Projection Bias in Effort Choices*

Marc Kaufmann (Central European University)

June 14, 2022

Abstract

I model individuals working on a long-term task who, due to projection bias, underestimate changes in marginal disutility. At the start of each day when current marginal disutility is low, such individuals overestimate how much they will work, and as they grow tired from working, they plan to work less and less. Despite these fluctuating plans, when they face decreasing returns to effort they work optimally, yet if they commit in advance, they overcommit. When facing increasing returns, they may repeatedly start and quit, working when rested only to give up as they grow tired. In all-or-nothing tasks, start-and-quitting can lead to wasting effort on never-to-be-completed tasks, or to completing them by working inefficiently little early on.

Keywords— projection bias, increasing returns, multi-tasking, time inconsistency

JEL — D03, D90, J22

1 Introduction

Our tastes fluctuate, often rapidly: we grow tired and thirsty from running, and we savor food or crave coffee more the longer we go without. Furthermore, evi-

*Central European University, Department of Economics and Business, Quellenstrasse 51-55, Vienna, Austria. Email: kaufmannm@ceu.edu. Declaration of interest: none.

dence from a variety of domains suggests that we misperceive future tastes as being closer to our current tastes than they will be (Loewenstein, O'Donoghue, and Rabin (2003)).[1] This *projection bias* can trigger undesirable and unintended habits and behaviors, such as buying too much when shopping on an empty stomach, or becoming addicted due to under-appreciating the future intensity of cravings. In this paper, I theoretically study effort choices where the distaste for work changes over time, such as when students grow bored of studying or employees become tired of working. Due to projection bias, individuals mispredict future disutility of work, which causes them to overcommit, to waste time on a task they do not complete, and to spend too much time on short-term at the expense of long-term tasks.

In my model, formalized in Section 2, an agent faces a single long-term task that she can work on by a fixed deadline. The agent starts each day rested, works continuously on the task, and stops working the moment she *perceives* stopping to be optimal given the monetary benefits and the perceived disutility from work. The instantaneous disutility from continuing to work is equal to $D'(s)$, which increases with the total time s she has worked so far that day. She misperceives her future disutility due to projection bias: she predicts that her marginal disutility after e hours of work lies between her current marginal disutility, $D'(s)$, and her true future marginal disutility, $D'(e)$. So when she is rested and marginal disutility is thus particularly low, she underestimates how hard it will be to work on future days, while when tired and marginal disutility is thus particularly high, she overestimates it. The main goal of the paper is to study the implications of the changing and inconsistent plans for future work that result from such misperceptions.

In Section 3, I analyse concave benefits. While the person stops working at the right time when deciding on the go, she will overcommit if making ex ante choices. Since she grows more tired the longer she works each day, she underestimates how unpleasant work will be at the time she stops working. Therefore she overestimates

[1]Evidence for projection bias has been found for food (Read and Van Leeuwen (1998); Nordgren, Pligt, and Harreveld (2008)), drink (Van Boven and Loewenstein (2003)), sexual arousal (Loewenstein, Nagin, and Paternoster (1997); Ariely and Loewenstein (2006)), effortful tasks (Augenblick and Rabin (2019)), heroin substitute cravings (Badger et al. (2007)), and the endowment effect (Loewenstein and Adler (1995)). Projection bias resembles immune neglect (Gilbert et al. (1998)) whereby people overestimate how long they will feel bad about negative events. Bushong and Gagnon-Bartsch (2020) additionally find evidence for *interpersonal* projection bias for choices over effort.

at all times how much she will, and should, work – right until the time at which she should stop. At this point, her current marginal disutility equals the marginal benefits, and since she perceives her *current* marginal disutility correctly, she stops working. At all earlier times, she held overly ambitious plans about how much she should work. If she had to commit to work, she would therefore overcommit, failing to realize how painful it would become. This suggests that people accept too many soft commitments that increase their workload, such as taking on additional work or promising higher levels of quality.

In Section 4, I consider *convex* benefits. Consider, for instance, an all-or-nothing task that pays a fixed benefit only if the person completes a target amount of work. At the start of the day, the person may plan to work, believing she will complete the task. Yet, as she grows tired from working, work becomes increasingly less appealing to the point where she may decide to drop the task and quit working. While she plans on not resuming work again if she quits, the next day she is rested and may once again start working, and the cycle repeats. It stops eventually: if the person makes too little progress on the task, even when rested she no longer starts working; if she makes enough progress, even when tired she no longer quits. In both cases, the behavior is optimal *given* her current progress on the task, but she may have wasted effort on a task she never actually completes, or worked inefficiently on a task she does complete.

If we only observe whether a person completed a task or not, does this reveal whether the task was worth doing? It turns out that task completion reveals little in all generality. Given the discussion above, if the person never works on the task, then it was not worth doing; if she works on it efficiently, then it was worth doing. However, depending on her disutility, she may complete some tasks that are not even worth doing efficiently (when $D''' < 0$), and she may fail to complete tasks that are worth doing efficiently (when $D''' > 0$). Only when the disutility is quadratic does the person fail all tasks that are not worthwhile and complete all tasks that are worthwhile – although the way in which she completes them may mean that she would have been better off not working at all.

In Section 5, I extend the basic setup to allow for multi-tasking when facing both a single long-term task (where benefits depend on total work by the final day) and a daily short-term task (where benefits depend on daily work). The person works too much overall, and under the plausible assumption that short-term tasks have small additional benefits from being completed earlier in the day, the person works too little, but increasingly more, on the long-term task. Such benefits occur, for instance, if early completion signals being valued to her client, a good work ethic to her boss, or improves coordination with colleagues. But for a projection-biased person, early completion has a cost: since the decision to stop working on the short-term task is a partial commitment – already completed work cannot be undone – the person overcommits by working too much on the short-term task. While working on the short-term task, she is more rested and overestimates how much she will work on the long-term task. Once she works on the long-term task, she becomes more tired, so she plans to allocate more time to the long-term and less to the short-term task on future days – a plan that she will not implement once rested. Over time, she realizes that she is falling behind on the long-term task, so she starts to work more overall every day, allocating more (yet still too little) time on the long-term task and less (yet still too much) time on short-term tasks.

Finally in Section 6, I compare the behavioral and welfare implications of this model to those of present bias, self-control, and overconfidence that are relevant in task management. I highlight some observable behaviors that are hard to reconcile with these models. Even when a behavior can be explained by other models, the welfare predictions can differ starkly: present bias and self control suggest that sticking to the ex ante plans is optimal, while under projection bias the less ambitious in-the-moment choice may be better. I discuss how data on plans and changes in willingness to work can be used to measure the contribution of different models. Section 7 concludes.

1.1 Literature review

My paper builds on the model of Loewenstein, O'Donoghue, and Rabin (2003) who formalize a model of projection bias and apply it to durable goods consumption, the endowment effect, and habit formation. I focus on the dynamic labor choices with projection bias. Since agents repeatedly re-optimize as they grow tired or rested, the final success of work depends on combining the many decisions made while having inconsistent plans.[2]

From a technical point of view, several related papers study stopping problems under time-inconsistent preferences, such as Quah and Strulovici (2013), Hsiaw (2013), and Huang and Nguyen-Huu (2018). In these models, the time inconsistency stems from present bias or dynamically inconsistent changes in patience, rather than from state changes that depend themselves on earlier decisions. In the multi-period settings that I study, the agent no longer faces a single stopping decision, but constantly re-optimizes. This setup is closest to the model of instant gratification of Harris and Laibson (2013) and the recursive present bias formulation in Ahn, Iijima, and Sarver (2020), both of which assume naiveté and allow for repeated decisions to combine over time to a final outcome such as final savings.

A related question is how partially sophisticated agents may respond to their time inconsistency due to present bias. They may try to commit to specific actions (Laibson (1997)), exert costly self-control over their impulsive selves (Fudenberg and Levine (2006)), or signal to their future selves that they are disciplined so as to incentivize future self-control. Such signalling can happen by taking costly actions today (Bénabou and Tirole (2004)) or by seeking peers who reflect positively on oneself (Battaglini, Bénabou, and Tirole (2005)). While I assume that the agent is unaware of her bias, I discuss in Section 2 that this allows for an agent who is aware of but underestimates her projection bias.

My approach may help integrate projection bias into specific economic settings, in particular to time (mis)management and personnel economics. For instance, Buehler, Griffin, and Ross (1994) find that students believe that they will finish their bachelor's thesis earlier than they actually do – which they explain by students underestimating the number of hours necessary for the task. Projection bias provides a complementary explanation that people overestimate how much they will work. Regarding my results on multi-tasking, Coviello, Ichino, and Persico (2015) and Bray et al. (2016) find empirical evidence that multi-tasking decreases productivity. Coviello, Ichino, and Persico (2014) show that workers may engage knowingly in intrinsically inefficient multi-tasking due to lobbying by co-workers and superiors.

[2] Herrnstein and Prelec (1991)'s model of melioration under distributed choices is somewhat related, but differs substantially in the sense that plans don't matter in their model, while they are central to the results in my paper.

With projection bias, even if multi-tasking is not intrinsically inefficient, such lobbying will lead workers to multi-task inefficiently. Given the possibly large welfare losses under inconsistent behavior, my results highlight the potential to expand the study of projection bias beyond domains with large swings in taste (Levy (2009); Chaloupka, Levy, and White (2019); Chang, Huang, and Wang (2018)) and binding choices (Conlin, O'Donoghue, and Vogelsang (2007); Busse et al. (2015); Buchheim and Kolaska (2017); Michel and Stenzel (2020)) to everyday choices with smaller, repeated changes in taste.

2 A Model of Projection Bias in Effort Choices

2.1 The Formal Model

Environment Consider a setup with two continuous dimensions of time: in each (continuous-time) period $x \in [0,1)$ – informally a *day* – a person works for a single block of time e_x – informally *hours*. Thus e_x is the flow effort in period x from which the person incurs a flow disutility equal to $D(e_x)$, where $D(\cdot)$ is continuously differentiable, with $D(0) = 0$, $D''(0) \geq 0$, $D'(0) \geq 0$, and $D'(e) > 0$, $D''(e) > 0$ for all $e > 0$. The marginal disutility $D'(e)$ is the instantaneous disutility of continuing to work after e hours – at a time when the person has worked for a duration e. Finally, in period $x = 1$, she earns a monetary benefit $B(E_1)$, increasing in $E_1 = \int_0^1 e_x dx$, the total work completed by period 1.

Perceived Disutility Projection bias as defined by Loewenstein, O'Donoghue, and Rabin (2003) leads a person to misperceive her future taste for work as more similar to her current taste for work than it will actually be. As people work in a given period, they grow more tired of working which is captured by increasing marginal disutility, and so they perceive future work as more onerous the more tired they currently are. Formally:

Definition 1 (Projection Bias)**.** *At time s – when a person has worked for a time*

s – the person mispredicts marginal disutility at time e in some period to be

$$\tilde{D}'(e|s) = (1-\alpha)D'(e) + \alpha D'(s) \tag{1}$$

where $\alpha \in [0,1]$ is the degree of projection bias. Moreover, she is naive *with respect to her projection bias: she does not anticipate that she will perceive her marginal disutility differently in the future.*

Taking the integral of this definition, we obtain the perceived disutility from total effort e as $\tilde{D}(e|s) = (1-\alpha)\cdot D(e) + \alpha \cdot D'(s) \cdot e$, since $D(0) = 0$.

Note that, since the taste for money (for consuming goods that money can buy) does not change substantially nor systematically with tiredness, the benefits $B(\cdot)$ are perceived correctly.

Behavior I assume that in each period x, the person stops working at the earliest time when she perceives continuing to work as suboptimal. Concretely, let $E_x = \int_0^x e_y dy$ be the total work she has done by the start of period x. Then it would require her e hours every remaining period to complete a total of $E_x + (1-x)\cdot e$ of work by period 1.[3] At time s, the perceived utility of working e hours in every future period is $B(E_x + (1-x)\cdot e) - \int_x^1 \tilde{D}(e|s)dt = B_x(e) - (1-x)\cdot\tilde{D}(e|s)$, where $B_x(e) := B(E_x + (1-x)\cdot e)$. Then we have the following behavior:

Definition 2 (Momentary Work Decision)**.** *A projection-biased person works in period x until $\tilde{e}_x$ given by*

$$\tilde{e}_x = \inf\{s : B_x(e) - (1-x)\cdot\tilde{D}(e|s) < \max_{\underline{e}\leq s}\left(B_x(\underline{e}) - (1-x)\cdot\tilde{D}(\underline{e}|s)\right), \forall e > s\}$$

In words, the person (roughly) stops at time s when the perceived utility from working strictly more (in this and all future periods) is strictly less than the perceived utility from working as much or less than s.

[3]This uses the fact that in this setup the person finds it optimal to work the same amount each day going forward, as is true for $\alpha < 1$. To extend the definition to the case $\alpha = 1$, I assume that even this fully projection-biased person perceives it to be preferable to work the same across all days. Section 5 generalizes this setup to multi-tasking, and momentary work decisions can be generalized further, but are notationally more cumbersome.

Momentary work decisions thus determine the flow effort in period x, which determines E_x, the amount of effort completed by period x. This yields the following initial value problem (IVP) as an ordinary differential equation (ODE):

Definition 3 (Continuous-Time IVP)**.**

$$\dot{E}_x = \tilde{e}_x$$
$$\tilde{e}_x = \inf\{s : B_x(e) - (1-x) \cdot \tilde{D}(e|s) < \max_{\underline{e} \leq s} \left(B_x(\underline{e}) - (1-x) \cdot \tilde{D}(\underline{e}|s) \right), \forall e > s\}$$

with $E_0 = 0$.

2.2 Discussion of the Model

By assuming that the marginal disutility is the current disutility of working, I implicitly assume that the person either works or does not work, but cannot choose the level of effort intensity each moment. And by assuming that the agent's work in a period is restricted to a single block, I implicitly rule out the possibility of resting within a period. If resting within a period were possible, the person might decide to resume work after a break. Of course, both resting and intensity of work are important, since a projection-biased person may misoptimize both. The current setup allows only exogenous resting between periods such that the person is fully rested at the start of each period.

The central feature that drives the results is the cycle between a high-effort target at the start of each day and a low-effort target later in the day: they drive the overoptimism for concave benefits and the repeated starting and quitting for convex benefits. As long as people grow more tired during each day and more rested between days, the results in Sections 4 and 5 will qualitatively hold. There are certainly situations that do not fit this template. It may be that resting between days does not reduce tiredness; or that daily work and rest maintains tiredness within a narrow and fairly constant range. In the first case, the person grows increasingly more tired, possibly reaching a plateau, while in the second she stays at the same level of tiredness, both of which preclude the repeated cycles that drive the results.

Notice that when $\alpha = 0$ the person has no projection bias, and the actual work done

$\tilde{e}_x$ equals the optimal work e_x^*, if it is unique.[4] Thus the setup nests the unbiased case.

Finally, it may appear as though the model assumes that the person is entirely unaware of her projection bias. Instead, we may wish to allow her to be partially aware, believing that she has projection bias $\hat{\alpha} \in [0, \alpha]$, similar to the model of partially naive present bias in O'Donoghue and Rabin (2001). Specifically, suppose that the person feels her current marginal disutility $D'(s)$ and her *perceived* future marginal disutility $\tilde{D}'(e|s)$. She believes that she is projection-biased with projection bias $\hat{\alpha}$. If this was the case, then she would know that $\tilde{D}'(e|s) = (1-\hat{\alpha})D'(e) + \hat{\alpha}D'(s)$, hence she could infer $\hat{D}'(e)$ via $D'(e) = \frac{\tilde{D}'(e|s) - \hat{\alpha}D'(s)}{1-\hat{\alpha}}$ where she knows all the terms on the right-hand side. However, since her true projection bias is α, we actually have that $\hat{D}'(e) = \frac{(1-\alpha)D'(e)+\alpha D'(s)-\hat{\alpha}D'(s)}{1-\hat{\alpha}} = \frac{(1-\alpha)}{1-\hat{\alpha}}D'(e) - \frac{\alpha-\hat{\alpha}}{1-\hat{\alpha}}D'(s) = (1-\alpha_0)D'(e)+\alpha_0 D'(s)$. Thus partial projection bias with actual bias α and perceived bias $\hat{\alpha}$ is the same as simple projection bias with bias $\alpha_0 = \frac{\alpha-\hat{\alpha}}{1-\hat{\alpha}} < \alpha$, with no projection bias when $\hat{\alpha} = \alpha$.

3 Concave Benefits

Let us start exploring concave benefits with $B''(e) \leq 0$. This allows for a fixed hourly wage rate, for decreasing returns to work, and as such covers cases where the primary question is not whether to work on the task, but how much. We will see that the person then behaves optimally, despite optimistic beliefs, yet if she had to choose a target beforehand, she would overcommit.

Consider Anna, a projection-biased student with $\alpha = 0.5$, who is studying towards an exam at the end of term. The benefits of every additional hour of studying are linear and equal to 3, and studying becomes more unpleasant the longer she studies. Specifically, Anna's daily disutility is quadratic in total time studied, thus $D(e) = \frac{e^2}{2}$ and $D'(e) = e$.[5] After having studied for s hours, Anna plans to study

[4] If $\underline{e} \leq s < e_x^*$, then $B_x(e_x^*) - (1-x) \cdot \tilde{D}(e_x^*|s) - (B_x(\underline{e}) - (1-x) \cdot \tilde{D}(\underline{e}|s)) = B_x(e_x^*) - (1-x) \cdot D(e_x^*) - (B_x(\underline{e}) - (1-x) \cdot D(\underline{e})) > 0$, since with $\alpha = 0$ we have $\tilde{D}(e|s) = D(e)$ and since e_x^* maximizes $B(e) - D(e)$. Hence $\tilde{e}_x \geq e_x$. But similarly, for $s = e_x^*$, then $B(s) - D(s) = B(e_x^*) - D(e_x^*) > B(e) - D(e)$ for every $e > e_x^* = s$, hence $\tilde{e}_x \leq s = e_x$. Thus $\tilde{e}^* = e_x$.

[5] In expositions, I refer to periods as *days* and to time within a period as *hours*.

until her currently perceived marginal disutility is equal to her marginal benefits of 3. I denote the time at which she plans to stop on day x by $\tilde{e}_x(s)$, the total hours she plans to work after having worked for s hours. She perceives her marginal disutility after studying for e hours to lie between her current marginal disutility, $D'(s)$, and her actual marginal disutility after e hours of studying, $D'(e)$:

$$\underbrace{\tilde{D}'(e|s)}_{\text{Perceived } D'} = (1-\alpha)\overbrace{D'(e)}^{\text{Actual } D'} + \alpha \underbrace{D'(s)}_{\text{Current } D'} = \frac{1}{2}(D'(e) + D'(s))$$

At the start of the first day – period $x = 0$ – Anna hasn't studied at all and $s = 0$.[6] So she thinks that her marginal disutility after e hours of studying will be $\tilde{D}'(e|0) = \frac{1}{2}D'(e)$. She plans to work for $\tilde{e}^*(0)$ hours (today and on all future days), and $\tilde{D}'(\tilde{e}^*(0)|0) = 3$ implies $1/2 \cdot D'(\tilde{e}^*(0)) = 3$, which implies $\tilde{e}^*(0) = 6$. Anna plans to study for 6 hours and thus starts studying. After 2 hours of studying, the current marginal disutility is $D'(2) = 2$. Anna now plans to study for $\tilde{e}^*(2)$ hours in total, with $\tilde{D}'(\tilde{e}^*(2)|2) = 3$ – the first order condition as she perceives it now. This leads to $1/2 \cdot (D'(\tilde{e}^*(2)) + D'(2)) = 3 \implies \tilde{e}^*(2) = 4$ hours. Finally, once she has completed 3 hours of studying, the current marginal disutility is $D'(3) = 3$, so that $\tilde{e}^* = 3$ and Anna stops studying. The same logic holds on all days (given that on past days she worked optimally), so that Anna stops every day after 3 hours of studying.

The same logic applies when the returns to effort are decreasing rather than constant, which leads to Proposition 1 that Anna works optimally, with overoptimistic plans until she stops. All proofs can be found in the appendix.

Proposition 1. *Let $D(.)$ be a strictly convex function with $D'(\cdot) \to \infty$, let $\alpha \in [0,1)$, and $B(.)$ be both differentiable and linear or concave. Let e^* be the optimal amount of work, $\tilde{e}_x$ be the actual amount of work done, and $\tilde{e}_x(s)$ the perceived optimal amount of work after s hours of work in period $x < 1$. Then $\tilde{e}_x = e^*$, and $\tilde{e}_x(s) > e^*$ $\forall s < e^*$.*

Proposition 1 relies on momentary work decisions. If Anna had to make an irreversible (or hard-to-reverse) choice, then she would commit to working too much.

