

Self-awareness of biases in time perception ^{*}

Isabelle Brocas

*University of Southern California
and CEPR*

Juan D. Carrillo

*University of Southern California
and CEPR*

Jorge Tarrasó

University of Southern California

November 2017

Abstract

We investigated self-awareness in time perception using three time production tasks with different reward structures, and collected self-assessments of performance. Participants had monetary incentives to target the true time in the first (baseline) task, not to exceed the true time in the second task and not to fall below the true time in the third task. We found that participants overestimated time in all tasks but responded correctly to incentives: they decreased their estimates in the second task and increased them in the third. Participants' self-assessment in the baseline task was in line with their time perception biases, and their behavior in the other tasks was consistent with their (correct) beliefs. Self-perceived over-estimators decreased their estimates in the second task significantly more relative to self-perceived under-estimators, while in the third task they increased their estimates significantly less. Last, we explored the effect of physiological stress by having half of our population complete the Cold Pressure Task and found no significant effect.

Keywords: time perception, decision-making, self-awareness, metacognition.

JEL Classification: C91, D83.

^{*}We are grateful to Andreas Aristidou, Calvin Leather and the members of the Los Angeles Behavioral Economics Laboratory (LABEL) for their insights and comments in the various phases of the project. We also acknowledge the financial support of the National Science Foundation grant SES-1425062. Address for correspondence: Juan D. Carrillo, Department of Economics, University of Southern California, 3620 S. Vermont Ave., Los Angeles, CA 90089, USA. E-mail: <juandc@usc.edu>.

1 Introduction

How did it get so late so soon? – Dr. Seuss

The literature in psychophysics documents how we make time evaluations and demonstrates that our subjective assessments of the time that passes may be inaccurate. When individuals are explicitly given time targets, discrepancies between objective time and subjective time are observed as a function of the time duration range – below or above 1 second –, but also the specific task design (Wearden and Lejeune (2008), Grondin (2010), Grondin (2014)). Furthermore, emotional states (Droit-Volet and Wearden (2002), Droit-Volet and Meck (2007), Fayolle et al. (2015)), attentional constraints (Buhusi and Meck (2006)), drug usage (Wittmann et al. (2007)) and incentives to be accurate (Çoskun et al. (2015), Akdogan and Balci (2016)) also shape our perception of time and perception accuracy. Even though the mechanisms of time-keeping are reasonably well understood, little is known about how *aware* people are of their potential biases and whether this knowledge affects their behavior when facing time-related incentives.

This question is related to our ability of being aware that we know something, a concept referred to as metacognition (Koriat (2007), Nelson (1996)). Metacognition has been applied to subjective time only recently (Sackett et al. (2010), Lamotte et al. (2012), Akdogan and Balci (2017)) to show that individuals hold beliefs about distortions in their experience of time that affect their time-related judgements. However, it is unclear how such beliefs impact decision making. For instance, if a person is typically late, is she bound to fail to meet important deadlines? Or is she conscious enough of her tendency to be able to correct it when the incentives are sufficiently high?

The objective of this study is to embed a classical time production paradigm into a decision-making paradigm, where participants are not given explicit time targets but, instead, *choose their time targets as a response to varying incentives*. With this approach, we can investigate the ability of individuals to internalize their perceived biases in their decision of when to act. This study complements the growing literature that links timing to decision-making. However, instead of investigating timing as a decision-making process itself as in traditional accumulator timing models (Treisman (1963), Gibbon (1977)) or in studies looking into the processing dynamics underlying temporal decisions (Balci and Simen (2014)), we investigate the decisions that result from timing. Said differently, we model timing properties as an input for decision-making rather than an output of decision-making.¹

¹There is also a small theoretical and experimental literature in economics that links time perception to time preferences (Ebert and Prelec (2007); Ray and Bossaerts (2011); Brocas et al. (2016); Capra and Park (2016)).

Our study faced four main challenges. First, to isolate biases in time-keeping behavior, it was crucial to choose a task that relied as little as possible on orthogonal concerns. To prevent interference with memory processes, we avoided retrospective paradigms (Fraisse (1984), Block and Zakay (1997)) and we opted for a prospective paradigm in which participants were informed beforehand that they had to make a time related judgement. We also favored time production (the length of the interval to produce is disclosed) to time reproduction (participants have to reproduce an interval of time of unknown length). Even though time production tasks require some form of memory regarding the length of the intervals involved in the experiment, they do not tax working memory to process and store information on the time interval to reproduce as time reproduction tasks do (Baudouin et al. (2006)). We finally designed a novel task that discouraged chronometric counting (Grondin and Killeen (2009), Wearden (2003), Hinton and Rao (2004)) to better capture intrinsic timing properties.

Second, to be able to focus on how participants adjust their time targets to respond to incentives, it was critical to design a task in which we varied the structure of the rewards and paid participants as a function of their performance. This design departs from traditional experiments on subjective time. Indeed, with a few exceptions (Wearden and Grindrod (2003), Akdogan and Balci (2016)) participants are usually asked to produce accurate reports and are compensated with a fixed payment for their effort. Incentivized payments are, however, crucial in decision-making studies because the objective is to trigger cognitive processes that govern real-life choices. Concretely, we included three tasks, each featuring specific time-related incentives. In our baseline task, participants completed a time production task in the range of 31 to 41 seconds, and they were rewarded for accuracy. Formally, they earned maximal payoff (\$20) when they correctly estimated the announced time and their payoffs was reduced symmetrically as their estimates were farther away from the announced time in either direction. We then asked participants to complete two time production tasks in which the reward structure was altered. In one of them, the announced time was a hard deadline: earnings increased as the estimate came close to the announced time and vanished afterwards. This reward scheme incentivized participants to avoid exceeding the deadline and to report lower estimates compared to the baseline task. In the other one, the announced time was a release time: earnings were nil for estimates below the release time, maximal at the release time and decreased afterwards. This scheme incentivized participants to wait past the announced time and to report higher estimates compared to the baseline.

Third, to assess whether participants were aware of their intrinsic timing attitudes, we needed to elicit their beliefs. There is a variety of belief elicitation methods. Recent studies show that eliciting beliefs with properly incentivized methods are often meaningful and

consistent with observed behavior in the laboratory (Palfrey and Wang (2009); Schotter and Trevino (2014)). Furthermore, the experimenter should minimize the chances that the procedure itself affected the behavior under study. In particular, we were not interested in feedback effects, which are known to affect timing behavior (Brown et al. (1995), Franssen and Vandierendonck (2002), Droit-Volet and Izaute (2005)). For these reasons, we opted for an incentivized method and a procedure that did not interfere with the main tasks. At the end of the experiment, we asked our participants to make a self-assessment regarding their performance in the baseline task and we rewarded them for accuracy.

