt{1+\delta}}{\delta}\right] \supset [0,1]$.

Differentiating the expression for w_1^{**} from Proposition 12 with respect to β, we have that if $\hat{\mu}^{**} = 0$ then

$$\frac{dw_1^{**}}{d\beta} \propto \beta\left(1+2a\right),$$

and if $\hat{\mu}^{**} > 0$ then

$$\frac{dw_1^{**}}{d\beta} \propto \beta^3\left(2+\beta\right)\left(1+\mu+2a\right).$$

Clearly, both expressions are positive for all parameter values.