[6] I write $\tilde{e}(s)$ rather than $\tilde{e}_0(s)$ in the exposition of the example to reduce notation and as there is no ambiguity about x.

Studying is unlikely to be about irreversible choices, but working on a common project in a team may consist of few and committed choices that lead to overcommitting. This shows that, even under concave benefits, overoptimistic beliefs over future work may be costly.

The reason for overcommitting rather than undercommitting stems from the fact that the optimal decision under concave benefits depends on her *final* (and hence highest) marginal disutility. Since the highest marginal disutility is necessarily always underestimated (at all other times the person is *less* tired) she is overoptimistic.

4 Convex Benefits

This section explores bounded convex benefits, both the general case and the special case of all-or-nothing tasks which cover situations with fixed thresholds for success and failure, where the question is whether to go all-in: passing an exam, getting a promotion, building a profitable startup.

4.1 Extended Example and Preliminary Results

This extended example provides intuition and context for the formal results in the following subsections.

Consider Beth, a student who is studying for an exam. Time in days is continuous: the first day is day 0 and the day of the exam is day 1. Beth knows that she will receive a B for sure in her final if she does nothing but attend the required lectures. If in addition she studies 5 hours a day on average, she receives an A for sure, which has an additional total benefit of 12.5 over a B.[7]

Suppose that $\alpha = 0.5$ and that the daily disutility is quadratic: $D(e) = \frac{e^2}{2}$ so that $D'(e) = e$. Her benefits from exerting on average E per day by the time of the exam are $B(E) = \mathbb{1}(E \geq E_F)B_F$, where $B_F = 12.5$ and $E_F = 5$. Let $\bar{e}_x$ be the

[7] Total benefits of B means that the person values these benefits as if they received a daily flow utility B on each day from $x = 0$ to $x = 1$. Thus a total benefit of 12.5 is equivalent to receiving a daily benefit of 12.5.

number of hours she would have to study daily starting on day x in order to get an A, i.e. $\bar{e}_x = \frac{E_F - E_x}{1-x}$.

Let us now solve how much Beth studies each day. Note that $B_F = 12.5 = \frac{5^2}{2} = D(5) = D(\bar{e}_0)$, so that studying efficiently has 0 net utility. Let us say that Beth faces a net-0-utility task on day x if $R(x) = 0$ where $R(x) := B_F - (1-x)D(\bar{e}_x)$ – that is, when the actual net utility from studying efficiently is 0. Since Beth underestimates the cost of the task at the start of the day, and because she would overestimate the cost of the task if she worked for $\bar{e}_x$ hours, she perceives it worth doing at the start of the day, and as not worth doing if she worked for $\bar{e}_x$ hours. She therefore starts studying, but stops before reaching $\bar{e}_x$, stopping the moment she is indifferent between studying and not studying. Thus $\tilde{e}_x$ – the time at which she stops – satisfies $B_F - (1-x)\tilde{D}(\bar{e}_x|\tilde{e}_x) = 0$. Since $R(x) = 0$, we can replace B_F by $(1-x)D(\bar{e}_x)$ to get $(1-x)D(\bar{e}_x) - (1-x)\tilde{D}(\bar{e}_x|\tilde{e}_x)$ which we can expand in several steps as $(1-x)\left(D(\bar{e}_x) - \frac{1}{2}D(\bar{e}_x) - \frac{1}{2}D'(\tilde{e}_x)\bar{e}_x\right) = (1-x)\frac{1}{2}\left(D(\bar{e}_x) - D'(\tilde{e}_x)\bar{e}_x\right)$. This utility is 0 if $D(\bar{e}_x) = D'(\tilde{e}_x)\bar{e}_x \iff \frac{\bar{e}_x^2}{2} = \tilde{e}_x\bar{e}_x \implies \tilde{e}_x = \frac{\bar{e}_x}{2}$.

Next, let us note the change in $R(x)$ when $\tilde{e}_x = \frac{1}{2}\bar{e}_x$:

$$\begin{aligned}
\frac{d}{dx}(B_F - (1-x)D(\bar{e}_x)) &= D(\bar{e}_x) - (1-x)D'(\bar{e}_x)\frac{d}{dx}\bar{e}_x \\
&= D(\bar{e}_x) - (1-x)D'(\bar{e}_x)\frac{\bar{e}_x - \tilde{e}_x(x)}{1-x} \\
&= D(\bar{e}_x) - D'(\bar{e}_x)(\bar{e}_x - \tilde{e}_x(x)) \\
&= \frac{\bar{e}_x}{2} - \bar{e}_x\frac{1}{2}\bar{e}_x \\
&= 0
\end{aligned}$$

Thus as long as Beth works in this manner on a net-0-utility task, the task remains a net-0-utility task in future periods.

Hence we have shown that Beth studies for $\frac{\bar{e}_x}{2}$ hours every day. Plugging this into the IVP, we get that $\dot{E}_x = \tilde{e}_x = \frac{\bar{e}_x}{2} = \frac{E_F - E_x}{2(1-x)}$. Solving this ODE, we get $\bar{e}_x = \frac{E_F}{\sqrt{1-x}}$.[8] Thus Beth daily studies for $\tilde{e}_x = \frac{E_F}{2\sqrt{1-x}}$ hours, which is plotted in Figure 1.

Thus, every day Beth starts studying with the intention of studying, only to give up half-way through when she becomes convinced that doing so is not worth doing. Because she does not study efficiently, the efficient effort $\bar{e}_x$ to achieve an A is strictly increasing over time – but she is studying enough every day, so that overall disutility to achieve an A stays constant, as there are fewer days left during which she has to study. This leads her to increase her effort ever higher, so high in fact that the disutility from completing the task goes to infinity.[9]

Figure 1 plots the numerical solutions when the benefits are slightly higher at $B_F = 13$ and slightly lower at $B_F = 12$. This highlights that the pattern of starting and stopping is not a special case due to picking a task where Beth should be indifferent.

Figure 1 also highlights the central role of day x_q. As it turns out, before day x_q, Beth perceives herself as working inefficiently, thereafter she perceives herself as working efficiently – and she in fact is working efficiently.

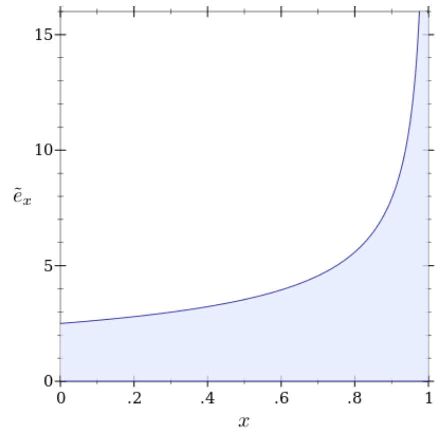
For example, when $B_F = 13$, Beth works efficiently from day $x_q \approx 0.66$ on. Despite studying less on the days leading up to x_q than planned, she studied enough so that

[8] We have

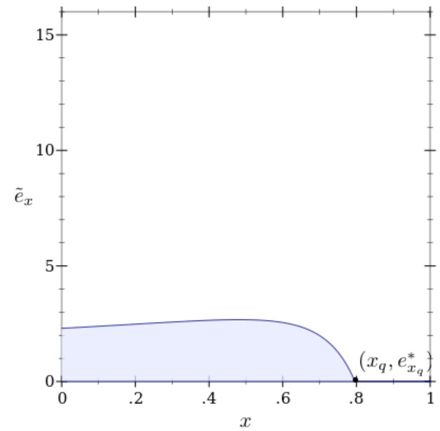
$$\begin{aligned}
\dot{E}_x = \frac{E_F - E_x}{2(1-x)} &\implies \frac{\dot{E}_x}{E_F - E_x} = \frac{1}{2(1-x)} \\
&\implies \log(E_F - E_x) = \frac{1}{2}\log(1-x) + \log(E_F) \\
&\implies E_F - E_x = E_F(1-x)^{\frac{1}{2}} \\
&\implies \frac{E_F - E_x}{1-x} = \frac{E_F}{\sqrt{1-x}}
\end{aligned}$$

[9]

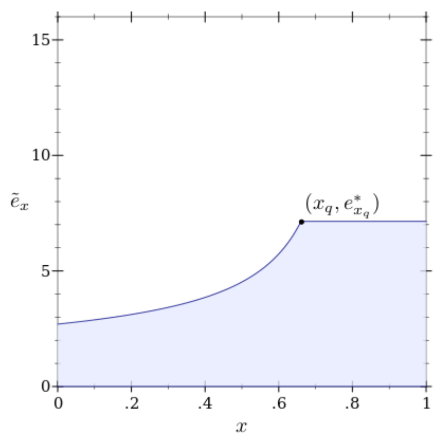
$$\begin{aligned}
\int_0^1 D(\tilde{e}_x)dx &= \int_0^1 \frac{\tilde{e}_x}{2}dx \\
&= \int_0^1 \frac{1}{2}\frac{E_F^2}{4(1-x)}dx \\
&= \frac{E_F^2}{8}\left(log(1) - log(0)\right) \\
&= \infty
\end{aligned}$$



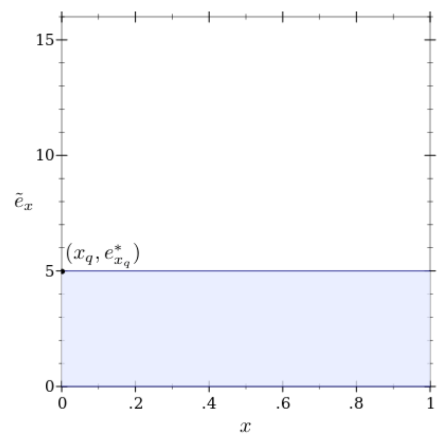
(a) $B_F = 12.5$, $E_1 = 5$, $\alpha = 0.5$



(b) $B_F = 12$, $E_1 \approx 1.87$, $\alpha = 0.5$



(c) $B_F = 13$, $E_1 = 5$, $\alpha = 0.5$



(d) $B_F = 13$, $E_1 = 5$, $\alpha = 0.03$

Figure 1: Plots the daily effort $\tilde{e}_x$ for different B_F and α. The x-axis plots the fraction of days until the final deadline. E_1 is the total effort done by the deadline. The plots for $B_F = 12.5$ and for $B_F = 13$ with $\alpha = 0.03$ are based on the exact solution $\tilde{e}_x = \frac{E_F - E_x}{2(1-x)^{\frac{1}{2}}}$ (see main text for derivation). All other plots are computed numerically. x_q is the first time where the person works efficiently, either by not working at all (as for $B_F = 12$) or by working $\bar{e}_x = \frac{E_F - E_x}{1-x}$ (as for $B_F = 13$). For $B_F = 12.5$, we have $x_q = 1$ – i.e. the person never works efficiently.

studying is still worth it and she perceives it as such even when she is at her most tired. However, she would be better off if she had smoothed the effort over time. When $B_F = 12$, Beth gives up studying once and for all around day $x_q \approx 0.80$. She has achieved so little that even when she is at her most rested, she (accurately) perceives studying as not worth doing.

4.2 Characterization for General Benefits

Notation: Let $e^*_x = \arg\max_e B_F \mathbb{1}(e \geq \bar{e}_x) - (1-x)D(e)$ – the *optimal effort* in period x, given the (potentially suboptimal) amount of effort E_x the person has completed so far. Let us parameterize the net perceived utility over degrees of bias by writing $U_\alpha(x, e, s) = B(E + (1-x)e) - (1-x)(1-\alpha)D(e) - (1-x)\alpha D'(s)e$ for the net perceived utility of a person with projection bias α. When the dependence on α is implicit, I simply write $U(x, e, s)$.

The results apply to general convex benefits subject to the following sufficient assumption:

Assumption 1 (Bounded Effort)**.** *Assume that there is some $E_F > 0$ s.t. $B(E) = B(E_F)$ for all $E \geq E_F$.*

The results also hold for all-or-nothing tasks (defined further below).

We finally get the following characterization of behavior in a given period.

Proposition 2 (Characterization for Convex Benefits)**.** *Consider B and D, with B satisfying either of the following:*

- *continuous and strictly convex on $[0, E_F]$ and such that $\tilde{e}_x$ is Lipschitz continuous in x*
- $B(E) = \mathbb{1}(E \geq E_F) \cdot B_F$ *for some* E_F, $B_F > 0$

and D strictly convex with $D'(e) \to \infty$ as $e \to \infty$. Then there exist $\bar{\alpha}$ and x_q s.t.

- $x_q = 0 \iff \alpha \leq \bar{\alpha}$
- $x \geq x_q$ *holds if and only if* $\tilde{e}_x \in \arg\max_e U_\alpha(x, e, \tilde{e}_x)$ *and both imply* $\tilde{e}_x = e^*_{x_q}$

Roughly, Proposition shows that the person's decision-making changes qualitatively on day x_q. Let us focus on the person's perception of her behavior at the end of each day. On any day x before x_q, she stops studying after $\tilde{e}_x$ hours, but she does not perceive that amount as optimal ($\tilde{e}_x \notin \arg\max_e U_\alpha(x, e, \tilde{e}_x)$). Instead she quits because she currently would prefer to have studied less. From day x_q on, she studies for the optimal $e^*_{x_q}$ hours, and she perceives it as optimal.

The proposition also states that the person perceives herself as making the optimal choice from the first day if her bias is low enough. When her bias is large, then until day x_q she stops each day believing she should have worked less. Having a larger bias means that the perception of the costs of studying go through larger swings: she underestimates the costs of studying more when rested, and overestimates them more when tired. If studying is worth doing, she will start studying but is more likely to mistakenly quit as she grows tired, overestimating the costs of studying. If studying is not worth doing, she is more likely to mistakenly start studying only to quit once she has grown sufficiently tired.

For Beth, making the optimal choice from day x_q on means that she either does not work at all, or that she works $\bar{e}_x$ hours, as illustrated. In Figure 1 (a), $x_q = 1$, Beth always quits believing she should have studied less, while in (b) and (c), she eventually perceives herself as making the optimal choice by not studying at all (in (b)) or studying fully (in (c)). Finally, in (d) Beth faces the same benefits as in (c), but because I assume that she is less projection-biased ($\alpha = 0.03$), she studies efficiently from the first day ($x_q = 0$).

4.3 Characterization for All-or-Nothing Benefits

All-or-nothing tasks pay total benefits B_F if the person completes E_F by the final period: $B(E) = \mathbb{1}(E \geq E_F) \cdot B_F$. This section explores how Beth's behavior on such a task changes as the benefits from completing such an all-or-nothing task increase. Here, I will provide an informal intuition for Proposition 3 and then state it. The formal proof is in the Appendix.

First, let us apply Proposition 2 to the case of all-or-nothing tasks. After day x_q, Beth makes the optimal choice, which for all-or-nothing tasks mean either 0 or $\bar{e}_{x_q}$ hours. Before day x_q, $\tilde{e}_x$ is *not* perceived to be optimal at the time of stopping

($\tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)$). Not perceiving 0 hours as optimal at the start implies that Beth must study at least a little. Similarly, not perceiving $\bar{e}_x$ as optimal after $\bar{e}_x$ hours of studying implies that Beth must quit at some strictly earlier time. We can summarize this by saying that before day x_q, Beth studies inefficiently, and after x_q she makes the optimal choice – either not to study at all, or to study fully.

Next, let us consider how x_q and $e^*_{x_q}$ change as the benefits increase. Clearly when benefits are really low, then even though Beth is biased, she realizes that the task isn't worth doing and doesn't study at all. Thus $x_q = 0$ and Beth makes the optimal choice from day 0 by not studying at all.

As the benefits become larger, Beth studies inefficiently until $x_q > 0$, and never studies thereafter. As the benefits increase, x_q increases and the total amount E_1 that Beth has studied by the deadline increases, yet she falls short of her target. She is therefore worse off as the benefits increase. This case is illustrated in Figure 1 (b).

As the benefits reach and pass the critical threshold B_C, x_q is 1 and Beth reaches her target. Yet as Figure 1 (a) highlights, she studies potentially very inefficiently. Increasing benefits past this point is unambiguously beneficial for Beth: Beth studies more on all days before x_q and studies efficiently after day x_q. Since x_q occurs sooner, Beth studies efficiently earlier on, so she incurs lower costs from studying while reaping higher benefits. This case is illustrated in Figure 1 (c).

Finally, for sufficiently high benefits, Beth studies and meets her daily targets from the first day on. This case is illustrated in Figure 1 (d).

Proposition 3. *The disutility of effort is strictly convex, with $D''(\cdot) > d$ for some $d > 0$ and $\lim_{e \to \infty} D'(e) = \infty$. Consider an all-or-nothing task (E_F, B) and some $\alpha < 1$. Then for $B_H = (1-\alpha)D(E_F) + \alpha D'(E_F) \cdot E_F$ and $B_L = (1-\alpha)D(E_F)$ there exists some B_C with $B_H > B_C > B_L$ such that:*

- *$B \in [B_H, \infty)$: then $x_q = 0$ and $\tilde{e}_x = \bar{e}_0$*
- *$B \in [B_C, B_H]$: then $x_q \in (0, 1)$, $\tilde{e}_x = \bar{e}_{x_q}$ for $x \geq x_q$ and $E_1 = E_F$*
- *$B \in [B_L, B_C]$: $x_q \in (0, 1)$, $\tilde{e}_x = 0$ for $x \geq x_q$, and $E_1 < E_F$*
- *$B \in [0, B_L]$: then $x_q = 0$, $\tilde{e}_x = 0$ for all x, and $E_1 = 0$*

Moreover, $x_q(B)$ is continuous in B, strictly increasing in B from $x_q(B_L) = 0$ to $x_q(B_C) = 1$, and then strictly decreasing to $x_q(B_H) = 0$. $E_x(B)$ is continuous and strictly increasing in B for $B \in [B_L, B_H]$.

The next proposition provides a partial characterization of B_C, the threshold for completing tasks:

Proposition 4. *Consider B_C from Proposition 3 in the all-or-nothing case. Then:*

- $D''' < 0 \implies B_C < D(E_F)$
- $D''' > 0 \implies B_C > D(E_F)$
- $D''' = 0 \implies B_C = D(E_F)$

The Proposition shows that when $D''' < 0$ the person completes some tasks that are not worth completing *even when done efficiently*. And that, when $D''' > 0$, there are some tasks that are worth doing efficiently that she fails to complete at all. Only under quadratic disutility does she complete a task if and only if it is worth doing efficiently. Thus observing a person complete or not complete a task does not allow us to infer whether her optimal choice is to commit to completing the task or not doing so. Of course, committing to completing the task efficiently is better than completing the task inefficiently. But it may be that the person would be even better off not doing the task at all.

5 Multi-Tasking with Concave Benefits

We know from Section 3 that a biased person works optimally on a single task with decreasing returns, despite overestimating how much she will work. To explore how such overestimation in one task can affect work on another task, let us now consider a person who divides her time each day between a short-term and a long-term task, each of which has decreasing returns to effort.

5.1 Example

To illustrate, suppose that Elaine has to finish her thesis (the long-term task, indexed by L) by the final day, and that every day she has some homework to do (the short-term task, indexed by S). She has projection bias $\alpha = 1/2$. The first day is $x = 0$, the final day is $x = 1$. The daily benefits from exerting effort $\tilde{e}_S$ are $B_S(\tilde{e}_S)$ and the benefits from exerting total effort $\int_0^1 \tilde{e}_L de$ on the long-term task are $B_L(\tilde{e}_L)$.[10] The disutility on some day from exerting $\tilde{e}_S$ and $\tilde{e}_L$ is $D(\tilde{e}_S + \tilde{e}_L)$. Suppose for illustration that $B_S(e) = B_L(e) = \frac{3}{2}\log(e)$, and suppose that $D(e) = \frac{e^2}{2}$.