Fourth, to be able to assess biases in time perception and to understand their effects on decision-making, we had to develop a normative theoretical framework that generated predictions which could then be compared to the empirical findings. In the tradition of neoclassical economics, we opted for a rational framework in which decision-makers maximize their expected payoffs but are subject to noisy time perception. More precisely, we modeled individuals as in traditional theories of subjective time (Gibbon et al., 1984) and we assumed that, when they target a given time interval, they produce time intervals on average equal to the target (“mean accuracy”). We derived the optimal behavior of such decision-maker in the context of our time-related incentives. We contrasted the predictions of the model with the actual behavior of participants and we determined whether deviations from the target intervals were due to biases in time perception or biases in decision-making.

Given the growing interest in the role of emotions on time perception (Droit-Volet and Meck (2007), Droit-Volet and Wearden (2002), Fayolle et al. (2015)) and decision-making (Damasio (1994), Schwartz (2000), Phelps et al. (2014)), we were also interested in the effect of emotions on the interaction between the two. Emotion, however, is a complex and rich concept and we opted for an exploration of the effect of physiological stress on time-related incentives. To this purpose, half of our population completed the Cold Pressor Task (CPT), a method that increases the participants’ cortisol levels and has been previously shown to affect decision-making (Porcelli and Delgado (2009), Lighthall et al. (2012)). In particular, in the context of psychological stress, increases in cortisol levels correlate with increases in dopamine release (Pruessner et al. (2004)), which are known to make time perceived as passing more slowly (Meck (1996)). This evidence suggests a possible mechanism linking stress to changes in time perception (through variations in dopamine concentration).² We therefore hypothesized that participants who completed the Cold Pressor Task would act as if time passed more slowly.

To answer our research questions, we followed a simple research strategy. We derived

²Other mechanisms may be at play (Dickinson and Kemeny, 2004), but the dopamine mechanism provides a clear testable prediction in terms of the direction of the effect.

hypotheses for a rational decision-maker who maximizes rewards, and fitted predictions to the data. We used time reports in the baseline task to assess the presence of a time perception bias, for each individual and in the overall population. By comparing the responses between the baseline and each of the two other tasks, we assessed whether participants responded to incentives conditional on their biases. That is, we determined whether biases in behavior in those two tasks were due to biases in time perception or biases in decision-making. Moreover, by comparing behavior between participants who did and did not complete the CPT, we studied the effect of physiological stress on time perception and decision-making. Finally, we used elicited beliefs to evaluate self-awareness and to further ascertain if behavior was contingent on beliefs.

2 Experimental design and procedures

We conducted two treatments of a single experiment in the Los Angeles Behavioral Economics Laboratory (LABEL) at the University of Southern California, with a large number of subjects (170 individuals, 79 Male and 91 Female, in 21 sessions of 6 to 10 participants each). Among those, 83 individuals (36 Male and 47 Female) participated in the “Stress” treatment and underwent the Socially Evaluated Cold Pressure Task (CPT). Subjects completed a survey before entering the laboratory to ensure all criteria for cortisol sampling were satisfied. The remaining 87 individuals participated in the “Control” (non-CPT) treatment. Sessions lasted for about 90 minutes and all started at 3pm to control for the circadian variability in cortisol levels. Instructions were read out loud at the beginning of each task. All methods were carried out in accordance with existing guidelines and the protocol was reviewed and approved by the Institutional Review Board of the University of Southern California (UP-14-00663). We obtained informed consent from each participant.

2.1 Socially Evaluated Cold Pressure Task

We induced stress by submerging the participant’s non-dominant hand in a bucket of ice water for 3 minutes. The procedure was coupled with a social pressure aspect. The level of physiological stress was assessed by collecting three saliva samples per participant (Base, Peak and End of experiment) in order to measure changes in cortisol levels. The details of the procedure are explained in Appendix A1.

2.2 Time perception tasks

The design was adapted from Brocas et al. (2016).

Prospective time estimate. Participants completed three time perception tasks that differ exclusively on the payoff incentives. In each task, participants were asked to report 10 time intervals t of similar but not identical length, without knowing in advance the number or length of intervals to report. This is called a *prospective* time estimate in a *production* paradigm. Prospective refers to a case where participants know in advance that they will be requested to estimate the elapsed time. Production occurs when participants are informed about the length of the interval they must produce (Nichelli, 1996). The intervals were of 31, 32, 33, 34, 35, 36, 37, 38, 39 and 41 seconds, respectively. We designed a Matlab-based program to implement the elicitation of the participants’ time perception. It presented the instructions on the screen and guided them to estimate time intervals. Participants were prompted the length of the interval t to be estimated. Then, participants marked the beginning and end of the interval by clicking on a button on the top right corner of the screen. The order for the 10 intervals was randomly selected but it was the same for all participants. At the beginning of the session, an experimenter explicitly asked participants to put away any time-keeping devices (watches, music players, cell phones, etc.) and made sure they complied. We used computer mice that produced no sound when clicked to ensure that participants could not use auditory cues to infer the behavior of others.

Interfering tasks. To ensure that participants did not count time, we asked them to solve a concurrent interfering task (the “table task”) while estimating the time intervals.³ In the table task, participants were sequentially presented a series of 4×6 tables where each row and column had a name, and they were instructed to click on a specific cell. For example, in one table participants were asked: “Please click the cell where the column to the right of the column labeled athena intersects the row above the biology row.” The names of the rows and columns as well as the phrasing of which cell to click on changed from table to table, to make sure participants would pay attention to the table task and therefore could not count at the same time. There was a random and unspecified time limit to complete each task (between 10 and 15 seconds) and failure to complete it counted as an incorrect answer. Overall, we designed the interfering task in a way that it required sufficient effort to prevent participants from counting but was easy enough to make sure all participants could successfully complete it if they put attention.⁴ In Appendix A2 we provide a screenshot of one such table. Finally, we informed participants beforehand that

³In the debriefing of the pilot sessions preceding the experiment we asked participants whether they could count time while performing the table task, and almost everyone answered negatively.

⁴The task typically used in the interval timing literature consists in asking participants to repeat aloud digits presented on a computer screen (Wearden et al. (1997)). Given we organized sessions with several participants, it was not possible to use such method. We also preferred a distractor task that could be easily incentivized.

performance in a task would not count as correct or incorrect whenever they reported the end of an interval during that task. This was meant to prevent the completion of a table task from interfering with time estimation.