Let us now solve this problem. In the proofs I show formally that Elaine switches or stops if she perceives this to be optimal *at the time of switching or stopping*. Therefore, to find the time of switching, we find a time at which Elaine perceives switching *right now* as optimal, and similarly for stopping. Then, consider the time s at which Elaine switches on day x from working on the thesis to homework. Clearly, $s = \tilde{e}_L$, since she switches once she is done with the thesis. At this time, she maximizes her current perceived utility and she plans to exert $\tilde{e}_L$ on the long-term task and $\tilde{e}_{S|P}$ (to indicate this is *planned*, not actual) on the short-term task every day. Since she maximizes $B_L(E_x + (1-x)e_L) + (1-x)B_S(e_S) - (1-x)\tilde{D}(e_L + e_S|s)$ (where $E_x = \int_0^x \tilde{e}_L(x)$ is effort spent on the long-term task by period x), the FOC yield:

$$
\begin{aligned}
& B_S'(\tilde{e}_{S|P}) = \tilde{D}'(\tilde{e}_L + \tilde{e}_{S|P}|s) = B_L'(E_x + (1-x)\tilde{e}_L) \\
\iff & \frac{3}{2}\frac{1}{\tilde{e}_{S|P}} = \frac{1}{2}\left(\tilde{e}_L + \tilde{e}_{S|P} + s\right) = \frac{3}{2}\frac{1}{E_x + (1-x)\tilde{e}_L}
\end{aligned}
$$

At $x = 0$, we have $E_x = E_0 = 0$ (she has not yet completed any work), so that $\tilde{e}_{S|P} = \tilde{e}_L$. Moreover, $s = \tilde{e}_L$, and from the equality of the left-most and the right-most expressions, we see that $\tilde{e}_{S|P} = \tilde{e}_L = s$. So the remaining equalities yield $\frac{3}{2}\frac{1}{\tilde{e}_L} = \frac{1}{2}(3\tilde{e}_L) \implies \tilde{e}_L = 1$. Now suppose that she works this amount on all days until x, i.e. $\tilde{e}_L(x') = 1$ for all $x' < x$. Then $E_x = x$, and so the FOC on day x becomes:

[10] Note that total effort is measured in the average hours per day: if daily effort is $\bar{e}$ on all days, then total effort is $\int_0^1 \bar{e} de = \bar{e}$.

$$\frac{1}{\tilde{e}_{S|P}} = \frac{1}{2}\left(\tilde{e}_L + \tilde{e}_{S|P} + s\right) = \frac{1}{x + (1-x)\tilde{e}_L}$$

It is easy to see that $\tilde{e}_L = s = \tilde{e}_{S|P} = 1$ solves this equation, since it is the same as we solved at time $x = 0$! This shows that Elaine works $\tilde{e}_L = 1$ on her thesis every day. Since $\tilde{e}_{S|P} = \tilde{e}_L$, this is also how much she plans, at the time of switching, to work on her homework.

However, as she starts her homework, she grows more tired and stops earlier than anticipated. In fact, she stops once $B'_S(\tilde{e}_S) = \tilde{D}'(\tilde{e}_L + \tilde{e}_S | \tilde{e}_L + \tilde{e}_S) = D'(\tilde{e}_L + \tilde{e}_S)$, which implies $\frac{3}{2}\frac{1}{\tilde{e}_S} = \tilde{e}_S + 1$, so that $\tilde{e}_S = \frac{\sqrt{7}-1}{2} \approx 0.82$.

The optimal effort has $e^*_L = e^*_S = e^*$ since $B_S(e) = B_L(e)$, where e^* solves the FOC $\frac{3}{2}\frac{1}{e^*} = 2e^* \implies e^* = \frac{\sqrt{3}}{2} \approx 0.87$.

Thus it turns out that Elaine works more than she should on the thesis – and this is because at the time she switches, she underestimates future costs and hence overestimates how much she will work in total. Because of this, she ends up working less than she should on her homework. In total, she works more than she should: at the time at which she would stop if she worked optimally, she has worked substantially less than is optimal on her homework, thus the marginal benefits from working on the homework are larger than the costs from working. As I show below, all of these results hold more generally.

5.2 Multi-Tasking Model

The model from Section 2 no longer provides a unique answer, because the person is indifferent about the order in which she works on the tasks, despite the fact that this order affects the final outcome. I therefore assume that the person works on the tasks in sequence, and extend the framework to determine both the time at which she switches between the tasks and the time at which she stops working on the second task.[11] At the end, I provide a natural condition that leads her to work on the short-term task first.

[11] Every time we add an (implicit or explicit) choice variable, the problem may allow many different solutions under projection bias. Here that choice variable is the order in which the person works on the tasks.

Environment A projection-biased person can work on 2 tasks each period. The flow disutility in period x depends on total effort in x, so that we can write it as $D(e_1(x) + e_2(x))$. The monetary benefits are task-specific, and one is a short-term task that yields a flow benefit $B_S(e_t(x))$ at the end of the period x, and one is a long-term task that yields a total benefit $B_L(\int_0^1 e_t(x)dx)$ in period $x = 1$.[12]

Behavior The person works consecutively on the tasks, first on task 1 for a duration e_1, then on task 2 for a duration e_2. I analyze both the situation when the short-term task is done first and when the long-term task is done first, and provide a sufficient condition (*time-sensitivity*) under which it is natural that the short-term task is done first.[13]

Under multi-tasking, we need to distinguish carefully between the optimal plan for *today* (the current period x) and the optimal plan for all future days (periods $x' > x$). The reason is that when working on the second task today, the person can no longer re-optimize the amount of work to exert on the first task, but she can re-optimize how much she plans to work on the first task on future days. We also need to formalize how the person trades off the benefits and costs from (the infinitesimally short) *today* against the future, which requires taking seriously that the continuous-time setup approximates a setup with T periods as T grows large and as the time per period shrinks.

Then today's *flow benefits* from working e_S on a short-term task are $B_S(e_s)$. Let $p_L(x)$ be the amount the person *plans* to work on the long-term task on all future days. Then the perceived flow benefits from working e_L on the long-term task are given by $B_L'(E + (1 - x)p_L(x)) \cdot e$. It is linear in today's effort e_L since today's effort is too small to affect the final stock of effort. Finally, the flow disutility from exerting $e_S + e_L$ is given by $\tilde{D}(e_S + e_L|s)$.[14]

[12] Subscript S stands for *short-term*, subscript L for *long-term*.

[13] I do not consider situations where a person switches repeatedly between the two tasks.

[14] Formally, the continuous-time setup is limiting case as $T \to \infty$ of a discrete-time setup where T is the number of periods and the time per period becomes shorter and shorter. If we consider a single period to be of length $\Delta \to 0$ with $T \cdot \Delta = 1 - x$, then the benefit is $B_L(E + \Delta e + (T-1)\Delta \cdot p_L(x, \Delta)) - B_L(E + (T-1)\Delta \cdot p_L(x, 0)) = B_L'(E + (T-1)\Delta \cdot p_L(x, 0))(\Delta \cdot e + (T-1)\Delta \cdot (p_L(x, \Delta) - p_L(x, 0))) + O(\Delta^2) \to B_L'(E + (1-x)p_L(x, 0))\Delta \cdot e$ if $p_L(x, \Delta) - p_L(x, 0) \to 0$ as $\Delta \to 0$ – i.e. if future plans are continuous in E as is the case here.

Let us write $U_S(x, E, e_S, e_L, s) := U_L(x, E, e_L, e_S, s) := B_S(e_S) + B_L'(E + (1 - x)p_L(x, E, s)) \cdot e_L - \tilde{D}(e_S + e_L|s)$ for the perceived net flow utility from exerting e_S on the short-term and e_L on the long-term task in period x after s hours of working when having completed E on the long-term task. Note that p_L is determined by optimization over future perceived utility, and that U_S and U_L merely flip the order of the two tasks for notational purposes.

Denote by $V_S(x, E, e_S, s) = \max_{e_L} U_S(x, E, e_S, e_L, s)$ the maximum (perceived net flow) utility when choosing e_S for the short-term task, and $V_L(x, E, e_L, s) = \max_{e_S} U_L(x, E, e_L, e_S, s)$ the maximum (perceived net flow) utility when choosing e_L for the long-term task. Let i be the first and $j \neq i$ be the second task, with $i, j \in \{S, L\}$. Then we get the following generalization of Definition 2:

Definition 4 (Multiple Effort Decisions)**.** *Consider a projection-biased person who works on two consecutive tasks $i, j \in \{S, L\}$ in that order, with $i \neq j$. Then she exerts effort $\tilde{e}_i$ on task i given by*

$$\tilde{e}_i(x) = \inf\{s : V_i(e, E_x, x, s) < V_i(s, E_x, x, s),\ \forall e > s\}$$

and she exerts effort $\tilde{e}_j$ on task j given by

$$\tilde{e}_j(x) = \inf\{s - \tilde{e}_i : U_j(x, E_x, e_j, \tilde{e}_i, s) < U_j(x, E_x, s - \tilde{e}_i, \tilde{e}_i, s),\ \forall e_j > s - \tilde{e}_i\}$$

The first task is determined as before, except that she switches (rather than stops) when the utility V of switching is perceived higher than that of continuing to work on the first task. The second task takes the amount of effort on the first task as a given, since that decision can't be undone. At time s, the person has exerted effort $s - \tilde{e}_i$ on the second task, which explains the condition $e_j > s - \tilde{e}_i$.

5.3 Multi-Tasking Results

Let us now state the general proposition for multi-tasking.

Proposition 5. *Consider the multi-tasking setup with two consecutive tasks S (short-term) and L (long-term). Let $\tilde{e}_S(x)$ and $\tilde{e}_L(x)$ be the actual amount of effort* *spent on the short-term and long-term task in period x, and $\tilde{f}(x) := \tilde{e}_S(x) + \tilde{e}_L(x)$ be the total effort in period x. Let e_s^*, e_L^*, and f^* be the optimal effort levels, and all strictly positive. Let $\tilde{E}_x$ be the actual total effort and E_x^* the optimal total effort done by period x. Then*

1. *If the long-term task is done first each period, $\tilde{e}_L(x) = \tilde{e}_L > e_L^*$, $\tilde{f}(x) = \tilde{f} > f^*$, $\tilde{e}_S(x) = \tilde{e}_S < e_S$.*
2. *If the short-term task is done first each period and $1 > x' > x$, then $\tilde{e}_S(x) \geq \tilde{e}_S(x')$, $\tilde{e}_L(x) \leq \tilde{e}_L(x')$, $f^* < \tilde{f}(x) \leq \tilde{f}(x')$, and $\tilde{E}_x < E_x^*$ with strict inequalities everywhere if and only if $\tilde{e}_S(x) > 0$ and $x(0,1)$.*

The Proposition states that the person works too much in total, working too much on the task done first each day and too little on the task done second. The case when the long-term task is done first is illustrated by the earlier example. When the short-term task is done first, the person works less and less over time on the short-term task, and increasingly more on the long-term task. She works less on the long-term task than she planned, so that the marginal benefit from working on the long-term task increases steadily. I provide one reason why people may naturally perform the short-term task before the long-term task.

To provide an intuition for why the effort is constant when the long-term task is done first, but not when the short-term task is done first, let us revisit the example above. Every day, Elaine works as much on her thesis as she planned to do *at the time of switching.* Thus on future days when she considers whether to switch at the same time, her perceived disutility is the same as at all past times when she switched, and she followed her plans according to that perceived disutility. Thus it is clear that she wants to continue following this plan, which requires constant effort. If instead Elaine works on the short-term task first, then on day 0 she switches at time $\bar{s}$, say, at which time she plans to work $\tilde{e}_{L|P}$ on her thesis. However, she ends up working less on her thesis than planned. Therefore on a future day, she will have worked less than planned at time $\bar{s}$, so the marginal benefits from working on the thesis are strictly higher, and so she stops working on her homework earlier.

Definition 5. *A task is time-sensitive if the benefit from the task not only depends on how well the task was done, but also how early in the period it was completed.*

Similarly, it is time-insensitive if the benefit is independent from the time of day when the task is completed. Consider two tasks that pay $B_1(e,s) = B_1(e) + b_1(s)$ and $B_2(e,s) = B_2(e)$ with $b_1(s) > 0$ and $b_1'(s) < 0$. Then benefits B_1 are time-sensitive and benefits B_2 are time-insensitive.

If task 1 is time-sensitive and task 2 is not, then it is always better to do task 1 in full first. Any amount of work spent on task 2 simply delays completion on task 1 and thus reduces overall benefits for no gain (holding total effort on both tasks constant). It seems natural that a long-term task is close to time-insensitive, since usually there will be no benefit from working on the long-term task at 11am rather than at 5pm – it is primarily about the total amount of work done. It may however be beneficial to send a report to a client at 2pm rather than at 5pm, if only to signal to the client how important they are. This leads to the following corollary:

Corollary 1. *Consider a person who works on a time-insensitive long-term task as well as on time-sensitive short-term tasks each day. Then the person works first on the short-term tasks each day and hence works too much on the short-term tasks at the expense of the long-term task.*

6 Discussion

Throughout the paper, I have illustrated the intuition of the formal statements with examples about students studying for exams. Despite this focus, the results hold for any sufficiently large long-term task such that a person working on it repeatedly grows tired and rested. Given the large-scale nature of the task, it applies particularly to important domains and projects. It applies to the student studying for a degree, the wannabe entrepreneur who founds a startup requiring long hours, the employee working towards a promotion, and the athlete who wants to get in shape for tournaments.

Let us now compare the behavioral and welfare implications of projection bias to those from other models that can generate time-inconsistent behaviors, such as present bias, self-control problems, or overconfidence. Since time-inconsistent preferences imply that a revealed preference approach cannot identify which of the choices is preferred, even less which is optimal, the welfare implications often rely on the model used. Projection bias may provide alternative welfare implications to the other models, in particular in relation to the benefit of commitment.

In this section, I first consider observable behaviors that are either inconsistent, or hard to reconcile, with other models. Next I show that, even when behaviors can be reconciled with other models, projection bias often leads to different welfare consequences and interpretations. Commitment makes the agent worse off under concave benefits, and completing or failing to complete a task does not reveal whether the task is worth doing. Thus different channels lead to different responses to time-inconsistent behavior. I therefore conclude with a discussion of richer data that can be used to measure the contribution of different channels.

Observable Behaviors Consider first the all-or-nothing setup in Section 4. When benefits are sufficiently high, a projection-biased person works inefficiently little in the early days, yet works efficiently on the later days.[15]

Let us now consider whether present bias, self-control, or overconfidence, can generate such behavior, and let us suppose that the person completes the task, but possibly inefficiently. Under present bias (O'Donoghue and Rabin (1999)), both naive and sophisticate, the person will never work the same amount on all remaining days – since then she would have a strict preference to work strictly less today. Now consider a model of self-control, by which I mean any model where a person may exert costly effort to overcome their present bias. It is certainly possible that such a person would exert self-control on all days. It seems rather unlikely, however, that the person would decide *not* to exert self-control in the early days, and yet *do* exert self-control in the final days. Since early self-control is a substitute for later self-control, if anything it is more plausible to expect more self-control early on. And finally, consider a person who is overconfident about all future, but not her current, effort costs. Then she always plans on working strictly more tomorrow than today (where she perceives the costs correctly).[16] Similarly, projection bias allows a person to work less and less each day until she stops working, which again

[15]Concretely, benefits B_F have to lie in (B_C, B_H), see Proposition 3.

[16]This is identical to a model of overconfidence in one's future output. If the person is overconfident in the value of *all* her output, including right now, then she always plans to work the same, but too much. In none of the cases can it generate the pattern of working inefficiently at first, efficiently later.

does not happen under the other models.

Second, consider the multi-tasking case, where the projection-biased person works too much on short-term tasks and too little on the long-term task – and as a consequence, she will work too much and increasingly more each day. A similar pattern can be reconciled with naive present bias when the short-term task yields immediate *utility* (not monetary) benefits, which incentivise the person to put more effort into the short-term task.

Overcommitment and Welfare Differences Another difference in the multi-task setup is that present bias leads to inefficiently low daily work (on both the short-term and the long-term task), while projection bias leads to inefficiently high daily work. Committing the person to her initial plans – as present bias would suggest – leads to overcommitment if projection bias is at play.

This highlights a general welfare difference between the models. Under projection bias, the person may over- or under-estimate her future disutility, depending on whether current marginal disutility is low or high. This is why it is not always the earlier, more ambitious plan that is better. Instead the less ambitious plan of quitting may be correct. With decreasing benefits, the person will be worse off under commitment, since she will overcommit. This is in contrast to present bias and self-control, where the person would be better off under commitment.[17]

Less Revealed Preference For all-or-nothing tasks, projection bias can cause a person to work too much by working on a task they should not work on, or to work too little by quitting a task that they should continue to work on. This limits how much we can learn about underlying preferences from observed choices. Proposition 4 shows that it is possible that a person completes a task despite it not being worth doing (if $D''' < 0$), or that she fails to complete some worthwhile tasks (if $D''' > 0$). Hence, absent strong assumptions on the disutility, the completion or failure to complete a task does not even allow us to infer whether the person should have completed the task efficiently.

[17] Another reason why people may mistakenly commit is when they underestimate their present bias and pick a commitment that is too weak. John (2020) finds evidence for people choosing too weak commitments for savings contracts. Even in this case, the welfare interpretation is still one in which sufficiently strong commitment is desirable.

Empirical Predictions The discussion so far highlights how difficult it may be to tell projection bias apart from other biases. This is easier if we also observe *planned* effort: the daily work planned is constant, but the level of effort fluctuates (with current willingness to work, if we observe it). Naive present bias on the other hand predicts no such fluctuations at the daily level, and it predicts lower work planned for today than for future days. Even when both types of biases are present, fluctuations in planned effort should be a more robust indication of projection bias. I am unaware of any finding that supports or refutes such fluctuations. While Kaur, Kremer, and Mullainathan (2015) have a long-running intervention, they did not collect data on the planned effort of workers at times of different tiredness, and the workers in their study faced different daily incentives.[18]

Given the central role of welfare, it may be helpful to measure how much people enjoy a given task – keeping in mind that such evaluations have their own challenges (Kahneman et al. (1993), Kahneman, Wakker, and Sarin (1997)). Bisin and Hyndman (2020) analyse data from Ariely and Wertenbroch (2002) and find that one of the student groups subject to deadlines performed better, but also disliked the assignment more than the student group without additional deadlines. While one has to be careful in how to interpret such retrospective evaluations, and the deadlines were set exogenously, such data may help separate the welfare impact of commitments from its behavioral impact.[19]

7 Conclusion

Throughout the paper, I illustrated situations where projection bias turns fluctuating tiredness into fluctuating plans, which leads to inefficient work. When rested, the misperceptions may taunt people with an optimistic vista of their goal, only to douse those hopes with exhaustion, in a repeating cycle. From the outside – and

[18]The closest they get to eliciting plans is to elicit commitments on the day of work and the night before – but only from those workers who are willing to commit, leading to a sample selected for stronger present bias. Contrary to both present bias and projection bias, they find no difference, which they attribute to uncertainty resolution.

[19]Galak, Kruger, and Loewenstein (2013) use retrospective evaluation to tease apart how much people liked consuming chocolates at a self-set pace versus an exogenously spread out in a setting where people also seem to misjudge their own preferences.

even from the inside – it is hard to tell whether to ride the optimistic highs or succumb to the pessimistic lows. Unlike models of present bias or self-control, the more ambitious plan may not be the virtuous path, but a fata morgana best to avoid. This contrasts with the view that ex ante plans reveal (long-term) preferences.

All this holds only if the fluctuations repeat and people naively learn little from these fluctuations. While such naiveté is more appropriate than may often be assumed, people regularly display some meta-sophistication: they realize and learn that they fall short of their own expectations, that they behave inconsistently, yet without a clearly articulated cause for this behavior.[20] While it is beyond this paper, it is an interesting question how projection-biased people respond when realizing their fluctuating plans, without realizing the source of the problem. They may mistakenly treat it as a pure self-control problem and attempt to exert self-control and self-signal (Bénabou and Tirole (2004), Battaglini, Bénabou, and Tirole (2005), Fudenberg and Levine (2006)). Personality traits such as perseverance and grit (Duckworth et al. (2007)) may also affect how quickly a person quits, and thus how well they do when facing changing plans. This might lead them to not let go when they should (Alaoui and Fons-Rosen (2021)). In all these cases, the welfare implications are ambiguous, since the mistake may be planning too much, rather than too little work. Thus a major challenge, relevant for many types of misperceptions, is to answer the question: what do people learn or do when we neither assume that people *must* learn nor that they *cannot* learn – and do they learn the right thing?

8 Acknowledgements

I am grateful to Matthew Rabin, David Laibson, Gautam Rao, and Josh Schwartzstein. I also thank Ben Bushong, Krishna Dasaratha, Anastassia Fedyk, Tristan Gagnon-Bartsch, James Hodson, Mats Köster, Botond Kőszegi, Annie

[20] See Le Yaouanq and Schwardmann (2019) for a nice experiment in which participants make initially overoptimistic forecasts about the effort they will put in, and adjust later forecasts appropriately. Note also however that in almost all the cases in which projection bias has been found to play a role, people involved had ample real-life experience, so that any persisting bias highlights that learning is not complete or may be context-specific. The adult subjects in Read and Van Leeuwen (1998) had a life-time of experience with eating foods, and yet they displayed the bias.