2.3 Earnings

In the time perception tasks, the amount earned was a function of (1) the distance between the announced and reported intervals, and (2) the proportion of correct answers in the interfering task. More precisely, for each participant, one time interval of each task was randomly selected for payment. For that time interval, the participant earned money only if at least 75% of the interfering tasks were correctly answered.⁵

As noted above, the three time-perception tasks differed in the payoff incentives. In the “baseline” task, hereafter B , participants earned \$20 if the report coincided exactly with the announced interval t . Participants lost \$0.5 for each 1% that their report was above or below the announced interval. In the “under-report” task, hereafter U , participants lost \$0.5 for each 1% that their report was below the announced interval and they lost everything if the report was above the announced interval. Last, in the “over-report” task, hereafter O , participants lost \$0.5 for each 1% that their report was above the announced interval and they lost everything if the report was below the announced interval. The three tasks are summarized in Figure 1. Task B was always performed first. The order of tasks U and O was randomized across sessions. The entire procedure was explained beforehand. While participants could potentially obtain as much as \$60, earnings based on the three tasks averaged \$16.9.

2.4 Beliefs

Participants were asked to assess their performance in task B by estimating how many of the 10 reported intervals were below the announced intervals and how many were above: 0 to 2 below and 8 to 10 above, 3 to 7 below and 3 to 7 above, or 8 to 10 below and 0 to 2 above. This question was incentivized with \$1 payment for a correct answer. It was asked at the end of the experiment to prevent any form of feedback from interfering with the choices in the estimation task (e.g., by reassessing own performance). For analysis, participants who responded ‘0 to 2 below and 8 to 10 above’ (respectively ‘3 to 7 below and 3 to 7 above’, respectively ‘8 to 10 below and 0 to 2 above’) were categorized as perceiving themselves as “Over-estimators” (respectively “Unbiased”, respectively

⁵In 83% of the intervals (compared to 84% in Brocas et al. (2016)), participants answered correctly at least 75% of the interfering tasks and therefore were eligible for payment.

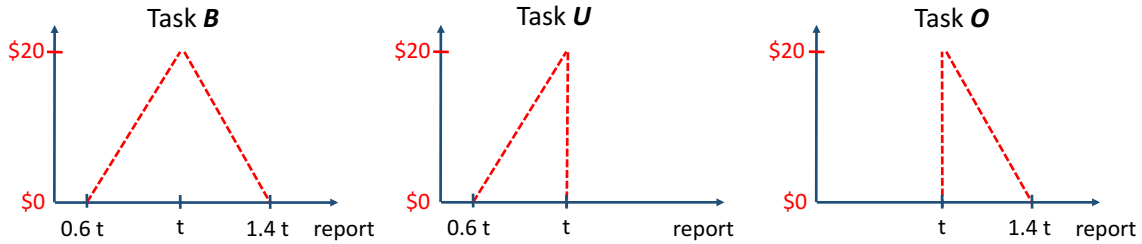


Figure 1: Payments. Given an announced time interval t , participants were paid according to the payoff function illustrated with the dotted line. Payment was maximized when the report was $r = t$. Payments were positive for reports $r \in [0.6t, 1.4t]$ in task B (left), $r \in [0.6t, t]$ in task U (center), and $r \in [t, 1.4t]$ in task O (right).

“Under-estimators”). We used the same categories to classify them based on their actual reports. Accuracy of beliefs was obtained by comparing the perceived category with the actual category of each participant.

2.5 Timeline and saliva sample collection

The precise timeline of the experiment was the following. First, subjects submitted the “Base” saliva sample. Next we read the instructions for task B and performed a few practice rounds of time estimation intervals. Subjects in the non-CPT treatment started task B immediately after. Subjects in the CPT treatment performed the Cold Pressure Task and then proceeded to task B . Exactly 20 minutes after the beginning of task B (so, for CPT subjects, 20 minutes after the end of the CPT procedure), we stopped the experiment and collected the “Peak” saliva sample. We did not need to interrupt any decision-making as in all sessions the timing coincided with the experimenter reading the instructions for the second task (either U or O depending on the session). Subjects completed the second and third tasks, after which they provided the “End” saliva sample. Finally, they completed a brief questionnaire (see below) which included the belief elicitation of their performance in task B . They were paid in private before leaving the experiment. The time between the “Peak” and “End” saliva samples was between 30 and 40 minutes depending on the sessions. A sample copy of the instructions and questionnaire is provided in Appendix B. Notice that, given our timing, maximal cortisol levels for CPT subjects occurs between the end of the first task and the beginning of the second. By the end of the third task, cortisol is expected to be back (or close) to baseline levels.

2.6 Questionnaire

We conducted a questionnaire also at the end of the session to collect demographic information such as gender and primary language spoken. These control questions are often added by default in experiments in Economics. Participants were also asked to report their usual stress level and their stress level on the day of the experiment. We used a scale from 1 (no stress) to 10 (highest stress). The belief elicitation of performance in task B was also part of the questionnaire.

3 Theory and predictions

We start with a theoretical exercise to assess how participants who believe that their subjective perception may be noisy and biased should behave. This exercise allows us to make testable predictions and to identify in our data whether behavior is driven by the existence of a bias that is not corrected, by mistakes in decision-making or by both.

3.1 Model

Background. Consider participant i who targets time interval τ and suppose that his perception p^i of τ is biased by an amount b^i . For simplicity, let us assume it takes the following additive linear form:⁶

$$p^i(\tau) = \tau + b^i \tag{1}$$

To formalize the idea that a participant does not know his true bias and is therefore unable to correct for it, we model b^i as a random variable with p.d.f. $g(b^i)$ and c.d.f. $G(b^i)$. Participant i knows only that his own bias is drawn from the distribution of the bias in the population. Additionally, when participant i wants to report his perception of an interval τ , he does so noisily. Formally:⁷

$$r^i = p^i(\tau) + \varepsilon^i \iff r^i = \tau + b^i + \varepsilon^i \quad \text{where } \varepsilon^i \text{ i.i.d. } N(0, \eta_i^2) \tag{2}$$

⁶We adopt an additive bias for technical simplicity. In our design, the specific functional form of $p^i(\tau)$ (additive, multiplicative, logarithmic, regression towards the mean, etc.) is not crucial since the variation in the announced intervals is small, namely between 31s and 41s. In the limit where there is no variation, all one-parameter functional forms are identical. After consideration, we decided against multiple measures of the same interval to avoid an excessively boring task for participants. For an estimation of a more general time perception function (two-parameter power function) over a wider set of intervals (24s to 196s) we refer to Brocas et al. (2016).

⁷It is important to introduce both a noise and a bias. Indeed, in the absence of noise our model would not allow any variance in bias within individuals (which is not what we empirically observe) and in the absence of bias our model would force the report of every individual to be unbiased (which, again, is not what we empirically observe).

Overall, the difference between the report r^i and the interval τ that the participant targets is due both to the perception bias b^i and the individual noise ε^i .

Assumptions. A desirable property of subjective time is mean accuracy, or the idea that reported time in the population is, on average, unbiased with respect to the target ($E[r^i] = \tau$). We capture this property in our model by assuming no expected bias in the population: $E[b^i] = 0$. This assumption corresponds to the standard “mean accuracy” requirement in the classical scalar expectancy theory. This is also a natural assumption in the context of a rational model, since if there was a known bias at the population level, participants would simply correct for it. In other words, any rational theory must presume the existence of no aggregate bias in the population.

Reported time is also usually modeled as being drawn from a normal distribution. We impose weaker assumptions, namely (i) a symmetric distribution of bias, $g(b^i) = g(-b^i)$, which implies that reports err symmetrically around the target ($\Pr[r^i = \tau + z] = \Pr[r^i = \tau - z]$) and (ii) log-concavity of the distribution of the random variable $a^i = b^i + \varepsilon^i$.⁸

Optimal target. When an interval t is announced in task $k \in \{B, U, O\}$, participant i chooses a target $\theta_k^i(t)$. Given the structure of payoffs, one could think that a risk-neutral participant who maximizes expected payoffs should always target the interval t that results in maximal payoff. However, the participant is aware that his perception may be biased (1) and that his estimation is noisy (2). These affect his incentives to target a certain interval. In Appendix A3, we show that the optimal target $\theta_k^{i*}(t)$ in task k is:

$$\theta_B^{i*}(t) = t; \quad \theta_U^{i*}(t) < t; \quad \theta_O^{i*}(t) > t \quad \text{with} \quad \theta_B^{i*}(t) - \theta_U^{i*}(t) = \theta_O^{i*}(t) - \theta_B^{i*}(t) \quad (3)$$

The intuition is as follows. When payoffs are symmetric around t (Figure 1, left), a payoff-maximizing participant who believes to be unbiased on average should target t , to minimize the expected (upward or downward) deviation. This is not true anymore when payoffs are asymmetric. In task U , participants have incentives to decrease their target because over-reporting ($r^i > t$) is significantly more costly than under-reporting ($r^i < t$). A lower target ensures that the final report remains within the range of positive rewards, that is, below t (Figure 1, center). Using an analogous argument, participants have incentives to increase their target in task O (Figure 1, right). By the symmetry of the problem, the incentives to decrease the target in U and to increase the target in O are identical. This implies that targets in U and O are symmetric with respect to the

⁸Formally, let $f(a^i)$ be the p.d.f. of the random variable $a^i = b^i + \varepsilon^i$. Log concavity requires $\left(\frac{f'(a)}{f(a)}\right)' < 0$. It is a technical condition satisfied by many common distributions, including but not limited to the normal distribution.

announced time t .⁹ Overall, the final report in task k and given the optimal target is:

$$r_B^i(t) = \theta_B^{i*}(t) + b^i + \varepsilon_B^i; \quad r_U^i(t) = \theta_U^{i*}(t) + b^i + \varepsilon_U^i; \quad r_O^i(t) = \theta_O^{i*}(t) + b^i + \varepsilon_O^i \quad (4)$$

Shading. We call shading the amount by which reports, are decreased in task U and increased in task O with respect to task B . For future reference, the shadings of participant i in interval t of tasks U and O are denoted by $h_U^i(t)$ and $h_O^i(t)$, respectively:

$$h_U^i(t) = r_B^i(t) - r_U^i(t) \quad \text{and} \quad h_O^i(t) = r_O^i(t) - r_B^i(t)$$

Using (4) it is immediate that $h_O^i(t) \sim N(\theta_O^i - \theta_B^i, 2\eta_i^2)$ and $h_U^i(t) \sim N(\theta_B^i - \theta_U^i, 2\eta_i^2)$. This means that while the report in each task depends on the bias of the participant, shading across tasks *does not*. This is based on the assumption that while the bias is unknown, it is constant across tasks.

Also, if the participant acts optimally, expected shadings h_U^{i*} and h_O^{i*} must satisfy:

$$h_U^{i*} = \theta_B^{i*} - \theta_U^{i*} \quad \text{and} \quad h_O^{i*} = \theta_O^{i*} - \theta_B^{i*}$$

3.2 Predictions

The theory offers a number of testable predictions. In each task and in each trial, we observe the announced time interval t as well as the report $r_k^i(t)$ made by each participant. We will use that information to estimate the individual biases b^i and the individual shadings $h_U^i(t)$ and $h_O^i(t)$. Let $I_k(r_k^i)$ be the empirical c.d.f. of reports in task k . Also, let $H_U(h_U^i)$ and $H_O(h_O^i)$ be the empirical c.d.f. of shading in U and O . We will test three main predictions of the model.

Prediction 1. *There is no bias in the population in task B : $E[b^i] = 0$.*

Prediction 2. *Participants shade downwards in task U and upwards in task O . Formally, the distributions of reports across tasks satisfy first-order stochastic dominance: $I_O(r) \leq I_B(r) \leq I_U(r)$ for all r .*

Prediction 3. *Participants shade the same amount (though in opposite directions) in U and O . Formally, the distribution of shading between B and U is the same as between O and B : $H_U(h) = H_O(h)$ for all h .*

⁹Trivially, as the variance of the bias and noise go to zero (e.g., because we allow time keeping devices), we get $\theta_U^{i*}(t) \rightarrow t^-$ and $\theta_O^{i*}(t) \rightarrow t^+$. Also, if the participant is risk-averse the incentives to target an interval below t in task U and above t in task O are increased, as it reduces the likelihood to obtain no payoff. However, the incentives to target t in task B remain unchanged since the cost of over- and under-reporting are still symmetric.

Notice that these predictions hold generically, that is, without having to specify a functional form for the distribution of biases in the population, $G(b^i)$.

Last, the theoretical model is not designed to predict how stress should affect behavior in our tasks. In particular, stress may interact not only with the perception of time but also with the ability to act rationally, which is a key element in the derivation of the results above. However, based on the evidence reviewed in the introduction regarding the relationship between stress, dopamine release and time perception, we make the following prediction.