Liang, Ben Lockwood, Rani Spiegler, Neil Thakral, Linh T. Tô, and Gal Wettstein, as well as seminar participants at CEU, Harvard, Humboldt, the Hungarian Academy of Sciences, Liser, and LMU for helpful comments.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

A Proofs

A.1 Proofs for Section 3

Lemma 1 (Stop at optimum for strict concavity)**.** *Consider a function $F(e, s)$ that is strictly concave in e and $\frac{\partial^2 F}{\partial e \partial s} < 0$, so that $e(s) = \arg\max_e F(e, s)$ exists and is unique. Then $e(s)$ is strictly decreasing in s. Let $\tilde{e} = A$ where $A := \inf\{s : F(e, s) < \max_{\underline{e} \leq s} F(\underline{e}, s), \forall e > s\}$. Assume that $e(0) > 0$. Then we have $\tilde{e} = e(\tilde{e})$.*

Proof. First we show that $s - e(s)$ is strictly increasing in s. By strict concavity, $e(s)$ satisfies $\frac{\partial}{\partial e} F(e(s), s) = 0$. Taking derivatives of this expression wrt s, we get $\frac{\partial^2 F(e(s),s)}{\partial e \partial e} \frac{\partial e(s)}{\partial s} = -\frac{\partial^2 F(e(s),s)}{\partial e \partial s} \implies \frac{\partial e(s)}{\partial s} < 0$ since F is strictly concave in e and the cross partials are negative.

Since $e(s)$ is continuous (as $e(s) \leq e(0)$ is bounded), $\Delta(s) := s - e(s)$ is continuous. We have that $\Delta(0) = 0 - e(0) < 0$ and since $e(e(0)) < e(0)$ (since $e(s)$ is strictly decreasing in s), we have that $\Delta(e(0)) = e(0) - e(e(0)) > 0$. Hence there is some unique s_0 s.t $\Delta(s_0) = 0$.

For a given s with $\Delta(s) > 0$, we have $s > e(s)$. Since $e(s)$ maximizes $F(e, s)$ and $F(e, s)$ is strictly concave in e, $F(e, s)$ is strictly decreasing in e for $e > s$, hence $F(e, s) < F(s, s)$ for all $e > s$ and $s \in A$. Since $\Delta(s) > 0$ for all $s > s_0$, we have that $\tilde{e} \leq s_0$.

For a given s with $\Delta(s) < 0$, we have $s < e(s)$, so that $F(e(s), s) > F(s, s)$. Since $\Delta(s) < 0$ for all $s < s_0$, we therefore have that no $s < s_0$ is in A. Hence $\tilde{e} \geq s_0$.

It follows that $\tilde{e} = s_0$, thus $e(s_0) = s_0 \implies e(\tilde{e}) = \tilde{e}$. □

Proof of proposition 1.

Proof. Note that $F(e, s) := B(E_x + (1 - x)e) - \tilde{D}(e|s)$ satisfies all the conditions of Lemma 1. So $\tilde{e}_x$ is the global maximum of $F(e, s)$ when $s = \tilde{e}_x$, and the global maximum is determined by the FOC $(1 - x)B'(E_x + (1 - x)e) = (1 - x)\tilde{D}'(e|s)$. Since $e = \tilde{e}_x$ is the unique solution when $s = \tilde{e}_x$, we can substitute $\tilde{e}_x$ for both e and s: $B'(E_x + (1 - x)\tilde{e}_x) = \tilde{D}'(\tilde{e}_x|\tilde{e}_x) = D'(\tilde{e}_x)$. We can check that $\tilde{e}_x = e^*$ solves the FOC for all x and hence the original IVP: $E_x = xe^*$, so $B'(E_x + (1 - x)e^*) =$

$B'(e^*) = D'(e^*)$, which holds since it is the FOC for the optimum effort level. When $\tilde{e}_x$ is Lipschitz continuous in x, the IVP has a unique solution, and it is easy to see that for a fixed $x < 1$, that $\tilde{e}_x$ is indeed Lipschitz continuous. This proves the main result. Lemma 1 also shows that $\tilde{e}_x(s)$ is strictly decreasing in s, so that for $s < e^* = \tilde{e}_x = \tilde{e}_x(e^*) < \tilde{e}_x(s)$. □

A.2 Proofs for Section 4

A.2.1 Convex Benefits

Write $U(x, e, s) := B_x(e) - (1 - x)\tilde{D}(e|s)$ where $B_x(e) := B(E_x + (1 - x)e)$. Thus $U(x, e, s)$ is the net utility perceived in period x at time s from exerting effort e in all remaining periods. Intuitively, it is net perceived utility on 'day' x after s 'hours' of work when planning to work e 'hours' on all remaining 'days'.

Lemma 2. *Let $e_H > e_L$ and $s_H > s_L$. Then $U(x, e_L, s_H) - U(x, e_H, s_H) > U(x, e_L, s_L) - U(x, e_H, s_L)$.*

Proof. We have:

$$U(x, e_L, s) - U(x, e_H, s) = B_x(e_L) - B_x(e_H) - (1 - \alpha)(D(e_L) - D(e_H)) - \alpha D'(s)(e_L - e_H)$$

Since $e_L - e_H < 0$, the expression is strictly increasing in s, proving the claim. □

Fixing some $\alpha < 1$, as well as some (x, E_x, s), by Assumption 1 we know that there is some E_F s.t. the person never intends to work beyond that. Hence we have that $e \leq \frac{E_F}{1-x} = K$, so that $\sup_e U(x, e, s) = \sup_{e \leq K} U(x, e, s)$. When B is continuous, then U is a continuous function in e, so we have $\sup_{e \leq K} U(x, e, s) = \max_{e \leq K} U(x, e, s)$ and $\arg\max_{e \leq k} U(x, e, s)$ is a closed set. When B is not continuous as in the all-or-nothing case, then the set is closed because the person is always only considering to not work at all ($e = 0$) or to work fully ($e = \frac{E_F - E_x}{1-x}$).

Therefore in all cases, we can define the following:

Definition 6. *Let $m_x(s) = \min \arg\max_e U(x, e, s)$ be the smallest effort level that maximizes perceived net utility in period x at time s.*

Lemma 3. *$m_x(s)$ is decreasing in s.*

Proof. Since $m_x(s)$ maximizes $U(x,e,s)$, we have that $U(x,m_x(s),s) \geq U(x,e,s)$ for all e. This implies that $U(x,m_x(s),s)-U(x,e,s) \geq 0$ for all e. For $s_H > s$ and all $e > m_x(s)$, Lemma 2 implies that $U(x,m_x(s),s_H) - U(x,e,s_H) > U(x,m_x(s),s) - U(x,e,s) \geq 0$. Therefore any e maximizing $U(x,e,s_H)$ must be smaller than $m_x(s)$. Since $m_x(s_H)$ does maximize $U(x,e,s_H)$, we have that $m_x(s_H) \leq m_x(s)$. This proves the claim. □

We can now simplify the decision to stop in terms of $m_x(s)$.

Lemma 4. $\tilde{e}_x = \inf\{s : m_x(s) \leq s\}$.

Proof. From Definition 2, we have

$$\begin{aligned}\tilde{e}_x &= \inf\{s : B_x(e) - (1-x)\cdot\tilde{D}(e|s) < \max_{\underline{e}\leq s} B_x(\underline{e}) - (1-x)\cdot\tilde{D}(\underline{e}|s),\ \forall e > s\} \\ &= \inf\{s : U(x,e,s) < \max_{\underline{e}\leq s} U(x,\underline{e},s)\forall e > s\}\end{aligned}$$

Let $A(x) = \{s : U(x,e,s) < \max_{\underline{e}\leq s} U(x,\underline{e},s)\forall e > s\}$. Then for every $s \in A(x)$, we have $U(x,e,s) < \max_{\underline{e}\leq s} U(x,\underline{e},s)$ for all $e > s$. Hence all effort levels that maximize $U(x,e,s)$ (for given x and s) must be less than or equal to s. So in particular $m_x(s) \leq s$, and hence $m_x(s) > s$ implies $s \notin A(x)$. Hence $A(x) \subset \{m_x(s) \leq s\}$, so that $\inf A(x) \geq \inf\{m_x(s) \leq s\}$.

Now consider $m_x(s) \leq s$. Since $m_x(s)$ maximizes $U(x,e,s)$, we have $\max_{\underline{e}\leq s} U(x,\underline{e},s) = U(x,m_x(s),s) \geq U(x,e,s)$ for all e. For every $s_H > s$ and every $e > s$, we also have that $e > m_x(s)$ since $s \geq m_x(s)$. Thus by Lemma 2 $U(x,m_x(s),s_H) - U(x,e,s_H) > U(x,m_x(s),s) - U(x,e,s) \geq 0$ for all $e > s$, which implies that any maximizer of $U(x,e,s_H)$ is smaller than s and hence strictly smaller than s_H. Hence $U(x,e,s_H) < U(x,m_x(s),s_H) \leq \max_{\underline{e}\leq s_H} U(x,\underline{e},s)$ for all $e > s_H$, since $m_x(s) \leq s < s_H$. Hence if $m_x(s) \leq s$, then $s_H \in A$ for every $s_H > s$. So $\inf\{m_x(s) \leq s\} \geq \inf A(x)$, which shows that $\inf A(x) = \inf\{m_x(s) \leq s\}$. This proves the claim. □

Lemma 5. *The limits as s converges to $\tilde{e}_x$ from below and from above do both exist: $m_x^- := \lim_{s\to\tilde{e}_x^-} m_x(s)$ and $m_x^+ := \lim_{s\to\tilde{e}_x^+} m_x(s)$. Moreover, $m_x^- \geq \tilde{e}_x$, $m_x^+ \leq \tilde{e}_x$, and $m_x(\tilde{e}_x) = m_x^+$.*

Notation: Denote $m_x(\tilde{e}_x)$ by m_x, the planned future effort at the time the person stops working. Then Lemma 5 shows that $m_x = m_x^+$.

Proof. By Lemma 4, we have that $\tilde{e}_x = \inf\{s : m_x(s) \leq s\}$, and we know that $m_x(s)$ is decreasing in s. Consider $s^- < \tilde{e}_x$, so that $s \notin \{s : m_x(s) \leq s\}$, and hence $m_x(s^-) > s^-$. I then prove that $\tilde{e}_x \leq m_x(s^-)$. Suppose not, so that $m_x(s^-) < \tilde{e}_x$. Then take $s' \in (m_x(s^-), \tilde{e}_x)$. Since $s' > m_x(s^-)$ and $m_x(s^-) > s^-$, we have $s' > s^-$. Since $m_x(s)$ is decreasing in s, we get $m_x(s') \leq m_x(s^-)$. But then $m_x(s^-) < s'$ implies $m_x(s') < s'$ which implies $s' \in \{s : m_x(s) \leq s\}$ which implies $\tilde{e}_x \leq s'$, which is a contradiction. Therefore we have that $\tilde{e}_x \leq m_x(s^-)$ for all $s^- < \tilde{e}_x$.

Since $m_x(s^-)$ is decreasing in s^- and bounded below, it is clear that the limit as s^- goes to $\tilde{e}_x$ exists and we denote it by m_x^-. Since $m_x(s^-) \geq \tilde{e}_x$ for all $s^- < \tilde{e}_x$, we must have that $m_x^- \geq \tilde{e}_x$.

Now consider $s^+ > \tilde{e}_x$. Since $\tilde{e}_x = \inf\{s : m_x(s) \leq s\}$, for every $s^+ > \tilde{e}$ there is some $s' \in (\tilde{e}_x, s^+)$ s.t. $s' \in \{s : m_x(s) \leq s\}$, i.e. s.t. $m_x(s') \leq s'$. Since $s' < s^+$ and since $m_x(s)$ is decreasing in s, we have that $m_x(s^+) \leq m_x(s') \leq s' < s^+$. Hence $m_x(s^+) < s^+$ for all $s^+ > \tilde{e}_x$.

Claim: $s^+ > \tilde{e}_x \implies m_x(s^+) \leq \tilde{e}_x$. Suppose not, then there is some s^+ s.t. $m_x(s^+) > \tilde{e}_x$. Hence picking $s' \in (\tilde{e}_x, m_x(s^+))$ implies that $m_x(s') \geq m_x(s^+)$ ($m_x(s)$ decreasing), so that $m_x(s') > \tilde{e}_x$. Since $s' > \tilde{e}_x$, we must also have that $m_x(s') < s'$, but since $s' < m_x(s^+)$ that implies $m_x(s') < m_s(s^+)$, which is a contradiction. Hence $m_x(s^+) \leq \tilde{e}_x$.

As $s \to \tilde{e}_x$ from above, $m_x(s)$ is strictly inreasing and bounded above by $\tilde{e}_x$, so it has a limit denoted by m_x^+. Since $m_x(s^+) \leq \tilde{e}_x$ for every $s^+ > \tilde{e}_x$, we have $m_x^+ \leq \tilde{e}_x$.

Finally, note that either $U(x, e, s)$ is continuous in s and in e or we are in the all-or-nothing case. First consider the continuous case. Let $m_x := m_x(\tilde{e}_x)$, the smallest maximizer at the time $s = \tilde{e}_x$. By definition of $m_x(s)$, we have that $U(x, m_x(s), s) \geq U(x, e, s)$ for all e. Hence by continuity of $U(x, e, s)$ in e and s, letting s converge to $\tilde{e}_x$ from above we have that $U(x, m_x^+, \tilde{e}_x) \geq U(x, e, \tilde{e}_x)$ and letting s converge to $\tilde{e}_x$ from below we get $U(x, m_x^-, \tilde{e}_x) \geq U(x, e, \tilde{e}_x)$. Hence both m_x^+ and m_x^- maximize $U(x, e, \tilde{e}_x)$ over e. But we know that $m_x^+ \leq \tilde{e}_x \leq m_x^-$, so the smallest maximizer is m_x^+, hence $m_x(\tilde{e}_x) = m_x^+$.

In the all-or-nothing case, $m_x(s) = 0$ or $m_x(s) = \frac{E_F - E_x}{1-x} = \bar{e}_x$. Thus either the two limits overlap or the limits above and below are different and equal to 0 and $\bar{e}_x$ respectively. In the first case, the result is clearly true with equality. In the second case, we must have that the limit from above is 0 and the limit from below is $\bar{e}_x$. Thus $U(x, \bar{e}_x, s) \geq U(x, 0, s)$ for all $s < \tilde{e}_x$ and $U(x, 0, s) \geq U(x, \bar{e}_x, s)$ for all $s > \tilde{e}_x$. Since $U(x, 0, s)$ and $U(x, \bar{e}_x, s)$ are both continuous in s, this implies that $U(x, \bar{e}_x, \tilde{e}_x) = U(x, 0, \tilde{e}_x)$, so that both m_x^+ and m_x^- maximize $U(x, e, \tilde{e}_x)$, hence $m_x = 0$ which is the smallest such maximizer. □

Corollary 2. *We have the following implications:*

- $s < \tilde{e}_x \implies m_x(s) \geq \tilde{e}_x$
- $s > \tilde{e}_x \implies m_x(s) \leq m_x \leq \tilde{e}_x$
- $s = \tilde{e}_x \implies m_x(s) = m_x \leq \tilde{e}_x$
- $m_x(s) > \tilde{e}_x \implies s < \tilde{e}_x$
- $m_x(s) < \tilde{e}_x \implies s \geq \tilde{e}_x.$
- $m_x(s) = s \implies s = \tilde{e}_x.$

Proof. These follow directly from Lemma 5 and the fact that $m_x(s)$ is decreasing in s. □

Lemma 6. *If $\tilde{e}_x \in \arg\max_e U(x, e, \tilde{e}_x)$ then $\tilde{e}_x = e_x^*$ where $e_x := \arg\max_e B_x(e) - (1-x)D(e)$ is the unique effort level maximizing true (not perceived) net utility.*

Proof. Suppose $\tilde{e}_x \in \arg\max_e U(x, e, \tilde{e}_x)$. Then $U(x, \tilde{e}_x, \tilde{e}_x) \geq U(x, e, \tilde{e}_x)$ for all e. So $B_x(\tilde{e}_x) - \tilde{D}(\tilde{e}_x|\tilde{e}_x) \geq B_x(e) - \tilde{D}(e|\tilde{e}_x)$, which implies the following:

$$\begin{aligned}
& B_x(\tilde{e}_x) - (1-\alpha)D(\tilde{e}_x) - \alpha D'(\tilde{e}_x)\tilde{e}_x \geq B_x(e) - (1-\alpha)D(e) - \alpha D'(\tilde{e}_x)e \\
\iff & B_x(\tilde{e}_x) - D(\tilde{e}_x) + \alpha(D(\tilde{e}_x) - D'(\tilde{e}_x)\tilde{e}_x) \geq B_x(e) - D(e) + \alpha(D(e) - D'(\tilde{e}_x)e) \\
\iff & B_x(\tilde{e}_x) - D(\tilde{e}_x) \geq B_x(e) - D(e) + \alpha(D(e) - D(\tilde{e}_x) - D'(\tilde{e}_x)(e - \tilde{e}_x))
\end{aligned}$$

When $\tilde{e}_x > e$, then $D(e) - D(\tilde{e}_x) = -\int_e^{\tilde{e}_x} D'(y)dy > -\int_e^{\tilde{e}_x} D'(\tilde{e}_x)dy = -(\tilde{e}_x - e)D'(\tilde{e}_x) = (e - \tilde{e}_x)D'(\tilde{e}_x)$. When $e > \tilde{e}_x$, then similarly $D(e) - D(\tilde{e}_x) > (e -$

$\tilde{e}_x)D'(\tilde{e}_x)$. Hence for all $e \neq \tilde{e}_x$, we have $D(e) - D(\tilde{e}_x) - (e - \tilde{e}_x)D'(\tilde{e}_x) > 0$. Substituting this in the equation above, we get

$$B_x(\tilde{e}_x) - D(\tilde{e}_x) \geq B_x(e) - D(e) + \alpha(D(e) - D(\tilde{e}_x) - D'(\tilde{e}_x)(e - \tilde{e}_x)) > B_x(e) - D(e)$$

which proves that $\tilde{e}_x$ is the unique maximizer of $B_x(e) - D(e)$. □

Lemma 6 and Corollary 2 lead to the following characterization:

Proposition 6. *In every period, either $m_x = \tilde{e}_x$ or $m_x < \tilde{e}_x$. When $m_x = \tilde{e}_x$, then $\tilde{e}_x = e^*_x$ is the unique effort level maximizing true net utility $B_x(e) - (1 - x)D(e)$.*

Proof. By Corollary 2, when $s = \tilde{e}_x$ we know that $m_x \leq \tilde{e}_x$. Hence either $m(x) = \tilde{e}_x$ or $m_x < \tilde{e}_x$. Since $m_x \in \arg\max_e U(x, e, \tilde{e}_x)$, when $m_x = \tilde{e}_x$, we have $\tilde{e}_x \in \arg\max_e U(x, e, \tilde{e}_x$. Hence Lemma 6 implies that $\tilde{e}_x = e^*_x$ is the unique effort level that maximizes true net utility $B_x(e) - (1 - x)D(e)$, which completes the proof. □

The next Lemma makes use of Assumption 1. Although it is not needed – except when $\alpha = 1$ – the assumption simplifies the proof.[21]

Lemma 7. *Consider $x_0 < 1$. Then E_x is Lipschitz continuous in x for $x \in [0, x_0]$.*

Proof. By assumption 1, there are no benefits from exerting more total effort than E_F. Thus, at any time s in any period x, the person plans to work at most $E_F - E_x$ over the remaining periods. Since the person always plans to exert the same amount of effort in all remaining periods, she plans to work at most $\frac{E_F - E_x}{1-x} \leq \frac{E_F - E_x}{1-x_0} \leq \frac{E_F}{1-x_0}$. We know that the actual effort $\tilde{e}_x \leq m_x(0)$, and we just showed that $m_x(0) \leq K$ where $K = \frac{E_F}{1-x_0}$. Since $\dot{E}_x = \tilde{e}_x$, this shows that E_x is Lipschitz continuous in x with Lipschitz constant K. While K depends on x_0, for any $x < 1$ we can find $x_0 \in (x, 1)$ so that E_x is Lipschitz continuous on $[0, x_0]$ and hence at x. □

[21] Here is a sketch for proving continuity of E_x directly. We know that $m_x(0) \geq m_x^- \geq \tilde{e}_x$. Now remember that the justification for the continuous-time setup is a discrete setup with T periods as $T \to \infty$ and the length of each period $\Delta = 1/T \to 0$. We still have that $\tilde{e}_t \leq m_t(0)$, which means that the person works less each period than planned. This means that returns from working are lower in period $t + 1$ than the person thought in period t: she worked less than she planned, returns are increasing ($B''(e) > 0$), hence the returns are lower. Therefore $m_{t+1}(0) \leq m_t(0)$, so that $\tilde{e}_{t+1} \leq m_t(0)$. By induction, this yields that $\tilde{e}_t \leq m_0(0)$, so that $E_t \leq t \cdot m_0(0)$.