Prediction 4. *Participants who completed the CPT report stochastically higher time estimates in task B than those who did not.*

3.3 Calibration

We can also obtain point predictions of the unobservable (optimal) targets in tasks U and O if we assume that the bias in the population is normally distributed, $b^i \sim N(0, \hat{\sigma}^2)$, where $\hat{\sigma}^2$ is the empirical variance of the bias of our participants in task B . By computing these optimal targets, we can further test the theory and determine whether the expected reports are in line with the targets. Details of the method are provided in Appendix A4.

4 Results

The objective of the empirical analysis was to test Predictions 1-4. We do not report the analysis of each time interval separately in the main text as it is not central to our study. However, the analysis of each interval is briefly addressed in Appendix A5. We also use a p-value of 0.05 as the benchmark threshold for statistical significance.

4.1 Stress

Figure 2 shows the evolution of cortisol levels throughout the experimental sessions in both treatments. Each dot represents the average level of salivary cortisol samples (ng/mL) taken at Base, Peak (20 minutes after the CPT procedure), and End of the experiment (on average, 55 minutes after the CPT procedure). The CPT and non-CPT groups start with statistically indifferent levels of average cortisol (3.07 vs. 3.17; two-sided Welch t -test, p -value = 0.724). The CPT group experiences a large and statistically significant increase in average cortisol at Peak (5.58 vs. 3.07, p -value < 0.001). In comparison, the non-CPT group experiences no statistical change in average cortisol at Peak (3.20 vs. 3.17; p -value = 0.918). We also observe higher cortisol levels in the CPT group than in the non-CPT

in the End sample (3.53 vs. 2.57, p -value < 0.001), suggesting that by the end of the experiment cortisol levels are not fully back to baseline. Overall, the data provides strong evidence that our procedure was successful in raising the cortisol level of the subjects under CPT.

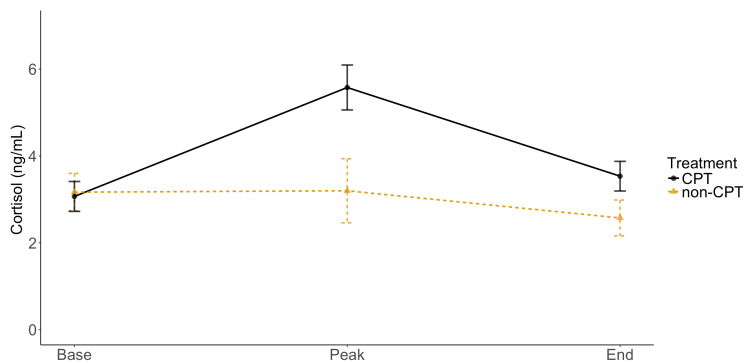


Figure 2: Cortisol levels over time

4.2 Model estimation

Task B: bias in report. For each participant, we estimated the bias in task B across all 10 intervals. The best estimate of the bias of each individual is simply the empirical mean of the difference between the observed report and the announced time. When we averaged the individual biases to obtain the population bias, we got $E[b^i] = 6.86$, which is significantly different from 0 (one sample, t-test, $t = 7.34$, $df = 169$, p -value < 0.001). The median bias was 6.05 and the distribution’s skewness was 0.48. Out of our 170 participants (and using a 5% significance level), 88 had a positive and significant bias, 22 had a negative and significant bias and 58 had no significant bias in either direction.

Overall, and consistent with the existing literature in production of similar time intervals (Matell and Meck (2000), Wittmann et al. (2007)), we found substantial heterogeneity in the bias of our participants, a symmetric distribution of the bias around the mean of the distribution, and a tendency to overestimate time.¹⁰ This result indicated that behavior was not consistent with Prediction 1. Said differently, mean accuracy is not supported by our data.¹¹ Finally, biases were not significantly different between CPT and non-CPT participants (t-test for comparison of means (6.47 vs. 7.27, p -value = 0.67) or Kolmogorov-

¹⁰Note however that such studies pay subjects a fixed fee and therefore lack the proper monetary incentives to be accurate.

¹¹Appendix A5 reports similar findings interval by interval.

Smirnov test for comparison of distributions (p-value = 0.65)). Thus, Prediction 4 was not supported by the data.¹²

Tasks U and O: shading. The data also revealed that the incentives to report intervals in tasks U and O were well-understood. Consistent with Prediction 2, participants provided lower reports in U than in B ($r_U^i < r_B^i$), and higher reports in O than in B ($r_O^i > r_B^i$). This result can best be noted with the cumulative distribution functions of the average report of each individual. As depicted in Figure 3 (left), the distribution of individual average report in O first-order stochastically dominated the distribution in B , which itself first-order stochastically dominated the distribution in U ($I_O(r^i) \leq I_B(r^i) \leq I_U(r^i)$, Wilcoxon rank sum test, p-value < 0.001 for all pairwise comparisons).

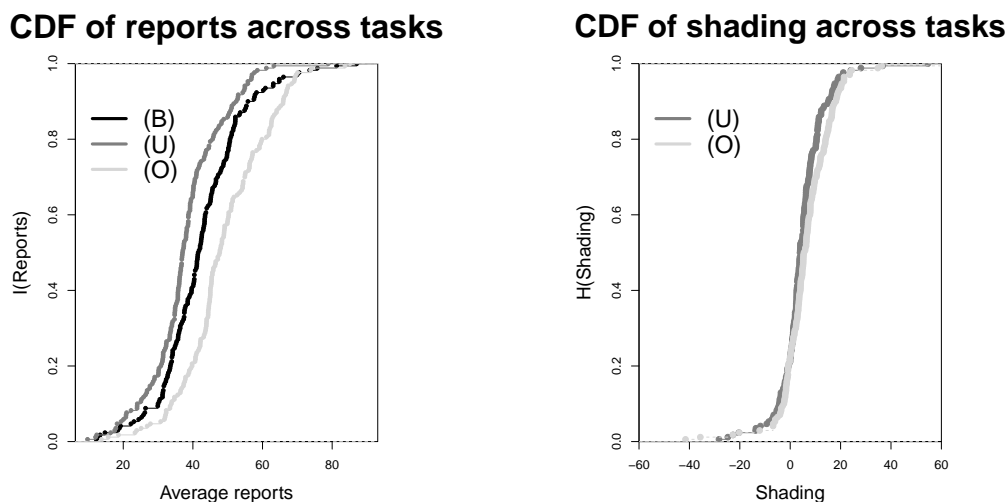


Figure 3: Distributions of reports (left) and shading (right). Participants' reports were significantly smaller in U and significantly higher in O compared to B ($I_O(r) \leq I_B(r) \leq I_U(r)$). Shading was similar in U and O ($H_U(h) \simeq H_O(h)$).