Lemma 8. *$\tilde{e}_x$ is continuous in x for all $x \in [0,1)$.*

Proof. Let $\varepsilon_0 > 0$, and $s_1 = \tilde{e}_x - \varepsilon_0 < \tilde{e}_x$. Since $m_x(s_1)$ is the smallest maximizer of $U(x, e, s_1)$ (over e), we have $U(x, m_x(s_1), s_1) > U(x, e, s_1)$ for all $e < m_x(s_1)$. Therefore $\max_{e \leq m_x(s_1) - \varepsilon_1} U(x, e, s_1) < U(x, m_x(s_1), s_1)$ for every $\varepsilon_1 > 0$. Pick $\varepsilon_1 < \varepsilon_0 = \tilde{e}_x - s_1 \leq m_x(s_1) - s_1$, where the last inequality holds by Corollary 2 since $s_1 < \tilde{e}_x$.

Let $\Delta^-(x, \bar{e}) := U(x, \bar{e}, s_1) - \max_{e \leq \bar{e} - \varepsilon_1} U(x, e, s_1)$, so that $\Delta^-(x, m_x(s_1)) = U(x, m_x(s_1), s_1) - \max_{e \leq m_x(s_1) - \varepsilon_1} U(x, e, s_1) > 0$. Since $U(x, e, s_1)$ is continuous in x, hence $\max_{e \leq m_x(s_1) - \varepsilon_1} U(x, e, s_1)$ is continuous in x, so that $\Delta^-(x, \bar{e})$ is continuous in x. We can therefore choose $\delta_1 > 0$ s.t. $|y - x| < \delta_1$ implies that $\Delta^-(x, m_x(s_1)) - \Delta^-(y, m_x(s_1)) < 1/2\Delta^-(x, m_x(s_1))$, which implies $1/2\Delta^-(x, m_x(s_1)) < \Delta^-(y, m_x(s_1))$ so that $0 < \Delta^-(y, m_x(s_1))$. By definition of Δ^-, this means that $U(y, m_x(s_1), s_1) > \max_{e \leq m_x(s_1) - \varepsilon_1} U(y, e, s_1)$, which implies that any maximizer of $U(y, e, s_1)$, such as $m_y(s_1)$, has to be greater or equal to $m_x(s_1) - \varepsilon_1$. Since $\varepsilon_1 < m_x(s_1) - s_1$, we have $m_x(s_1) - \varepsilon_1 > s_1$, so that $m_y(s_1) \geq m_x(s_1) - \varepsilon_1 > s_1$, so that $m_y(s) > s$ for all $s < s_1$. Therefore $\tilde{e}_y \geq s_1$.

Now consider $s_2 = \tilde{e}_x + \varepsilon_0 > \tilde{e}_x$. We have that $U(x, m_x^+, \tilde{e}_x) \geq U(x, e, \tilde{e}_x)$ for all e, hence it holds for all $e > m_x^+$. Since $s_2 > \tilde{e}_x$, when $e > m_x^+$ we can apply Lemma 2 to obtain that $U(x, m_x^+, s_2) - U(x, e, s_2) > U(x, m_x^+, \tilde{e}_x) - U(x, e, \tilde{e}_x)$. Since the RHS is greater than 0, the LHS is strictly greater than 0: $U(x, m_x^+, s_2) - U(x, e, s_2) > 0$ for all $e > m_x^+$. Hence $\Delta^+(x, m_x(s_2)) > 0$ for all ε_2, where $\Delta^+(x, \bar{e}) := U(x, \bar{e}, s_2) - \max_{e \geq \bar{e} + \varepsilon_2} U(x, e, s_2)$. Since $s_2 > \tilde{e}_x$, we have $m_x(s_2) \leq \tilde{e}_x < s_2$, so we can pick ε_2 with $0 < \varepsilon_2 < s_2 - m_x(s_2)$. By continuity of Δ^+ in x (as in the previous paragraph), we can find $\delta_2 > 0$ s.t. $|y - x| < \delta_2$ implies that $\Delta^+(y, m_x(s_2)) - \Delta^+(x, m_x(s_2)) > -1/2\Delta^+(x, m_x(s_2)) \implies \Delta^+(y, m_x(s_2)) > 0 \implies U(y, m_x(s_2), s_2) - \max_{e \geq m_x(s_2) + \varepsilon_2} U(y, e, s_2) > 0$. This shows that all maximizers of $U(y, e, s_2)$ are strictly less than $m_x(s_2) + \varepsilon_2$, which by our choice of ε_2 is strictly less than s_2. Hence $m_y(s_2) < s_2$, so that by definition of $\tilde{e}_y$, we have $\tilde{e}_y \leq s_2$.

In summary, for any $\varepsilon > 0$, let $\varepsilon_0 = 1/2\varepsilon$. Then we can pick δ_1 and δ_2 both strictly positive s.t. when $|y - x| < \min\{\delta_1, \delta_2\}$, then $\tilde{e}_y \in [\tilde{e}_x - \varepsilon_0, \tilde{e}_x + \varepsilon_0] \in (\tilde{e}_x - \varepsilon, \tilde{e}_x + \varepsilon)$, so that $\tilde{e}_x$ is continuous in x. This completes the proof. □

Proposition 7. *Suppose* $\tilde{e}_{x_0} \in \arg\max_e U(0, e, \tilde{e}_{x_0})$ *for some* $x_0 < 1$. *If* $\tilde{e}_x$ *is Lipschitz continuous in* x*, then* $\tilde{e}_x = \tilde{e}_{x_0}$ *for all* $x \geq x_0$ *and* $\tilde{e}_x = m_x$ *for all* $x > x_0$*.*

Later on, we prove that $\tilde{e}_x$ is Lipschitz continuous in x for all-or-nothing tasks.

Proof. Let us prove that $\tilde{e}_x = \tilde{e}_{x_0}$ for all $x \geq x_0$ solves the IVP. Then, assuming that $\tilde{e}_x$ is Lipschitz continuous in x, we know that there is a unique solution to the IVP by the Picard-Lindelöf Theorem. So the solution we found is the only solution, which proves the proposition.

To declutter the notation, let us write c for $\tilde{e}_{x_0}$.

We will now show that $\tilde{e}_x = c$ for all $x > x_0$ solves the IVP. When $\tilde{e}_x = c$ for all $x \geq x_0$, then $E_x = E_{x_0} + (x - x_0)c$. Let us now show that when $s = c$, the person perceives effort c as strictly better than some other effort $\bar{e}$ considered. Let e' be s.t. $e'(1 - x_0) = (x - x_0)c + (1 - x)\bar{e}$. Then working at effort level e' from period x_0 onwards leads to the same total effort as working c for all periods in $[x_0, x]$, followed by working $\bar{e}$ from x onwards. Hence it achieves the same total benefit, but at a lower disutility since it smoothes out the effort:

$$(x - x_0)\tilde{D}(c|c) + (1 - x)\tilde{D}(\bar{e}|c) > (1 - x_0)\tilde{D}(e'|c) \tag{2}$$

Since $\tilde{e}_{x_0} = c \in \arg\max_e U(0, e, \tilde{e}_{x_0})$, we have $B(E_{x_0} + (1 - x_0)c) - (1 - x_0)\tilde{D}(c|c) \geq B(E_{x_0} + (1 - x_0)e) - (1 - x_0)\tilde{D}(e|c)$ for every e, so it holds in particular for e'. Hence $B(E_{x_0} + (1 - x_0)c) - (1 - x_0)\tilde{D}(c|c) \geq B(E_{x_0} + (1 - x_0)e') - (1 - x)\tilde{D}(e'|c) > B(E_{x_0} + (1 - x_0)e') - (x - x_0)\tilde{D}(c|c) - (1 - x)\tilde{D}(e|c)$, where the last inequality holds by equation 2. By the definition of e', we have that $E_{x_0} + (1 - x_0)e' = E_{x_0} + (x - x_0)c + (1 - x)e = E_x + (1 - x)e$. Similarly, $E_{x_0} + (1 - x_0)c = E_x + (1 - x)c$. Substituting both of these into the equation after subtracting $(x - x_0)\tilde{D}(c|c)$ on both sides, we get:

$$B(E_x + (1 - x)c) - (1 - x)\tilde{D}(c|c) > B(E_x + (1 - x)e) - (1 - x)\tilde{D}(e|c)$$

This shows that $m_x(c) = c$. Therefore by Corollary 2, we must have $\tilde{e}_x = c$. This proves that $\tilde{e}_x = c$ is a solution to the IVP, and under Lipschitz continuity of $\tilde{e}_x$ it is the unique solution. This proves the proposition. □

Lemma 9. *Let* $x_q := \sup(\{0\} \cup \{x : \tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)\})$. *Then* $x \geq x_q \iff \tilde{e}_x \in \arg\max_e U(x, e, \tilde{e}_x)$.

Proof. For any $x_H > x_q$, there is some $x_L \in [x_q, x_H)$ s.t. $\tilde{e}_{x_L} \in \arg\max_e U(x_L, e, \tilde{e}_{x_L})$. Hence by Proposition 7, for all $x > x_L$ we have $\tilde{e}_x = m_x$, so in particular $\tilde{e}_{x_H} = m_{x_H} \in \arg\max_e U_\alpha(x_H, e, \tilde{e}_{x_H})$.

If $x_q = 0$, then this proves the claim, so assume that $x_q > 0$. Then for any $x_L < x_q$, we must have that $\tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)$): otherwise Proposition 7 would imply that $\tilde{e}_x = m_x \in \arg\max_e U_\alpha(x, e, \tilde{e}_x)$ for all $x > x_L$, so that $x_q \leq x_L < x_q$, a contradiction. □

The following proves Proposition 2.

Proof. Let us write $U_\alpha(x, e, s) := B(E_x + (1-x)e) - (1-x)\left((1-\alpha)D(e) - \alpha D'(s)e\right)$, i.e. the perceived net utility for a person with projection bias α. Then $U_\alpha(0, e, s) = B(e) - (1-\alpha)D(e) - \alpha D'(s)e$, since $E_0 = 0$, and $U_0(0, e, s)$ is the true net utility, as the person has no projection bias ($\alpha = 0$). Let $e^* = \min \arg\max_e U_0(0, e, s)$ be the smallest optimal effort, which is independent of s, so e^* maximizes in particular $U(0, e, e^*)$. Define $\bar{\alpha} := \sup A$ where $A := \{\alpha \in [0,1] : U_\alpha(0, e^*, e^*) \geq U_\alpha(0, e, e^*) \forall e\}$. Note that $0 \in A$, since e^* maximizes $U_0(0, e, e^*)$, so $\bar{\alpha} \in [0, 1]$.

Case 1: $\alpha \leq \bar{\alpha}$

Claim: $\alpha_L < \bar{\alpha} \implies U_{\alpha_L}(0, e^*, e^*) > U_{\alpha_L}(0, e, e^*)$ for all $e \neq e^*$.

Proof of claim: When $\bar{\alpha} = 0$, the claim is true since there is no $\alpha_L < \bar{\alpha}$. So consider $\bar{\alpha} > 0$ and $\alpha_L < \bar{\alpha}$. Let us compute $U_\alpha(0, e^*, e^*) - U_\alpha(0, e, e^*)$:

$$
\begin{aligned}
&U_\alpha(0, e^*, e^*) - U_\alpha(0, e, e^*) \\
=&B(e^*) - (1-\alpha)D(e^*) - \alpha D'(e^*)e^* - (B(e) - (1-\alpha)D(e) - \alpha D'(e^*)e) \\
=&B(e^*) - D(e^*) - (B(e) - D(e)) - \alpha\left(D(e) - D(e^*) - D'(e^*)(e - e^*)\right)
\end{aligned}
$$

Note that $D(e) - D(e^*) - D'(e^*)(e - e^*) = \int_{e^*}^{e} D'(y) - D'(e^*)dy = (e - e^*)(D'(\bar{y}) - D'(e^*))$ for some $\bar{y}$ that lies between e and e^*. Independent of whether $e < e^*$ or $e^* <$

e, this expression is strictly positive for all $e \neq e^*$. This shows that $U_\alpha(0, e^*, e^*) - U_\alpha(0, e, e^*)$ is strictly decreasing in α for $e \neq e^*$.

By definition of $\bar{\alpha}$, there is some $\alpha_H \in (\alpha_L, \bar{\alpha}]$ with $\alpha_H \in A$: $U_{\alpha_H}(0, e^*, e^*) \geq U(0, e, e^*)$ for all e, i.e. $U_{\alpha_H}(0, e^*, e^*) - U_{\alpha_H}(0, e, e^*) \geq 0$ for all e. Since $\alpha_L < \alpha_H$, we therefore have that $U_{\alpha_L}(0, e^*, e^*) - U_{\alpha_L}(0, e, e^*) > U_{\alpha_H}(0, e^*, e^*) - U_{\alpha_H}(0, e, e^*) \geq 0$ for all $e \neq e^*$. This proves the claim.

Since U_α is continuous in α (total effort is bounded by assumption), it is clear that $U_{\bar{\alpha}}(0, e^*, e^*) \geq U_{\bar{\alpha}}(0, e, e^*)$ for all $e \neq e^*$.

Therefore for $\alpha_L \leq \bar{\alpha}$, we know that e^* is a maximizer of $U_{\alpha_L}(0, e, e^*)$. So by Proposition 7, $\tilde{e}_x = e^* \in \arg\max_e U_\alpha(0, e, e^*) = \arg\max_e U_\alpha(0, e, \tilde{e}_x)$. For $x_q := \sup\left(\{0\} \cup \{x : \tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)\}\right)$, we have that $x_q = 0$. Since $\tilde{e}_x = e^* = e^*_0 = e^*_{x_q}$, this proves the result for this case.

Case 2: $\alpha > \bar{\alpha}$ Now consider $\alpha_H > \bar{\alpha}$. Then by definition of $\bar{\alpha}$ as $\sup A$, this implies that $\alpha_H \notin A$. Hence there is some $\bar{e}$ s.t. $U_{\alpha_H}(0, \bar{e}, e^*) > U_{\alpha_H}(0, e^*, e^*)$. Thus by Lemma 6 we know that $\tilde{e}_0 \notin \arg\max_e U_{\alpha_H} U(0, e, \tilde{e}_0)$: if it was, then $\tilde{e}_0$ would have to be the unique maximizer e^*, hence it would have to maximize $U_{\alpha_H}(0, e, \tilde{e}_0) = U_{\alpha_H}(0, e, e^*)$. But we know that $U_{\alpha_H}(0, e^*, e^*) < U_{\alpha_H}(0, \bar{e}, e^*)$, so that can't be the case. Therefore $\tilde{e}_0 \notin \arg\max_e U_{\alpha_H}(0, e, \tilde{e}_0)$, so that $U_{\alpha_H}(0, m_0, \tilde{e}_0) > U_{\alpha_H}(0, \tilde{e}_0, \tilde{e}_0)$, which shows that $m_0 \neq \tilde{e}_0$. Hence, since $m_x \leq \tilde{e}_x$ for all x, we must have that $m_0 < \tilde{e}_0$.

By continuity of $\tilde{e}_x$ in x and of U in its arguments, there is some $\delta > 0$ s.t. for all $x < \delta$ we have $U_{\alpha_H}(x, m_0, \tilde{e}_x) > U_{\alpha_H}(x, \tilde{e}_x, \tilde{e}_x)$ and s.t. $\tilde{e}_x > m_0$, since this holds for $x = 0$. This shows that $\tilde{e}_x$ does not maximize $U_{\alpha_H}(x, e, \tilde{e}_x)$, so that $m_x \neq \tilde{e}_x$. Since $m_x \leq \tilde{e}_x$ for all x, this implies that $m_x < \tilde{e}_x$.

Hence $\sup\{x : \tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)\} \geq \delta > 0$, so $x_q > 0$. Lemma 9 establishes the remaining properties for x_q, which proves this case. □

A.2.2 All-or-Nothing Tasks

First, note that all-or-nothing tasks clearly have bounded effort, so it satisfies Assumption 1. To explore the impact of raising the benefit B_F received for completing the task, we parameterize the benefit as $\mu \cdot B_F$ for $\mu \geq 0$. We write $\bar{e}_x = \frac{E_F - E_x}{1-x}$ for the daily effort needed from period x onwards to complete the task given how much work is left to do. Hence we write $U(x, e, s, \mu)$ for $\mu B_F \mathbb{1}(e \geq \bar{e}_x) - (1-x)\tilde{D}(e|s)$.

Note that $m_x(s) := \min \arg\max_e U(x, e, s, \mu)$ is well-defined, since there are only two effort levels (for any given x) that can possibly maximize utility: 0 or $\bar{e}_x$.

Proposition 8. *Consider an all-or-nothing task, with benefit μB_F for the effort E_F. Let $\bar{e}_x = \frac{E_F - E_x}{1-x}$. Then the IVP is*

$$\dot{E}_x = \tilde{e}_x$$

$$\tilde{e}_x = e(x, E_x, \mu) = \begin{cases} 0, \textit{ if } \mu B_F - (1-x)\tilde{D}(\bar{e}_x|0) \leq 0 \\ \bar{e}_x, \textit{ if } \mu B_F - (1-x)\tilde{D}(\bar{e}_x|\bar{e}_x) > 0 \\ s_q, \textit{ if } \mu B_F - (1-x)\tilde{D}(\bar{e}_x|0) > 0 \geq \mu B_F - (1-x)\tilde{D}(\bar{e}_x|\bar{e}_x) \end{cases}$$

with initial condition $E_0 = 0$, $B_F = B(E_F)$ and s_q satisfies $\mu B_F - \tilde{D}(\bar{e}_x|s_q) = 0$.

Proof. By Lemma 4, we know that $\tilde{e}_x = \inf\{s : m_x \leq s\}$. For all-or-nothing tasks, the person either plans to do no further work or to complete the task, since all other levels require strictly more work for exactly the same gain. Hence $m_x \in \{0, \bar{e}_x\}$.

Suppose that $\mu B_F - (1-x)\tilde{D}(\bar{e}_x|0) \leq 0$. Then 0 effort is a maximizer of the perceived utility at $s = 0$, hence it is the smallest maximizer. Therefore $m_x(0) = 0$, so that by Corollary 2, $\tilde{e}_x = 0$.

Suppose that $\mu B_F - (1-x)\tilde{D}(\bar{e}_x|\bar{e}_x) > 0$. Then $\bar{e}_x$ is the only maximizer at $s = \bar{e}_x$, i.e. $m_x(\bar{e}_x) = \bar{e}_x$. Thus by Corollary 2, $\tilde{e}_x = \bar{e}_x$.

Finally suppose that $\mu B_F - (1-x)\tilde{D}(\bar{e}_x|0) > 0 \geq \mu B_F - (1-x)\tilde{D}(\bar{e}_x|\bar{e}_x)$. Let s_q satisfy $\mu B_F - \tilde{D}(\bar{e}_x|s_q) = 0$. Since $\mu B_F - \tilde{D}(\bar{e}_x|s)$ is strictly decreasing in s, this s_q satisfies $s_q > 0$ and $s_q \leq \bar{e}_q$. Since it is strictly decreasing, for all $s < s_q$ we also have $\mu B_F - (1-x)\tilde{D}(\bar{e}_x|s) > \mu B_F - (1-x)\tilde{D}(\bar{e}_x|s_q) = 0$, hence $m_x(s) = \bar{e}_x \geq s_q > s$, so that $\tilde{e}_x \geq s_q$. Since at s_q we have that both 0 and $\bar{e}_x$ are maximizers, $m_x(s_q) = 0$ is the smallest maximizer, so that $\tilde{e}_x = s_q$. □

I will use the following theorem (from https://www.math.washington.edu/~burke/crs/555/555_notes/continuity.pdf):

Theorem 1. *Consider the initial value problem*

$$x' = f(t, x, \mu), \; x(t_0) = y$$

where x' is the derivative of $x(t)$ with respect to time. If f is continuous in t, x, μ and Lipschitz in x with Lipschitz constant independent of t and μ, then $x(t, \mu, y)$ is continuous in (t, μ, y) jointly.

Lemma 10. *Consider $e(x, E, \mu)$ from Proposition 8. Suppose that $D''(x) > d$ for some $d > 0$. Then $e = e(x, E, \mu)$ restricted to $x \in [0, 1-\varepsilon]$ with $\varepsilon > 0$ exists and is Lipschitz continuous in x, E, and μ, on $[0, 1-\varepsilon] \times [0, \bar{E}] \times [0, \bar{u}]$ for any finite $\bar{E} > 0$ and $\bar{u} > 0$.*

Proof. It is clear that $\tilde{e}_x$ is Lipschitz continuous for all x, μ satisfying $\mu B_F - (1 - x)\tilde{D}(\bar{e}_x|0) \leq 0$ since it is constant and equal to 0 in this region by 8. The same is true for all x, μ satisfying $\mu B_F - (1-x)\tilde{D}(\bar{e}_x|\bar{e}_x) > 0$, since it is equal to $\bar{e}_x = \frac{E_F - E_x}{1-x}$ in this region, which is Lipschitz continuous for $x < 1 - \varepsilon$. Finally, we have $\tilde{e}_x = s_q$ for all other x and μ, with s_q satisfying $\mu B_F - \tilde{D}(\bar{e}_x|s_q) = 0$. This implies:

$$\mu B_F - (1-\alpha)D(\bar{e}_x) = \alpha D'(s_q)\bar{e}_x \iff s_q = (D')^{-1}\left(\frac{\mu B_F - (1-\alpha)D(\bar{e}_x)}{\alpha \bar{e}_x}\right)$$

The argument on the RHS is Lipschitz in μ and $\bar{e}_x$, and $\bar{e}_x$ is Lipschitz continuous in E_x (since $x \leq x_0 < 1$). The inverse function of D' is Lipschitz continuous if $D'' \geq d$ for some $d > 0$. Hence $\tilde{e}_x$ is Lipschitz also in this domain.

It is staightforward to show that where the closures of the three regions overlap, $\tilde{e}_x$ is continuous, hence it is Lipschitz continuous on the whole domain. □

This shows that $\tilde{e}_x$ is Lipschitz continuous in E_x, so that Proposition 7 applies. Lemma 10 means we can apply Theorem 1 to obtain that $E(x, \mu)$ is continuous in x and μ jointly.

Next, we prove that there are thresholds below which the person never works and above which the person works efficiently on the task.

Lemma 11. *Let μ_L be s.t. $\mu_L B_F - \tilde{D}(\bar{e}_0|0) = 0$ and μ_H s.t. $\mu_H B_F - \tilde{D}(\bar{e}_0|\bar{e}_0) = 0$. Then $\mu \leq \mu_L$ implies that $\tilde{e}_x = 0$ for all $x < 1$, and $\mu \geq \mu_H$ implies that $\tilde{e}_x = \bar{e}_0 = E_F$ for all $x < 1$.*

Proof. When $\mu \leq \mu_L$, we have $\mu B_F - \tilde{D}(\bar{e}_0|0) \leq 0$, so that $m_0 = 0$, since this maximizes the perceived utility at $s = 0$. Therefore $\tilde{e}_0 = 0$ maximizes perceived utility at $s = 0$, so that Proposition 7 implies that $\tilde{e}_x = \tilde{e}_0 = 0$ for all $x < 1$.

When $\mu > \mu_H$, we have that $\mu B_F - \tilde{D}(\bar{e}_0|\bar{e}_0) > 0$. Thus by Proposition 8 we know that $\tilde{e}_0 = \bar{e}_0$ which thus maximizes perceived utility at $s = \bar{e}_0$. So by Proposition 7, $\tilde{e}_x = \tilde{e}_0 = \bar{e}_0 = E_F$ for all $x < 1$.

Finally, when $\mu = \mu_H$, we have $\mu B_F - \tilde{D}(\bar{e}_0|\bar{e}_0) = 0$. Hence by Proposition 8, we know that $\tilde{e}_0 = s_q$ with s_q satisfying $\mu B_F - \tilde{D}(\bar{e}_0|s_q) = 0$. Since this is solved by $\bar{e}_0$, we have $s_q = \bar{e}_0$, so that $\tilde{e}_0 = \bar{e}_0$. This maximizes the perceived utility at $\bar{e}_0 = s_q$, thus Proposition 7 again implies that $\tilde{e}_x = \tilde{e}_0$.

This completes the proof. □

Lemma 12. *Suppose that $\mu > \mu'$, with $x_q(\mu') > 0$. Then $E_x(\mu) > E_x(\mu')$ for all $x \in (0, 1)$. Letting $q = x_q(\mu)$ and $q' = x_q(\mu')$, we have that either $q \leq q'$ and $\tilde{e}_q(\mu) = \bar{e}_q(\mu)$ (the μ-type completes the task efficiently) or $q' < q$ and $\tilde{e}_q(\mu') = 0$ (the μ'-type never works at all after period q'). Moreover, when $q = q'$, then $\tilde{e}_q(\mu') = 0$.*

Proof. I will refer to the agent facing benefits μB as the μ-type. As a reminder, $x_q := \sup\{x : \tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)\}$.

First, notice that independent of μ, when $x \leq x_q$, we know from Proposition 2 that $m_x < \tilde{e}_x$, and in general we have $m_x^- \geq \tilde{e}_x$ (Lemma 5). Hence we have $m_x < m_x^-$ and at $s = \tilde{e}_x$, we have $U(x, m_x, \tilde{e}_x) = U(x, m_x^-, \tilde{e}_x)$ so that there are at least two different maximizers at $s = \tilde{e}_x$. For all-or-nothing tasks, these must be 0 and $\bar{e}_x$, hence $m_x = 0$ and $m_x^- = \bar{e}_x = \frac{E_F - E_x}{1-x}$. Thus for $x \leq x_q$, we have $U(x, 0, \tilde{e}_x) = U(x, \bar{e}_x, \tilde{e}_x) \iff 0 = B_F - (1 - x)\tilde{D}(\bar{e}_x|\tilde{e}_x)$.

Let $q = x_q(\mu)$ and $q' = x_q(\mu')$. Then for all $x \leq \min\{q, q'\}$, we have that

$$\mu B_F - (1 - x)\tilde{D}(\bar{e}_x(\mu)|\tilde{e}_x(\mu)) = 0 = \mu' B_F - (1 - x)\tilde{D}(\bar{e}_x(\mu')|\tilde{e}_x(\mu')) \qquad (3)$$

At $x = 0$, we have $E_0(\mu) = E_0(\mu') = 0 \iff \bar{e}_0(\mu) = \bar{e}_0(\mu')$. Since $\mu' < \mu$, it is clear that $\tilde{e}_0(\mu') < \tilde{e}_0(\mu)$.

Claim: $\tilde{e}_x(\mu') < \tilde{e}_x(\mu)$ for all $x \leq \min\{q, q'\}$. Suppose not, so that there is $x \leq \min\{q, q'\}$ with $\tilde{e}_x(\mu') \geq \tilde{e}_x(\mu)$. Then, since $\tilde{e}_0(\mu') < \tilde{e}_0(\mu)$ and since $\tilde{e}_x(\mu)$ is continuous in x, there is some minimum x_0 s.t. $\tilde{e}_{x_0}(\mu') = \tilde{e}_{x_0}(\mu)$ with $x > 0$. So for all $x < x_0$, we have $\tilde{e}_x(\mu') < \tilde{e}_x(\mu)$, hence $E_{x_0}(\mu') < E_{x_0}(\mu)$. Then equation 3 implies $\tilde{e}_{x_0}(\mu') < \tilde{e}_{x_0}(\mu)$, a contradiction. This proves the claim.

By Proposition 2, we know that at x_q the person starts working efficiently, which in the case of all-or-nothing tasks implies that the person either stops working entirely, or works efficiently on completing the task by working $\bar{e}_x$.

Case 1: Suppose $q \leq q'$. We have that $\tilde{e}_x(\mu) = \tilde{e}_q(\mu)$ for all $x \geq q$, and $\tilde{e}_q(\mu)$ is either 0 or $\bar{e}_q(\mu)$. If $\tilde{e}_q(\mu) = 0$, then the claim above implies that $\tilde{e}_q(\mu') < \tilde{e}_q(\mu) = 0$, which is not possible. Thus we must have $\tilde{e}_q(\mu) = \bar{e}_q(\mu)$ and the person works efficiently towards completing the task.

We can now show that $E_x(\mu) > E_x(\mu')$ for all $x < 1$, not just all $x \leq q$. Consider $x = q$. From this period on, the μ-type is working efficiently. The μ'-type is behind this person (meaning $E_{x_q}(\mu') < E_{x_q}(\mu)$), but always plans to work efficiently, even though they may end up quitting before having worked as planned. If they follow their plan, they catch up with the μ-type in period $x = 1$, so they never plan on catching up before time $x = 1$. Yet, they end up working at most as much as planned, and initially less than planned, so they catch up less or at most as much as planned. So they never catch up before $x = 1$.

Finally note that we cannot have $q = q'$ with the μ'-type working efficiently from q' on, since then $q = q' > 0$ (by the Lemma's assumption). That implies $E_q(\mu) > E_q(\mu')$ which implies $\bar{e}_q(\mu) < \bar{e}_q(\mu')$, contradicting our claim above. Thus when $q = q'$, then the μ'-type never works from $x = q$ on.

This proves the result when $q \leq q'$.

Case 2: Suppose that $q > q'$. Since $q' > 0$, we know that $E_{q'}(\mu') < E_{q'}(\mu)$. Suppose that $\tilde{e}_{q'}(\mu') = \bar{e}_{q'}(\mu')$, meaning that the person works towards completing the task efficiently. The above claim shows that $\tilde{e}_{q'}(\mu) > \tilde{e}_q(\mu')$ so that $\tilde{e}_{q'}(\mu) > \tilde{e}_{q'}(\mu') = \bar{e}_{q'}(\mu') = \frac{E_F - E_q(\mu')}{1-q} > \frac{E_F - E_q(\mu)}{1-q} = \bar{e}_q(\mu)$. This would imply that the μ-type works more than is efficient, which cannot happen. Hence $\tilde{e}_{q'}(\mu') = 0$: the μ'-type stops working entirely. That implies that $E_x(\mu') = E_{q'}(\mu')$ for all $x \geq q'$. Thus $E_x(\mu) \geq E_{q'}(\mu) > E_{q'}(\mu') = E_x(\mu')$ for all $x \geq q'$, which proves the claim when $q > q'$.

Toghether, these prove that $E_x(\mu) > E_x(\mu')$ for all $x \in (0, 1)$. □

Lemma 13. *Let D be convex with $D'(e) \to \infty$ as $e \to \infty$. Then $\forall K > 0$, $\exists E$ s.t. $D(e) > K \cdot e \; \forall e > E$. That is, $D(e)/e \to \infty$ as $e \to \infty$.*

Proof. Since $D'(e) \to \infty$, pick E s.t. $D'(\frac{E}{2}) > 2 \cdot K$. Then for $e > E$

$$D(e) = \int_0^e D'(s)ds \geq \int_{E/2}^e D'(s)ds \geq \int_{E/2}^E 2 \cdot K ds \geq \frac{e}{2} 2 \cdot K = e \cdot K$$

□

Lemma 14. *Consider the bounded effort case, then $\mu_C := \inf\{\mu : E_1(\mu) = 1\}$ satisfies $\mu_C \in (\mu_L, \mu_H)$, where $E_1(\mu) = \lim_{x \to 1} E_x(\mu)$.*

Proof. Pick s_M so large that $\alpha D'(s_M)\frac{1}{4}E_F > B_F$. Then pick $\bar{x}$ so close to 1 that $(1 - \bar{x}) \cdot s_M < \frac{1}{4}E_F$ with $\bar{x} > \frac{3}{4}$. We have $E_{\bar{x}}(\mu_L) = 0$ and that $E_{\bar{x}}(\mu_H) = \bar{x} \cdot E_F > \frac{1}{2}E_F$ since $\bar{x} > \frac{3}{4} > \frac{1}{2}$. By continuity of $E_x(\mu)$ in μ for all $x < 1$, there is some μ' s.t. $E_{\bar{x}}(\mu') = \frac{1}{2}E_F$.

Claim: If $E_F - E_x > \frac{1}{4}E_F$, then $\tilde{e}_x < s_M$. $E_F - E_x > \frac{1}{4}E_F$ implies that $(1 - x)\tilde{D}(\bar{e}_x|s_M) = (1 - x)(1 - \alpha)D(\bar{e}_x) + (1 - x)\alpha D'(s_M)\bar{e}_x \geq \alpha D'(s_M)\frac{(E_F - E_x)}{1-x} \geq \alpha D'(s_M)(E_F - E_x) > \alpha D'(s_M)\frac{1}{4}E_F > B_F$. Thus $B_F - (1 - x)\tilde{D}(\bar{e}_x|s_M) < 0$ and the person stops working before reaching s_M, proving the claim.

Claim: for all $x > \bar{x}$, $\tilde{e}_x < s_M$. Note that we have that $E_F - E_{\bar{x}} = 1/2E_F > 1/4E_F$, so the previous claim shows that $\tilde{e}_{\bar{x}} < s_M$. Suppose, by contradiction, that the claim does not hold. Thus, since $\tilde{e}_{\bar{x}} < s_M$ and by continuity of $\tilde{e}_x$, there is minimum x' s.t. $\tilde{e}_{x'} = s_M$. For $x'' \in (\bar{x}, x')$, we have $\tilde{e}_{x''} < s_M$, hence $E_{x'} < E_{\bar{x}} + s_M(x' - \bar{x}) < E_{\bar{x}} + s_M(1 - \bar{x}) < E_{\bar{x}} + \frac{1}{4}E_F < \frac{3}{4}E_F$. Thus $E_F - E_{x'} > \frac{1}{4}E_F$ and the previous claim then implies that $\tilde{e}_{x'} < s_M$, a contradiction. This proves the claim.

We have thus established that $E_{\bar{x}}(\mu') = \frac{1}{2}E_F$ and $\tilde{e}_{x'} < s_M$ for all $x' \geq \bar{x}$, so that $E_{x'} < \frac{1}{2}E_F + (x' - \bar{x})s_M < \frac{3}{4}E_F$. Since $E_x = \frac{1}{2}E_F \neq 0$, and since $E_1(\mu)$ is increasing in μ, we have $\mu' > \mu_L$. Moreover, since $E_1(\mu') \leq \frac{3}{4}E_F < E_1$, we have $\mu' \leq \mu_C$, since $E_1(\mu)$ is increasing in μ (as $E_x(\mu)$ is increasing in μ.). Thus $\mu_C > \mu_L$.

To show that $\mu_C < \mu_H$, let us fix x. Then by continuity of $\tilde{D}(e|e)$ in e, we can pick a $\delta > 0$ s.t. for $\mu' < \mu_H$ with $|\mu_H - \mu'| < \delta$, we have $B_F - (1 -$

$x)\tilde{D}(\bar{e}_x(\mu')|\bar{e}_x(\mu')) > B_F - (1-x)\tilde{D}(\bar{e}_x(\mu_H)|\bar{e}_x(\mu_H)) - \varepsilon = x\tilde{D}(\bar{e}_x(\mu_H)|\bar{e}_x(\mu_H) - \varepsilon$, since $B_F - \tilde{D}(\bar{e}_x(\mu_H)|\bar{e}_x(\mu_H)) = 0$ by definition of μ_H. When ε is small enough, this expression is strictly positive, so the person works efficiently towards completing the task in period x and hence in all future periods by Proposition 2. Hence she completes it, which implies that $\mu' \geq \mu_C$, so that $\mu_H > \mu' \geq \mu_C$. □

We now prove Proposition 3.

Proof. Write $B = \mu B_F$, for some fixed B_F, then we will establish the result for the range of μs instead of for B.

Let μ_L be determined via $\mu_L B_F - \tilde{D}(\bar{e}_0|0) = 0 \iff \mu_L = (1-\alpha)D(E_F)$ since $\bar{e}_0 = E_F$. Let μ_H be determined via $\mu_H B_F - \tilde{D}(\bar{e}_0|\bar{e}_0) = 0 \iff \mu_H = (1-\alpha)D(E_F) + \alpha E_F \cdot D'(E_F)$. Then by Lemma 11, we know that for $\mu \leq \mu_L$ the person exerts 0 effort in all periods, and for $\mu \geq \mu_H$ the person exerts effort $\bar{e}_0 = e^*$ in all periods.

From Lemma 14, we know that $\mu_C \in (\mu_L, \mu_H)$. We only need to establish the results when $\mu \in (\mu_L, \mu_C)$ and when $\mu \in (\mu_C, \mu_H)$.

Claim: Consider $\mu \in (\mu_L, \mu_H)$. Then $x_q > 0$.

$\mu \in (\mu_L, \mu_H)$ implies that $\mu B_F - \tilde{D}(\bar{e}_0|0) > \mu_L B_F - \tilde{D}(\bar{e}_0|0) = 0$ and that $\mu B_F - \tilde{D}(\bar{e}_0|\bar{e}_0) < \mu_H B_F - \tilde{D}(\bar{e}_0|\bar{e}_0) = 0$. Hence by Proposition 8, we know that $\tilde{e}_0 = s_q$ with $B_F - \tilde{D}(\bar{e}_0|s_q) = 0$. Given the inequalities just establishe, such s_q is strictly between 0 and $\bar{e}_0$, so that $\tilde{e}_0 \in (0, \bar{e}_0)$. Since $\tilde{e}_x$ and $\bar{e}_x$ are both continuous in x, there is some $\delta > 0$ s.t. $\tilde{e}_x \in (0, \bar{e}_x)$ for all $x < \delta$, therefore $\tilde{e}_x \notin \arg\max_e U(x, e, \tilde{e}_x)$, since the only possible maxima are 0 and $\bar{e}_x$. Therefore x_q from Proposition 2 is strictly greater than δ, which proves the claim.

Claim: $\mu \in (\mu_L, \mu_C)$ implies $x_q < 1$, and $\tilde{e}_x = 0$ for all $x \geq x_q$.

$\mu \in (\mu_L, \mu_C) \implies E_1(\mu) \in (0, E_F)$. Let $\varepsilon > 0$ be s.t. $\varepsilon < E_F - E_1(\mu)$. Then for every $x < 1$, $E_x(\mu) \leq E_1(\mu) \implies \bar{e}_x = \frac{E_F - E_x(\mu)}{1-x} \geq \frac{E_F - E_1(\mu)}{1-x} > \frac{\varepsilon}{1-x}$. Hence $B_F - (1-x)\tilde{D}(\bar{e}_x|0) = B_F - (1-x)(1-\alpha)D(\bar{e}_x) < B_F - (1-\alpha)(1-x)D(\varepsilon/(1-x)) = B_F - (1-\alpha)\varepsilon \cdot D(y)/y$ where $y = \varepsilon/(1-x)$. As $x \to 1$, $y \to \infty$, so by Lemma 13, $D(y)/y \to \infty$. This implies that there is some $x_0 < 1$ with $y = \varepsilon/(1-x_0)$ s.t. $B_F - (1-\alpha)\varepsilon D(y)/y < 0$. Hence $B_F - \tilde{D}(\bar{e}_{x_0}|0) < 0$, so that by Proposition 7 we have that $\tilde{e}_x = 0$ for all $x \geq x_0$. Thus $x_q \leq x_0 < 1$. Since $\tilde{e}_x = e^*_{x_q}$ for all $x > x_q$ by Proposition 2 and since $\tilde{e}_x = 0$ for all $x \geq x_0 \geq x_q$, we must have that $\tilde{e}_x = 0$ for all $x \geq x_q$ (not just for all $x \geq x_0$). This proves the claim.

Claim: $\mu \in (\mu_C, \mu_H)$ implies $x_q < 1$ and $\tilde{e}_x = \tilde{e}_{x_q} = \bar{e}_{x_q}$ for all $x \geq x_q$.

$\mu \in (\mu_C, \mu_H)$ implies that $E_1(\mu) = E_F$, and that $x_q(\mu_C) > 0$. Therefore Lemma 12 implies that $E_x(\mu) > E_x(\mu_C)$ for all $x \in (0, 1)$. Moreover, $E_x(\mu) < E_x(\mu_H) = x \cdot E_F < E_F$, that is the person never finishes the task before period $x = 1$. That means that $\tilde{e}_x > 0$ for all $x < 1$, since if $\tilde{e}_x = 0$ for some x, then it is 0 for all following periods (Proposition 7), so that $E_1 = E_x < E_F$, a contradiction.

Consider $\mu' = \mu - \varepsilon > \mu_C$. Then for both μ and μ', we know that $E_1 = E_F$, so we know that they both work on all days. Let $q = x_q(\mu)$ and $q' = x_q(\mu')$. We know from Lemma 12 that either $q < q'$ and thus the μ-type completes the task efficiently, as we want. Or $q' < q$ and the μ-type does not work at all from q' onwards, which contradicts the fact that they work on all days (necessary for completing the task). Or $q' = q$, in which case μ' stops working entirely from q' on, which also contradicts that $E_q(\mu') = 1$. This proves the claim.

Let us now show that $x_q(\mu)$ is continuous in (μ_L, μ_C). Consider $\mu \in (\mu_L, \mu_C)$, then we know that $x_q(\mu) \in (0, 1)$ and $\tilde{e}_x(\mu) = 0$ for all $x \geq x_q(\mu)$. At $x = x_q(\mu)$, the person at time $s = 0$ is indifferent between 0 work and working towards the optimal amount of work: $U(x, \bar{e}_x(\mu), 0, \mu) = 0$. Consider $\mu_C > \mu' > \mu$, so that $E_x(\mu') > E_x(\mu)$, which implies $\bar{e}_x(\mu') < \bar{e}_x(\mu)$. Therefore $\mu' B_F - (1-x)\tilde{D}(\bar{e}_x(\mu')|0) > \mu B_F - (1-x)\tilde{D}(\bar{e}_x(\mu)|0) = 0$, so $\tilde{e}_x(\mu') > 0$. Hence $x_q(\mu') > x_q(\mu)$. However at $x = x_q(\mu) + \varepsilon$, we have $U(x, E_x(\mu), \bar{e}_x(\mu), 0) < 0$, so that when μ' is sufficiently close to μ, we also get $U(x, E_x(\mu'), \bar{e}_x(\mu'), 0) < 0$, so that the person never works past $x_q + \varepsilon$, proving that it is continuous.