Also consistent with Prediction 3, the amount of downward shading in U ($h_U^i \equiv r_B^i - r_U^i$) was positive and similar in magnitude to the amount of upward shading in O ($h_O^i \equiv r_O^i - r_B^i$), that is, $h_U^i \simeq h_O^i$. Both shadings were significantly above 0 on average ($E[h_U^i] = 4.62$, t-test, $t = 6.43$, $df = 169$, p-value < 0.001 and $E[h_O^i] = 6.55$, t-test, $t = 7.74$, $df = 169$,

¹²The only noticeable treatment effect was that the variance of the distribution of biases was significantly higher under CPT (F-test, $F(82,86) = 1.84$, p-value = 0.005). However, this was exclusively due to the fact that 7 CPT participants had extreme biases, outside the range of the non-CPT participants (F-test, $F(75,86) = 1.01$, p-value = 0.97 after removing these participants).

p-value < 0.001) and not significantly different from each other (t-test, $t = 1.34$, $df = 169$, p-value = 0.181). Again, the result can be seen most clearly with the cumulative distribution functions. As depicted in Figure 3 (right), the distributions of upward and downward shading revealed only marginally more shading in O than in U ($H_U(h) = H_O(h)$, Wilcoxon rank sum test, p-value = 0.070).

Finally, we compared the average shading in U and O in the control (non-CPT) and treatment (CPT) conditions, pooling all observations (all) as well as separating them by the order in which tasks were performed (task U before task O or U-O and task O before task U or O-U). The results are compiled in Table 1.¹³

	all		U-O		O-U	
	non-CPT	CPT	non-CPT	CPT	non-CPT	CPT
$E[h_U^i]$	4.15 (0.99)	5.11 (1.05)	4.82 (1.31)	4.30 (0.86)	3.32 (1.52)	5.80 (1.80)
$E[h_O^i]$	6.20 (1.23)	6.90 (1.17)	5.30 (1.79)	8.59 (2.07)	8.87 (1.36)	4.17 (1.40)

(standard errors in parenthesis)

Table 1: Shading across treatments

Differences in shading between CPT and non-CPT subjects are statistically significant in only 1 out of 6 cases, namely for upward shading when task O is performed before U (8.87 vs. 4.17, p-value = 0.02). In that case, the difference is opposite to the one expected, with stressed subjects shading less than non-stressed ones. Among CPT subjects, differences in shading are not significantly different depending on the order in which tasks U and O were performed. Overall and just like for task B , we found no strong evidence that stress affected behavior in the time perception tasks.

Individual heterogeneity. Although shading followed the basic theoretical predictions at the aggregate level (downwards in U and upwards in O), there was substantial heterogeneity across individuals. Indeed, only 96 participants shaded in the correct direction in both tasks. Among the rest, 37 shaded downwards in both cases, 36 shaded upwards in both cases and 1 shaded upwards in U and downwards in O . When comparing the bias in task B of these participants, we found that participants who shaded consistently with theory had a bias of 6.3 on average. Those who systematically decreased their report had a significantly higher bias (13.2 on average, t-test, $t = -2.43$, $df = 49.1$, p-value = 0.02)

¹³Since cortisol levels peak before the second task (see section 2.5), it is possible that the CPT procedure had an effect in the second but not the third task. Similarly, fatigue may have also had an effect on choices.

while those who systematically increased their report had a significantly lower bias (1.9 on average, t-test, $t = 2.29$, $df = 68.8$, $p\text{-value} = 0.02$). This finding is suggestive of a link at the individual level between behavior in task B and correction in the subsequent tasks.

We were also interested in analyzing the relationship between the amount shading in U and shading in O of each individual. If shading was entirely driven by a self-assessment of the individual noise ε_i (see equation 2), the correlation would be positive and close to 1: subjects with a noisier estimation would depart more from the target in both directions to avoid obtaining no payoff. Instead we found a strong, negative and statistically significant correlation (PCC $(h_U^i, h_O^i) = -0.67$, $p\text{-value} < 0.001$). This result is a first indication of an important finding that will be further studied in sections 4.4 and 5: individuals have a certain belief about their bias in task B , and try to correct it by under-shading in one task and over-shading in the other task.

We also noticed that participants with a smaller absolute bias were also less volatile (PCC $(|b^i|, \eta_B^i) = 0.43$, $p\text{-value} < 0.001$) suggesting that, at the individual level, the ability to produce accurate intervals on average went hand in hand with the ability to not deviate too much from them. The relationship between bias and volatility was not modeled by our theory. It was however consistent with previous research (Brocas et al. (2016)).

Calibration. Using the calibration proposed in section 3.3, we could obtain point estimates of optimal targets (see Appendix A4 for the formal derivation). Given the estimates of the variance of the sample ($\hat{\sigma}^2$) and the individual noise ($\hat{\eta}_i^2$), we found that, on average, participants should target $\theta_U^* = 30.88$ and $\theta_O^* = 40.32$. The empirical average reports were significantly higher than these optimal targets: 37.84 in U and 49.02 in O . However, they were remarkably consistent if we account for the sample’s positive bias. Indeed, given the bias estimated in task B , participants would be expected to report on average $\theta_U^* + E[b^i] = 37.74$ in U and $\theta_O^* + E[b^i] = 47.18$ in O . The first one was not significantly different from the empirical average report (t-test, $t = 0.12$, $df = 169$, $p\text{-value} = 0.906$) and the second one was marginally different (t-test, $t = 1.94$, $df = 169$, $p\text{-value} = 0.055$). This calibration exercise confirmed that, on average, participants were acting (close to) optimally in tasks U and O given the bias estimated in task B .

Earnings. Participants earned on average \$8.4 in B , \$4.7 in U and \$3.9 in O (Figure 4). The distributions of earnings showed significantly lower payoffs in O than in U (Wilcoxon signed rank test, $p\text{-value} = 0.028$). The main reason was that the standard deviation in reports in task O was higher than in task U , resulting in an increased likelihood of falling outside the positive reward zone ($\hat{\sigma}_O = 12.34$ and $\hat{\sigma}_U = 10.57$, F-test for comparison of

variance, $F(169,169) = 0.733$, $p\text{-value} = 0.044$; Figure 4, center and right).¹⁴

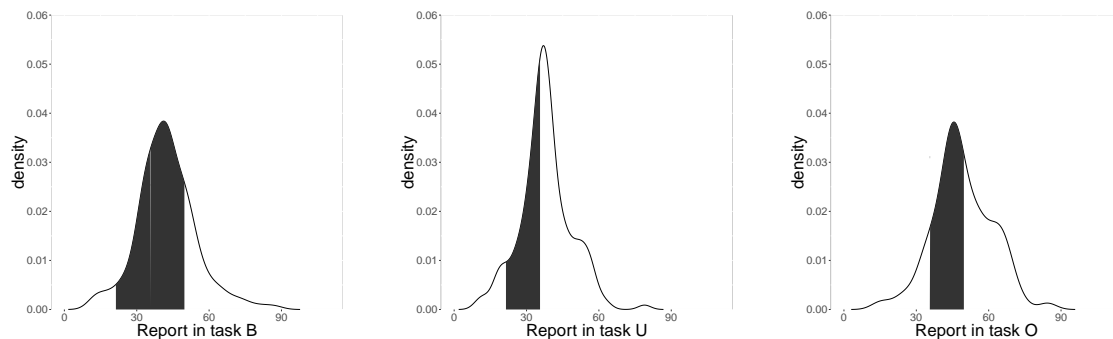


Figure 4: Distribution of earnings in B (left), U (center) and O (right). The curves represent the p.d.f. of reports in each task. The dark area represents the region in which participants earned money in each task: $r_B^i \in [0.6t, 1.4t]$ (task B), $r_U^i \in [0.6t, t]$ (task U) and $r_O^i \in [t, 1.4t]$ (task O).

Questionnaire. We did not find any effect of gender, language or reported stress levels.

Summary. Contrary to Prediction 1, reported intervals were excessively high, resulting in a positive expected bias in the population. Contrary to Prediction 4, reports were not significantly different between stressed and non-stressed subjects. By contrast, in support of Predictions 2 and 3, marginal changes in reports across tasks closely followed the predictions of theory at the aggregate level.

4.3 Regression analysis

An alternative (but formally identical) method to estimate the population bias is to run a simple OLS regression of the difference between individual i 's report and interval in task B , $r_B^i(t) - t$, on a constant. An advantage of this method is that we can include control variables to test for differential effects of treatment variables and demographic characteristics (naturally, once we include the control variables, it is not a structural estimation of the model so the coefficients cannot be interpreted as bias or shading). Table 2 (column 1) presents the result of this regression, with dummy variables for gender ($Male = 1$) and stress ($CPT = 1$), and with robust errors clustered at the individual

¹⁴Naturally, earnings were highest in task B , but these are not comparable since the interval with positive rewards was twice as large in B as in U or O (Figure 4). Standard deviations between B and the other tasks are not comparable either, since task B is always performed first.

level.¹⁵

	(1) $r_B^i(t) - t$	(2) $r_k^i(t) - t$	(3) $r_k^i(t) - t$	(4) $r_k^i(t) - t$
<i>Constant</i>	6.99*** (1.52)	7.69*** (1.29)	7.43*** (1.35)	8.28*** (1.67)
<i>Task O</i>	—	6.57*** (0.85)	6.90*** (1.17)	6.57*** (0.85)
<i>Task U</i>	—	-4.62*** (0.72)	-4.17*** (0.99)	-4.62*** (0.72)
<i>CPT</i>	0.69 (1.89)	0.08 (1.57)	0.61 (1.87)	-1.50 (2.45)
<i>Male</i>	-1.06 (1.86)	-1.93 (1.56)	-1.93 (1.56)	-1.91 (1.56)
<i>CPT* Task O</i>	—	—	-0.68 (1.70)	—
<i>CPT* Task U</i>	—	—	-0.91 (1.44)	—
<i>U-O</i>	—	—	—	-1.09 (1.96)
<i>CPT* U-O</i>	—	—	—	3.25 (3.09)
Observations	1697	5092	5092	5092
adj. R ²	0.001	0.152	0.152	0.156

Standard errors in parenthesis. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 2: OLS regressions of reports and shading

The results in (1) confirmed our previous findings: the bias in task B was positive and highly significant (the difference between the estimated constant in the regression and the coefficient of $E[b^i]$ in the previous section was just due to the control variables). Also, biases were not significantly different between CPT and non-CPT participants nor between males and females.

Next, we ran the same regression of the difference between report and interval, except that we included all three tasks ($k \in \{B, U, O\}$) and added dummy variables for task O (*task O*) and task U (*task U*). The outcome is reported in Table 2 (column 2). Reinforcing the previous results, we found a similar coefficient for the constant. We also found a

¹⁵For simplicity of exposition, we do not include the other answers to our questionnaire. They do not affect any of the results provided in this section.

positive and highly significant coefficient for *Task O*, which captured the amount of upward shading in task *O* relative to task *B*, and a negative and highly significant coefficient for *Task U*, which captured the amount of downward shading in task *U* relative to task *B*. Again those coefficients were similar in magnitude to the shading in the previous section.

Finally, we added some extra terms to the regression in column 2. We first included interaction terms between stress and task ($CPT * Task\ O$ and $CPT * Task\ U$, Table 2 column 3), to determine whether stress had a differential effect across tasks. Then, we added a dummy variable that took value 1 if task *U* was performed before *O* ($U-O$, Table 2 column 4) to determine whether shading depended on the order in which tasks *O* and *U* were performed (due, for example, to fatigue). We also included an interaction term between stress and order of tasks ($CPT * U-O$) since cortisol peaked during the second task and was close to baseline in the third. Consistent with the average findings in Table 1, none of these variables had a significant impact on reports, suggesting that cortisol increases did not significantly affect choices in our time perception tasks. Choices were not significantly different across genders.

4.4 Bias awareness and correction

Given the observed heterogeneity and negative correlation in shading across individuals, we studied the relationship between behavior and self-assessed biases. Under full rationality, every participant should take his possible bias into account, de-bias it and end up producing accurate reports, at least on average. Our analysis indicated that this was not the case, as we observed significant over-reporting in the population. For the same reason, all individuals should believe that they are equally likely to under-report than over-report, so perceive themselves as “Unbiased”.

The question is: assuming that participants were not sophisticated enough to de-bias their behavior, were their beliefs consistent with their behavior (“I think I over-estimated and, indeed, I did over-estimate”)? Also, did they act according to those beliefs in tasks *U* and *O* (“I think I over-estimated and therefore I adjusted my shading accordingly in *U* and *O*”)?