Virtually identical arguments work when $\mu \in (\mu_C, \mu_H)$.

Finally, let us prove that $x_q \to 1$ as $\mu \to \mu_C$. Fix some $\bar{x} < 1$. Let $E^-_{\bar{x}}$ be the quantity for which the person is exactly indifferent between no effort and full effort at time $\bar{x}$ and $s = 0$. When $\mu > \mu_L$, we know that $\mu B_F - \tilde{D}(E_F|0) > 0$, so $\mu B_F - (1-\bar{x})\tilde{D}(E_F|0) > 0$. Hence $E^-_{\bar{x}} < (1-\bar{x})E_F$. By continuity of $E_{\bar{x}}(\mu)$ in μ, and given that $E_{\bar{x}}(\mu_H) = (1-\bar{x})E_F$, we can pick some μ^- s.t. $E_{\bar{x}}(\mu) = E^-_{\bar{x}}$. Then $0 \in \arg\max_e\{U(\bar{x}, e, 0, \mu^-)\}$, and thus by Proposition 7, we know that the person never works in the following periods. It is also clear that the person works in all preceding periods: we know that by period $\bar{x}$ the person has completed $E^{-}_{\bar{x}}$, so if they no longer did any work from period $x' < \bar{x}$ onwards, then $E_{x'}(\mu) = E_{\bar{x}}(\mu)$, but then they would not be indifferent between no work and full work at the start of period x', but strictly prefer working.

This proves that $\mu < \mu_C$ (the person does not complete the task) and $x_q = \bar{x}$. Since $\bar{x}$ was arbitrary, we can let it converge to 1. Since $x_q(\mu)$ is increasing in μ, that implies that $x_q(\mu) \to 1$ as $\mu \to \mu_C$ from below.

The same line of reasoning works from above, showing that $x_q \to 1$ as $\mu \to \mu_C$ from above.

This completes the proof. □

I now state a Lemma that used to prove Proposition 4.

Lemma 15. *Let $R_x := B_F - (1-x)D(\bar{e}_x)$ be the utility as of period x from doing the remaining effort to complete the task (ignoring past sunk costs). Then:*

- *$D''' > 0$ and $R_x \geq 0$ implies $\dot{R}_x > 0$, with*
- *$D''' < 0$ and $R_x \leq 0$ implies $\dot{R}_x < 0$*
- *$D''' = 0$, then $sign(R_x) = sign(\dot{R}_x)$*

Proof. First I prove that when $D''' < 0$, a task that is (weakly) worth doing in period x becomes strictly more worth doing in times $x' > x$ given how the person works on it.

The result is trivial when the person works strictly efficiently in period x. The task being worth doing means $B_F - (1-x)D(\bar{e}_x) \geq 0$, so that the person exerts strictly positive effort on it in period x, but not full. This means that in the IVP from Proposition 8, we are in the case where $\tilde{e}_x = s_q(x)$.

Then the change in actual (not perceived) net utility over time is

$$\begin{aligned}\frac{d}{dx}(B_F - (1-x)D(\bar{e}_x)) &= D(\bar{e}_x) - (1-x)D'(\bar{e}_x)\frac{d}{dx}\bar{e}_x \\ &= D(\bar{e}_x) - (1-x)D'(\bar{e}_x)\frac{\bar{e}_x - s_q(x)}{1-x} \\ &= D(\bar{e}_x) - D'(\bar{e}_x)(\bar{e}_x - s_q(x))\end{aligned}$$

where I plugged in the value for the derivative of $\bar{e}_x$ from the following derivation:

$$\frac{d}{dx}\bar{e}_x = \frac{d}{dx}\frac{E_F - E_x}{1-x} = \frac{-\dot{E}_x(1-x) + (E_F - E_x)}{(1-x)^2} = \frac{\bar{e}_x - \dot{E}_x}{1-x} = \frac{\bar{e}_x - s_q(x)}{1-x}$$

since $\dot{E}_x = \tilde{e}_x = s_q(x)$.

We know that $B_F \geq D(\bar{e}_x)$, which we use in the implicit definition of $s_q(x)$: $0 = B_F - (1-\alpha)D(\bar{e}_x) - \alpha D'(s_q(x)) \cdot \bar{e}_x \geq D(\bar{e}_x) - (1-\alpha)D(\bar{e}_x) - \alpha D'(s_q(x)) \cdot \bar{e}_x = \alpha\,(D(\bar{e}_x) - D'(s_q(x)) \cdot \bar{e}_x) \implies D'(s_q(x)) \cdot \bar{e}_x \geq D(\bar{e}_x)$. (Note that similarly $B_F \leq D(\bar{e}_x) \implies D'(s_q(x)) \cdot \bar{e}_x \leq D(\bar{e}_x)$.)

Looking at Figure 2 (a), we see that the region under the horizontal line through the point $(s_q, D'(s_q))$ has the area $D'(s_q(x)) \cdot \bar{e}_x$, and the blue area under the curve D' has the area $D(\bar{e}_x)$. We have thus shown above that when $B_F \geq D(\bar{e}_x)$, the blue area is smaller than the area under the horizontal line. But the difference in these two areas is $X - Y$, where X and Y are the labelled areas Figure 2 (b). We thus have $X \leq Y$.

But from the concavity of D' (since $D''' < 0$), it is clear in Figure 2 (b) that $Z > Y$ and $X > W$, hence $Z > Y \geq X > W \implies Z > W$. Finally, since $D(\bar{e}_x) = X + Z + V$ and $D'(\bar{e}_x)(\bar{e}_x - s_q) = W + X + V$, we have that $D(\bar{e}_x) - D'(\bar{e}_x) \cdot (\bar{e}_x - s_q(x)) = Z - W > 0$. Hence the net utility strictly increases with x.

When $B_F \leq D(\bar{e}_x)$ and $D''' > 0$, an exactly analoguous argument shows that the net utility strictly decreases with x.

Finally, when $D''' = 0$, the arguments above yield weak inequalities when $R_x = 0$. But when using $R_x > 0$, they yield strict inequalities, as needed for the Lemma. □

We can now prove Proposition 4.

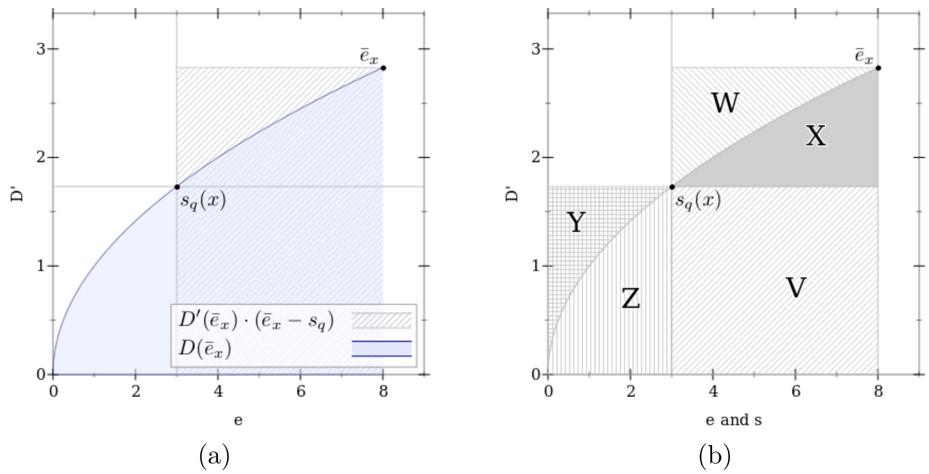


Figure 2: Comparing areas of regions

Proof. Suppose we start with a task such that $B_I = D(\bar{e}_0)$ in period 0 – the unbiased person is indifferent between doing the task efficiently and not doing it. Then when $D''' < 0$, we know that the task becomes strictly better with x from Lemma 15. This means that the person always works at least a little each day, since at time $s = 0$ in period x, she overestimates the value of the task. Therefore for every $x < 1$, the person exerts some effort on the task, which implies that $B_F \geq B_C$ from Proposition 3, since if $B_F < B_C$ then the person does not work at all for all $x > x_q$ with $x_q < 1$.

To show that $B_I > B_C$, note time $x = 0.5$. The remaining task has strictly positive value from being completed efficiently (ignoring past sunk costs). That is, $B_I - (1 - 0.5)D(\bar{e}_{0.5}(B_I)) > 0$. By continuity of $\bar{e}_{0.5}(B)$ in B, there is some $B'_I < B_I$ s.t. $\bar{e}_{0.5}(B'_I)$ satisfies $B'_I - (1 - 0.5)D(\bar{e}_{0.5}(B'_I)) = 0$. Since this task becomes strictly better as the person works on it, it becomes strictly worth doing after some time, and therefore the person works on all days, so this task is also completed. Hence $B_I > B'_I \geq B_C$. So there are $B_F \in (B_C, B_I)$ with $B_F < B_I = D(E_F)$ that she completes, despite them not being worth completing even efficiently.

An analoguous argument shows that when $D''' > 0$, then $B_I < B_C$, which shows that there are some tasks that are strictly worth doing, yet the person does not complete them. The same is true for $D''' = 0$, which completes the proof. □

A.3 Proofs for Section 5

Lemma 16. *Let $F(x, y)$ be a strictly concave real-valued function on some convex domain. Then $G(x) := \max_y F(x, y)$ is strictly concave in x.*

Proof. Let $y(x) = \arg\max_y F(x, y)$ which exists and is unique by strict concavity over a convex domain. Consider a and b with $\beta a + (1 - \beta)b = x$ for some $\beta \in (0, 1)$, and $c \in \arg\max_y F(a, y)$ and $d \in \arg\max_y F(b, y)$. So $G(a) = F(a, c)$ and $G(b) = F(b, d)$. Let $z = \beta c + (1 - \beta)d$. Then

$$\beta G(a) + (1 - \beta)G(b) = \beta F(a, c) + (1 - \beta)F(b, d) < F(x, z) \leq \max_y F(x, y) = G(x)$$

where the strict inequality holds by strict concavity of $F(\cdot, \cdot)$. This proves the claim. □

Proof of proposition 5:

Proof. Since x and E_x are constants during period x, I will drop them from the list of arguments in what follows. I remind the reader that the period-x flow utility is $U_S(e_S, e_L, s) = U_L(e_L, e_S, s) = B_S(e_S) + B_L'(E_x + (1 - x)p_L(s)) \cdot e_L - \tilde{D}(e_S + e_L|s)$, where $p_L(s)$ is the effort the person plans (at time s) to exert on the long-term task in future periods. Note that U_i is strictly concave as a function of e_S and e_L. In what follows, $i, j \in \{S, L\}$, with $i \neq j$, i being the task done first, j the task done second. From Definition 4 we have that $\tilde{e}_i(x) = \inf\{s : V_i(e, s) < V_i(s, s), \forall e > s\}$, where $V_i(e, s) = \max_{e_j} U_i(e_i, e_j, s)$. Since U_i is strictly concave in e_i and e_j, $V_i(e_i)$ is strictly concave in e_i by Lemma 16. It is clear that V is also strictly decreasing in s, and since $\frac{\partial^2}{\partial e \partial s} U = -\alpha D''(s) < 0$, we also have that $\frac{\partial^2}{\partial e \partial s} V < 0$.

Let $\tilde{e}_i(s)$ be the unique maximizer of $V_i(e_i, s)$ – the amount of effort planned after s hours of work – and $\tilde{e}_i$ the actual amount of effort done in period x. Then from Lemma 1, we know that $\tilde{e}_i = \tilde{e}_i(\tilde{e}_i)$. It thus satisfies the FOCs at time $s = \tilde{e}_i$, because it is the planned (and hence perceived optimal) effort. So we will first solve the FOCs for planned effort for arbitrary s, then plug in the correct value for s.

From Definition 4 we have that

$$\begin{aligned}\tilde{e}_j(x) &= \inf\{s-\tilde{e}_i : U_j(e_j,\tilde{e}_i,s) < U_j(s-\tilde{e}_i,\tilde{e}_i,s),\, \forall e_j > s-\tilde{e}_i\}\\ &= \inf\{s : U_j(e_j,\tilde{e}_i,s+\tilde{e}_i) < U_j(s,\tilde{e}_i,s+\tilde{e}_i),\, \forall e_j > s\}\\ &= \inf\{s : U_j^+(e_j,\tilde{e}_i,s) < U_j^+(s,\tilde{e}_i,s),\, \forall e_j > s\}\end{aligned}$$

where $U_j^+(e_j,e_i,s) = U_j(e_j,e_i,s+\tilde{e}_i))$, a translation of U_j by the constant $\tilde{e}_i$ in the s-dimension. Lemma 1 applies, so that $\tilde{e}_j = \arg\max_{e_j} U_j^+(e_j,\tilde{e}_i,\tilde{e}_j) = \arg\max_{e_j} U_j(e_j,\tilde{e}_i,\tilde{e}_j+\tilde{e}_i)$.

Case 1: Long-term task done first First we solve for the effort on the long-term task and the planned (not actual) effort on the short-term task at the time the person stops working on the long-term task.

We have 4 unknowns, two e_i's that maximize period x flow utility and two p_i's that maximize total future utility denoted by $T(x,E,p_L,p_S,s) = B_S(p_S(s)) + B_L(E + (1-x)\cdot p_L(s)) - (1-x)\cdot \tilde{D}(p_S(s)+p_L(s)|s)$. (Note: $\partial_x F$ denotes $\partial F/\partial x$.)

$$\begin{aligned}&\partial_{e_L} V_L(\tilde{e}_L(s),s) = 0 \implies B_L^{'}(E_x + (1-x)p_L(s)) = \tilde{D}(\tilde{e}_S(s)+\tilde{e}_L(s)|s)\\ &\partial_{e_S} U_L(\tilde{e}_L(s),\tilde{e}_s(s),s) = 0 \implies B_S^{'}(\tilde{e}_S(s)) = \tilde{D}^{'}(\tilde{e}_S(s)+\tilde{e}_L(s)|s)\\ &\partial_{p_L} T(p_L(s),p_S(s),s) = 0 \implies B_L^{'}(E_x + (1-x)p_L(s)) = \tilde{D}^{'}(p_S(s)+p_L(s)|s)\\ &\partial_{p_S} T(p_L(s),p_S(s),s) = 0 \implies B_S^{'}(p_S(s)) = \tilde{D}^{'}(p_S(s)+p_L(s)|s)\end{aligned}$$

The last two conditions do not depend on effort in period x (which contribute only infinitesimally). Any solution $(p_L(s),p_S(s))$ to the final two equations is also a solution to the first two FOCs, with $e_L(s) = p_L(s)$ and $e_S(s) = p_S(s)$, which are the unique solutions by strict concavity. Hence we can rewrite the FOCs for period-x effort while working on the long-term task as

$$B_S^{'}(\tilde{e}_S(s)) = \tilde{D}^{'}(\tilde{e}_S(s)+\tilde{e}_L(s)|s) = B_L^{'}(E_x + (1-x)\tilde{e}_L(s))$$

Plugging in $s = \tilde{e}_i = \tilde{e}_L$ for the time at which the person stops working on the long-term task, actual long-term effort $\tilde{e}_L$ and planned short-term effort (at that time) are determined by:

$$B_S^{'}(\tilde{e}_S(\tilde{e}_L)) = \tilde{D}^{'}(\tilde{e}_S(\tilde{e}_L) + \tilde{e}_L|\tilde{e}_L) = B_L^{'}(E_x + (1-x)\tilde{e}_L) \tag{4}$$

Let $E = 0$ and $x = 0$ to solve for $e_{L,0}$, the effort on the long-term task in period $x = 0$.

Claim: long-term effort is constant at $\tilde{e}_{L,x} = \tilde{e}_{L,0}$. We will show that $\tilde{e}_{L,x} = \tilde{e}_{L,0}$ and $\tilde{e}_S(\tilde{e}_{L,x}) = \tilde{e}_S(\tilde{e}_{L,0})$ solve the IVP. Substituting both into the FOC (equation 4), we get

$$B_S^{'}(\tilde{e}_S(\tilde{e}_{L,0})) = \tilde{D}^{'}(\tilde{e}_S(\tilde{e}_{L,0}) + \tilde{e}_{L,0}|\tilde{e}_{L,0}) = B_L^{'}(E_x + (1-x)\tilde{e}_{L,0})$$

Plugging $\tilde{E}_{x'} = \int_0^{x'} \tilde{e}_{L,x}dx = x' \cdot \tilde{e}_{L,0}$ into this, we have

$$B_S^{'}(\tilde{e}_S(\tilde{e}_{L,0})) = \tilde{D}^{'}(\tilde{e}_S(\tilde{e}_{L,0})+\tilde{e}_{L,0}|\tilde{e}_{L,0}) = B_L^{'}(E_x+(1-x)\tilde{e}_{L,0}) = B_L^{'}(x\cdot\tilde{e}_{L,0}+(1-x)\tilde{e}_{L,0}) = B_L^{'}(\tilde{e}_{L,0})$$

which is the FOC solved by $\tilde{e}_{L,0}$ and $\tilde{e}_S(\tilde{e}_{L,0})$, so it holds for all x, proving the claim.

Moving to the actual effort spent on the short-term task, $\tilde{e}_S$, we showed that it satisfies $\tilde{e}_j = \arg\max_{e_j} U_j(e_j, \tilde{e}_i, \tilde{e}_j + \tilde{e}_i)$, where in our current case $i = L$ and $j = S$. The FOC is:

$$\partial_{e_S} U_S(\tilde{e}_S, \tilde{e}_L, \tilde{e}_S + \tilde{e}_L) = 0 \iff B_S^{'}(\tilde{e}_S) = \tilde{D}^{'}(\tilde{e}_S + \tilde{e}_L|\tilde{e}_S + \tilde{e}_L) = D^{'}(\tilde{e}_S + \tilde{e}_L)$$

where the last equality follows because $\tilde{D}^{'}(s|s) = (1-\alpha)D^{'}(s) + \alpha D^{'}(s) = D^{'}(s)$. Since we established that $\tilde{e}_L = \tilde{e}_{L,0}$, we immediately see that $\tilde{e}_S = \tilde{e}_{S,0}$, so that short-term effort is the same in every period as well.

We can compare these to the FOCs for the optimal effort levels:

$$B_S^{'}(e_S^*) = D'(e_S^* + e_L^*) = B_L^{'}(e_L^*) \tag{5}$$

This is similar to the FOCs at the time the person stops working on the long-term task, except that the marginal disutility is replaced by $\tilde{D}^{'}(\cdot|\tilde{e}_L)$. Denote by $f^* := e_S^* + e_L^*$ the total optimal level. It is straightforward to show that $\tilde{e}_L < f^*$: if not, then $D'(\tilde{e}_S(\tilde{e}_L) + \tilde{e}_L|\tilde{e}_L) > \tilde{D}^{'}(e_S^* + e_L^*|\tilde{e}_L) \geq D'(e_S^* + e_L^*)$. But then

$B_L^{'}(e_L^*) < B_L^{'}(\tilde{e}_L) \implies e_L^* > \tilde{e}_L$, which is a contradiction. So $\tilde{e}_L < f^*$. Let $\tilde{f}(\tilde{e}_L) := \tilde{e}_L + \tilde{e}_S(\tilde{e}_L)$ be the total effort the person plans at time $s = \tilde{e}_L$.

Claim: $\tilde{D}^{'}(\tilde{f}(\tilde{e}_L)|\tilde{e}_L) < D^{'}(f^*)$**.** Suppose not. Then $\tilde{D}^{'}(\tilde{f}(\tilde{e}_L)|\tilde{e}_L) \geq D^{'}(f^*) \implies B_L^{'}(\tilde{e}_L) = B_S^{'}(\tilde{e}_S(\tilde{e}_L)) \geq B_L^{'}(e_L^*) = B_S^{'}(e_S^*) \implies \tilde{e}_L \leq e_L^*$ and $\tilde{e}_S(\tilde{e}_L) \leq e_S^*$. So $\tilde{f}(\tilde{e}_L) \leq f^*$. Then $\tilde{D}^{'}(\tilde{f}(\tilde{e}_L)|\tilde{e}_L) < \tilde{D}^{'}(\tilde{f}(\tilde{e}_L)|\tilde{f}(\tilde{e}_L)) = D'(\tilde{f}(\tilde{e}_L)) \leq D'(f^*)$, so that $B_L(\tilde{e}_L) < B_L(e_L^*) \implies \tilde{e}_L > e_L^*$ and similarly $\tilde{e}_S(\tilde{e}_L) > e_S^*$, contradicting $\tilde{f}(\tilde{e}_L) \leq f^*$. This proves the claim by contradiction.

Using the claim, we have $\tilde{D}^{'}(\tilde{f}(\tilde{e}_L)|\tilde{e}_L) < D^{'}(f^*) \implies B_L^{'}(\tilde{e}_L) < B_L^{'}(e_L^*) \implies \tilde{e}_L > e_L^*$.

The FOC for the short-term task is $B_S^{'}(\tilde{e}_S) = D^{'}(\tilde{e}_S + \tilde{e}_L)$ compared to the one for the optimum which is $B_S^{'}(e_S^*) = D^{'}(e_S^* + e_L^*)$, with $\tilde{e}_L > e_L^*$. We have $D^{'}(e_S^* + \tilde{e}_L) > D^{'}(e_S^* + e_L^*) = B_S^{'}(e_S^*)$, and $D^{'}(e + \tilde{e}_L) - B_S^{'}(e)$ is strictly increasing in e (by strict convexity and concavity of D and B), equal to 0 at $\tilde{e}_S$ and strictly positive at e_S^*. Hence $\tilde{e}_S < e_S^*$, which in turn implies that $D^{'}(\tilde{e}_S + \tilde{e}_L) = B_S^{'}(\tilde{e}_S) > B_S^{'}(e_S^*) = D^{'}(e_S^* + e_L^*) \implies \tilde{e}_S + \tilde{e}_L > e_S^* + e_L^*$, proving the results for case 1.

Case 2: Short-term task done first Following identical arguments as in case 1, but with the order of tasks changed, we get that the amount of short-term work is determined first via the following FOCs:

$$B_S^{'}(\tilde{e}_S) = \tilde{D}^{'}(\tilde{e}_S + \tilde{e}_L(\tilde{e}_S)|\tilde{e}_S) = B_L^{'}(E_x + (1-x)\tilde{e}_L(\tilde{e}_S)) \tag{6}$$

The FOCs for $\tilde{e}_L$ depend on planned $p_L(s)$ and $p_S(s)$ which still change after deciding on $\tilde{e}_S$, so it is determined by the following three FOCS, where $\tilde{f} = \tilde{e}_S + \tilde{e}_L$ determines $\tilde{e}_L = \tilde{f} - \tilde{e}_S$ since $\tilde{e}_S$ is fixed once the person works on the long-term task:

$$\begin{aligned}
&B_L^{'}(E_x + (1-x)p_L(\tilde{f})) = \tilde{D}^{'}(\tilde{f}|\tilde{f}) = D^{'}(\tilde{f}) \\
&B_L^{'}(E_x + (1-x)p_L(\tilde{f})) = \tilde{D}^{'}(p_S(\tilde{f}) + p_L(\tilde{f})|\tilde{f}) \\
&B_S^{'}(p_S(\tilde{f})) = \tilde{D}^{'}(p_S(\tilde{f}) + p_L(\tilde{f})|\tilde{f})
\end{aligned}$$

From the first two equations, we see that $\tilde{f} = p_L(\tilde{f}) + p_S(\tilde{f})$, i.e. the person plans to exert the same total effort in all periods when stopping – but without necessarily splitting it across the two tasks the same as today. So we get relevant FOCs:

$$B_S'(p_S(\tilde{f})) = D'(p_S(\tilde{f}) + p_L(\tilde{f})) = B_L'(E_x + (1-x)p_L(\tilde{f})) \tag{7}$$

with $\tilde{e}_L = \tilde{f} - \tilde{e}_S = p_L(\tilde{f}) + p_S(\tilde{S}) - \tilde{e}_S$.

We are going highlight the dependence on x explicitly again and write $\tilde{e}_i(x)$ for the actual effort exerted on task i in period x, and $\tilde{f}(x)$ for the total effort across both tasks.

First, note that $\tilde{e}_S(x) > p_S(\tilde{f}(x))$ as long as $\tilde{e}_S(x) > 0$: in that case, $\tilde{e}_S(x)$ is interior and solves the FOC exactly. Both $\tilde{e}_S(x)$ and $p_S(x)$ solve the same FOCs but with a higher s for $p_S(\tilde{f}(x))$ than for $\tilde{e}_S(x)$, i.e. a higher perceived marginal disutility for $\tilde{e}_S(x)$ than $p_S(\tilde{f}(x))$ – so they plan to work less.

Then $\tilde{e}_L(x) = p_L(\tilde{f}(x)) + p_S(\tilde{f}(x)) - \tilde{e}_S(x) < p_L(\tilde{f}(x))$. By continuity in x, there is some δ s.t. $\tilde{e}_L(x') < p_L(\tilde{f}(x))$ for all $x' \in [x, x+\delta)$, so that

$$E_{x+\delta} < E_x + \delta p_L(\tilde{f}(x)) \iff E_{x+\delta} - E_x < \delta p_L(\tilde{f}(x)) \tag{8}$$

Claim: $\tilde{f}(x+\delta) > \tilde{f}(x)$ when $\tilde{e}_S > 0$, with equality if $\tilde{e}_S = 0$. Suppose the claim does not hold, so that $\tilde{f}(x+\delta) \leq \tilde{f}(x)$. The relevant FOCs are:

$$\begin{aligned} &B_L'(E_{x+\delta} + (1-x-\delta)p_L(\tilde{f}(x+\delta))) = D'(\tilde{f}(x+\delta)) = B_S'(p_S(\tilde{f}(x+\delta))) \\ &B_L'(E_x + (1-x)p_L(\tilde{f}(x))) = D'(\tilde{f}(x)) = B_S'(p_S(\tilde{f}(x))) \end{aligned}$$

If $\tilde{f}(x+\delta) \leq \tilde{f}(x)$, then $D'(\tilde{f}(x+\delta)) \leq D'(\tilde{f}(x))$ so that $B_S'(p_S(\tilde{f}(x+\delta))) \leq B_S'(p_S(\tilde{f}(x))) \implies p_S(\tilde{f}(x+\delta)) \geq p_S(\tilde{f}(x))$ and also

$$\begin{aligned} &B_L'(E_{x+\delta} + (1-x-\delta)p_L(\tilde{f}(x+\delta))) \leq B_L'(E_x + (1-x)p_L(\tilde{f}(x))) \\ \implies &E_{x+\delta} + (1-x-\delta)p_L(\tilde{f}(x+\delta)) \geq E_x + (1-x)p_L(\tilde{f}(x)) \\ \implies &E_{x+\delta} - E_x - \delta p_L(\tilde{f}(x)) \geq (1-x-\delta)(p_L(\tilde{f}(x)) - p_L(\tilde{f}(x+\delta))) \\ \implies &0 > (1-x-\delta)(p_L(\tilde{f}(x)) - p_L(\tilde{f}(x+\delta))) \text{ using equation 8} \\ \implies &p_L(\tilde{f}(x+\delta)) > p_L(\tilde{f}(x)) \end{aligned}$$

where the strict inequality holds only when $\tilde{e}_s > 0$. Together these imply that $\tilde{f}(x+\delta) = p_S(\tilde{f}(x+\delta)) + p_L(\tilde{f}(x+\delta)) > p_S(\tilde{f}(x)) + p_L(\tilde{f}(x)) = \tilde{f}(x)$, which is a contradiction and thus shows the claim.

Claim: $\tilde{e}_S(x) > \tilde{e}(x+\delta)$. We prove this claim like the previous claim via contradiction: suppose that $\tilde{e}_S(x) \leq \tilde{e}(x+\delta)$. Note that in the following FOCs, I write with abuse of notation $\tilde{f}(x)$ for $\tilde{f}(x|\tilde{e}(s))$, i.e. the total work *planned* at time $\tilde{e}_S(x)$ to reduce notation. The relevant FOCs are:

$$
\begin{aligned}
&B_L^{'}(E_{x+\delta} + (1-x-\delta)p_L(\tilde{e}_S(x+\delta))) = \tilde{D}^{'}(\tilde{f}(x+\delta)|\tilde{e}_s(x+\delta)) = B_S^{'}(\tilde{e}_S(x+\delta)) \\
&B_L^{'}(E_x + (1-x)p_L(\tilde{e}_S(x))) = \tilde{D}^{'}(\tilde{f}(x)|\tilde{e}_s(x)) = B_S^{'}(\tilde{e}_S(x))
\end{aligned}
$$

If $\tilde{e}_S(x+\delta) \geq \tilde{e}_S(x)$ then:

$$
\begin{aligned}
&B_S^{'}(\tilde{e}_S(x)) \geq B_S^{'}(\tilde{e}_S(x+\delta)) \\
\Longrightarrow\ & B_L^{'}(E_{x+\delta} + (1-x-\delta)p_L(\tilde{e}_S(x+\delta)) \leq B_L^{'}(E_x + (1-x)p_L(\tilde{e}_S(x))) \\
\Longrightarrow\ & E_{x+\delta} + (1-x-\delta)p_L(\tilde{e}_S(x+\delta) \geq E_x + (1-x)p_L(\tilde{e}_S(x)) \\
\Longrightarrow\ & E_{x+\delta} - E_x - \delta p_L(\tilde{e}_S(x)) \geq (1-x)\left(p_L(\tilde{e}_S(x) - p_L(\tilde{e}_S(x+\delta))\right) \\
\Longrightarrow\ & 0 > p_L(\tilde{e}_S(x) - p_L(\tilde{e}_S(x+\delta))
\end{aligned}
$$

Adding this to the assumed inequality, noting that $p_S(\tilde{e}_S(x)) = \tilde{e}_S(x)$, we have $\tilde{f}(x+\delta) = p_L(\tilde{e}_S(x+\delta)) + p_S(\tilde{e}_S(x+\delta)) = p_L(\tilde{e}_S(x+\delta)) + \tilde{e}_S(x+\delta) > p_L(\tilde{e}_S(x)) + \tilde{e}_S(x) = \tilde{f}(x)$. So $\tilde{D}^{'}(\tilde{f}(x+\delta)|\tilde{e}_S(x+\delta)) > \tilde{D}^{'}(\tilde{f}(x)|\tilde{e}_S(x))$, which from the FOCs implies that $B_S^{'}(\tilde{e}_S(x+\delta)) > B_S^{'}(\tilde{e}_S(x))$ so that $\tilde{e}_S(x+\delta) < \tilde{e}_S(x)$, contradicting our initial assumption and thus proving the claim.

Now let us denote again by $\tilde{f}(x)$ the total effort done in period x. Then since $\tilde{f}(x)$ is strictly increasing in x and $\tilde{e}_S(x)$ is strictly decreasing, we must have that $\tilde{e}_L(x) = \tilde{f}(x) - \tilde{e}_S(x)$ is strictly increasing over time.

This proves all the claims except that $E_x < E_x^*$. (Once $\tilde{e}_S = 0$, it is easy to show that effort on the task is optimal given E_x – so the effort stays constant.) We proved above that $\tilde{e}_L < p_L(\tilde{f})$, and the latter is the optimal effort level conditional on the amount of effort exerted so far. Thus at $x = 0$, the effort is initially lower than optimal, so by continuity we can find δ such that effort is lower than optimal, so $E_\delta < E_\delta^*$.

This never changes: suppose not, so that there was a first time (by continuity of E_x) $\bar{x}$ with $E_{\bar{x}} = E^*_{\bar{x}}$. We know that $\tilde{e}_L(\bar{x}) \leq p_L(\bar{x})$. Thus since $\tilde{e}_L(x)$ is increasing over time (and strictly so initially when $\tilde{e}_S > 0$), we have $\tilde{e}_L(x') \leq p_L(\tilde{f}(\bar{x}))$ for all $x' < \bar{x}$. Since $p_L(\tilde{f}(\bar{x}))$ is the optimal effort going foward in period $\bar{x}$ when having done $E_{\bar{x}} = E^*_{\bar{x}}$, it must be equal to the optimal effort level $\bar{x}e^*_L$, since at time $\bar{x}$ the person has completed the optimal effort level, albeit in a non-constant fashion. But then $\tilde{e}_L(x') \leq p_L(\tilde{f}(\bar{x})) = e^*_L(\bar{x})$ for all $x' < \bar{x}$ with strict inequality for all $x' < \delta$, so that $E_{\bar{x}} = \int_0^{\bar{x}} \tilde{e}_L(x')dx' < \bar{x}e^*_L = E^*_{\bar{x}}$, a contradiction.

This completes the proof. $\square$

References

Ahn, David S, Ryota Iijima, and Todd Sarver. 2020. "Naivete About Temptation and Self-Control: Foundations for Recursive Naive Quasi-Hyperbolic Discounting." *Journal of Economic Theory* 189. Elsevier: 105087.

Alaoui, Larbi, and Christian Fons-Rosen. 2021. "Know When to Fold'em: The Flip Side of Grit." *European Economic Review* 136. Elsevier: 103736.

Ariely, Dan, and George Loewenstein. 2006. "The Heat of the Moment: The Effect of Sexual Arousal on Sexual Decision Making." *Journal of Behavioral Decision Making* 19 (2): 87–98.

Ariely, Dan, and Klaus Wertenbroch. 2002. "Procrastination, Deadlines, and Performance: Self-Control by Precommitment." *Psychological Science* 13 (3). SAGE Publications: 219–224.

Augenblick, Ned, and Matthew Rabin. 2019. "An Experiment on Time Preference and Misprediction in Unpleasant Tasks." *Review of Economic Studies* 86 (3). Oxford University Press: 941–975.

Badger, Gary J., Warren K. Bickel, Louis A. Giordano, Eric A. Jacobs, George Loewenstein, and Lisa Marsch. 2007. "Altered States: The Impact of Immediate Craving on the Valuation of Current and Future Opioids." *Journal of Health Economics* 26 (5): 865–876.

Battaglini, Marco, Roland Bénabou, and Jean Tirole. 2005. "Self-Control in Peer Groups." *Journal of Economic Theory* 123 (2). Elsevier: 105–134.

Bénabou, Roland, and Jean Tirole. 2004. "Willpower and Personal Rules." *Journal of Political Economy* 112 (4). The University of Chicago Press: 848–886.

Bisin, Alberto, and Kyle Hyndman. 2020. "Present-Bias, Procrastination and Deadlines in a Field Experiment." *Games and Economic Behavior* 119. Elsevier: 339–357.

Bray, Robert L, Decio Coviello, Andrea Ichino, and Nicola Persico. 2016. "Multitasking, Multiarmed Bandits, and the Italian Judiciary." *Manufacturing & Service Operations Management* 18 (4). Informs: 545–558.

Buchheim, Lukas, and Thomas Kolaska. 2017. "Weather and the Psychology of Purchasing Outdoor Movie Tickets." *Management Science* 63 (11). Informs: 3718–3738.

Buehler, Roger, Dale Griffin, and Michael Ross. 1994. "Exploring the" Planning Fallacy": Why People Underestimate Their Task Completion Times." *Journal of Personality and Social Psychology* 67 (3). American Psychological Association: 366.

Bushong, Benjamin, and Tristan Gagnon-Bartsch. 2020. "An Experiment on Interpersonal Projection Bias."

Busse, Meghan R, Devin G Pope, Jaren C Pope, and Jorge Silva-Risso. 2015. "The Psychological Effect of Weather on Car Purchases." *The Quarterly Journal of Economics* 130 (1). Oxford University Press: 371–414.

Chaloupka, IV, Frank J, Matthew R Levy, and Justin S White. 2019. *Estimating Biases in Smoking Cessation: Evidence from a Field Experiment*. Working Paper 26522. Working Paper Series. National Bureau of Economic Research. doi:10.3386/w26522.

Chang, Tom Y, Wei Huang, and Yongxiang Wang. 2018. "Something in the Air: Pollution and the Demand for Health Insurance." *The Review of Economic Studies* 85 (3). Oxford University Press: 1609–1634.

Conlin, Michael, Ted O'Donoghue, and Timothy J. Vogelsang. 2007. "Projection Bias in Catalog Orders." *American Economic Review* 97 (4): 1217–1249.

Coviello, Decio, Andrea Ichino, and Nicola Persico. 2014. "Time Allocation and Task Juggling." *American Economic Review* 104 (2): 609–623. doi:10.1257/aer.104.2.609.

Coviello, Decio, Andrea Ichino, and Nicola Persico. 2015. "The Inefficiency of Worker Time Use." *Journal of the European Economic Association* 13 (5): 906–947. doi:10.1111/jeea.12129.

Duckworth, Angela L, Christopher Peterson, Michael D Matthews, and Dennis R Kelly. 2007. "Grit: Perseverance and Passion for Long-Term Goals." *Journal of Personality and Social Psychology* 92 (6). American Psychological Association: 1087.

Fudenberg, Drew, and David K Levine. 2006. "A Dual-Self Model of Impulse Control." *American Economic Review* 96 (5): 1449–1476.

Galak, Jeff, Justin Kruger, and George Loewenstein. 2013. "Slow down! Insensitivity to Rate of Consumption Leads to Avoidable Satiation." *Journal of Consumer Research* 39 (5): 993–1009.

Gilbert, Daniel T., Elizabeth C. Pinel, Timothy D. Wilson, Stephen J. Blumberg, and Thalia P. Wheatley. 1998. "Immune Neglect: A Source of Durability Bias in Affective Forecasting." *Journal of Personality and Social Psychology* 75 (3): 617.

Harris, Christopher, and David Laibson. 2013. "Instantaneous Gratification." *The Quarterly Journal of Economics* 128 (1). MIT Press: 205–248.

Herrnstein, Richard J, and Dražen Prelec. 1991. "Melioration: A Theory of Distributed Choice." *The Journal of Economic Perspectives.* JSTOR, 137–156.

Hsiaw, Alice. 2013. "Goal-Setting and Self-Control." *Journal of Economic Theory* 148 (2). Elsevier: 601–626.

Huang, Yu-Jui, and Adrien Nguyen-Huu. 2018. "Time-Consistent Stopping Under Decreasing Impatience." *Finance and Stochastics* 22 (1). Springer: 69–95.

John, Anett. 2020. "When Commitment Fails: Evidence from a Field Experiment." *Management Science* 66 (2). INFORMS: 503–529.

Kahneman, Daniel, Barbara L Fredrickson, Charles A Schreiber, and Donald A

Redelmeier. 1993. "When More Pain Is Preferred to Less: Adding a Better End." *Psychological Science* 4 (6). SAGE Publications: 401–405.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *The Quarterly Journal of Economics* 112 (2): 375–406. http://qje.oxfordjournals.org/content/112/2/375.abstract.

Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan. 2015. "Self-Control at Work." *Journal of Political Economy* 123 (6). University of Chicago Press Chicago, IL: 1227–1277.

Laibson, David. 1997. "Golden Eggs and Hyperbolic Discounting." *The Quarterly Journal of Economics* 112 (2). MIT Press: 443–478.

Levy, Matthew. 2009. "An Empirical Analysis of Biases in Cigarette Addiction." Working Paper.

Le Yaouanq, Yves, and Peter Schwardmann. 2019. "Learning About One's Self." CEPR Discussion Paper No. DP13510.

Loewenstein, George, and Daniel Adler. 1995. "A Bias in the Prediction of Tastes." *The Economic Journal*. JSTOR, 929–937.

Loewenstein, George, Daniel Nagin, and Raymond Paternoster. 1997. "The Effect of Sexual Arousal on Expectations of Sexual Forcefulness." *Journal of Research in Crime and Delinquency* 34 (4). Sage Publications: 443–473.

Loewenstein, George, Ted O'Donoghue, and Matthew Rabin. 2003. "Projection Bias in Predicting Future Utility." *The Quarterly Journal of Economics* 118 (4): 1209–1248.

Michel, Christian, and André Stenzel. 2020. "Model-Based Evaluation of Cooling-Off Policies."

Nordgren, Loran F, Joop van der Pligt, and Frenk van Harreveld. 2008. "The Instability of Health Cognitions: Visceral States Influence Self-Efficacy and Related Health Beliefs." *Health Psychology* 27 (6). American Psychological Association: 722.

O'Donoghue, Ted, and Matthew Rabin. 2001. "Choice and Procrastination." *Quarterly Journal of Economics*, 121–160.

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *The American Economic Review* 89 (1): 103–124.

Quah, John K-H, and Bruno Strulovici. 2013. "Discounting, Values, and Decisions." *Journal of Political Economy* 121 (5). University of Chicago Press Chicago, IL: 896–939.

Read, Daniel, and Barbara Van Leeuwen. 1998. "Predicting Hunger: The Effects of Appetite and Delay on Choice." *Organizational Behavior and Human Decision Processes* 76 (2): 189–205.

Van Boven, Leaf, and George Loewenstein. 2003. "Social Projection of Transient Drive States." *Personality and Social Psychology Bulletin* 29 (9). Sage Publications: 1159–1168.