We found that 45 participants (26%) perceived themselves as “Under-estimators,” 78 participants (46%) perceived themselves as “Unbiased,” and 47 participants (28%) perceived themselves as “Over-estimators.” Although, as we already know, the majority of participants reported excessively high time intervals, they had on average correct beliefs about their biases in *B* relative to the population: the distribution of biases among Over-estimators first-order stochastically dominated the distribution of Unbiased and Under-estimators (Wilcoxon rank sum test to compare Over-estimators with Unbiased and Under-estimators respectively, p-value = 0.021 and 0.016). Under-estimators had lower biases

compared to Unbiased participants, although differences in the distributions were not statistically significant (Figure 5, left).

Participants also acted correctly upon their beliefs in tasks U and O . In task U , Under-estimators shaded downwards less than Unbiased and Unbiased shaded downwards less than Over-estimators (Figure 5, center). Symmetrically, in task O Under-estimators shaded upwards more than Unbiased and Unbiased shaded upwards more than Over-estimators (Figure 5, right).¹⁶ These differences were statistically significant, except for the difference between Unbiased and Under-estimators (Wilcoxon rank sum test to compare Under-estimators with Unbiased in U and O respectively, p-value= 0.52 and 0.16; Unbiased with Over-estimators in U and O respectively, p-value= 0.002 and 0.01; Under-estimators with Over-estimators in U and O respectively, p-value= 0.004 and 0.002).¹⁷

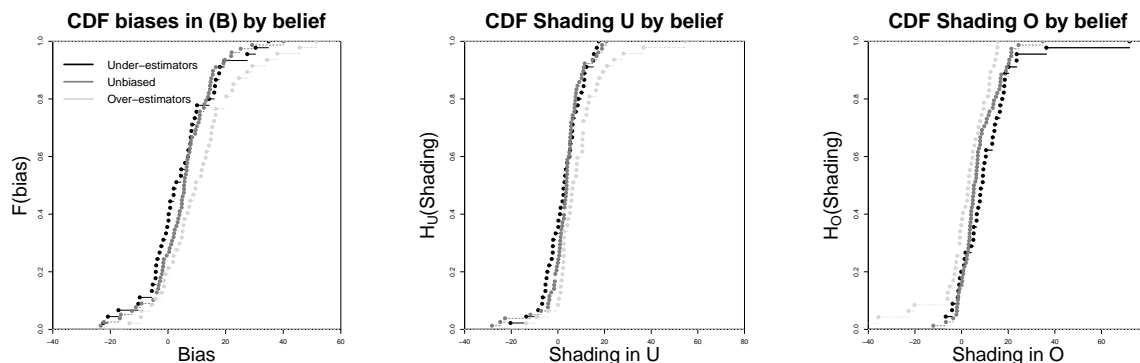


Figure 5: Distributions of biases in B (left) and distributions of shading in U (center) and O (right) as a function of self-perceived beliefs. Biases in B are statistically different across beliefs and consistent with them: Under-estimators are less biased than Unbiased who are less biased than Over-estimators. In U and O , participants shade as a function of their belief.

Overall, even though shading in U and O were of equal magnitude on aggregate, once we conditioned on the self-assessed bias of the participants they were not. Consistent with their beliefs, Over-estimators shaded significantly more in U than in O (Wilcoxon signed rank test, p-value = 0.042) and Under-estimators shaded significantly more in O than in

¹⁶As a referee pointed out, some risk-averse individuals could report to be Under-estimators or Over-estimators against their own beliefs for hedging motives. However, if subjects reported untruthfully as a mechanism for hedging, we should not observe a correlation between the belief reported and their actual behavior in task B . We should not observe a correlation either between the belief reported in task B and the amount of shading in U and O .

¹⁷In most cases, all the differences were significant or close to significant after running a boxplot analysis and removing extreme outliers.

U (Wilcoxon signed rank test, p-value = 0.005). However, we also found that unbiased participants shaded more in O than in U (Wilcoxon signed rank test, p-value= 0.014).

The corrective behavior implied that while average reports were significantly different in task B between Over-estimators and the other two groups (t-test, p-value = 0.009 and 0.012), they were not different between any two groups in tasks U or O (t-test, all p-values > 0.4), see Figure 6 (left).

To sum up, our results indicated that participants had correct beliefs about their time perception bias relative to others in the population. Even though they were not able to anticipate their bias in task B , they correctly acted upon it in tasks U and O . There was again no effect of the CPT procedure on the distribution of self-assessed biases or on the behavior of participants within groups.

To further analyze the relationship between reports and beliefs, we ran similar regressions as in Table 2, where we controlled for the self-assessed beliefs. More precisely, the baseline were the Under-estimators and we included dummies for *Unbiased* and *Over-estimators*.

	(1) $r_B^i(t) - t$	(2) $r_O^i(t) - t$	(3) $r_U^i(t) - t$	(4) $r_O^i(t) - r_B^i(t)$	(5) $r_B^i(t) - r_U^i(t)$
<i>Constant</i>	4.62* (2.05)	16.08*** (2.15)	3.09 (1.66)	11.43*** (2.30)	1.50 (1.65)
<i>Unbiased</i>	1.08 (2.12)	-2.15 (2.30)	0.44 (1.82)	-3.20 (2.15)	0.63 (1.49)
<i>Over-estimators</i>	7.07** (2.65)	-1.77 (2.62)	0.68 (2.16)	-8.91*** (2.60)	6.40** (2.02)
<i>CPT</i>	0.54 (1.82)	-0.12 (1.87)	-0.34 (1.58)	-0.67 (1.63)	0.95 (1.39)
<i>Male</i>	-1.05 (1.79)	-2.47 (1.85)	-2.36 (1.58)	-1.38 (1.63)	1.31 (1.40)
Observations	1697	1698	1697	1697	1697
adj. R ²	0.086	0.054	0.072	0.053	0.049

Standard errors in parenthesis. * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Table 3: OLS regressions of reports and shading on self-assessed beliefs

From column (1) we noticed the by-now familiar positive bias of the population in task B (as captured by the constant term), and a correct relative belief of the self-assessed bias of Over-estimators, who provided higher reports than the other two groups. By contrast, the difference between Unbiased and Under-estimators was not statistically significant.

Reports in task O (column (2)) and task U (column (3)) were not different as a function of the self-assessed beliefs, reinforcing the idea that participants compensate for their biases. This is best understood with columns (4) and (5), where we noticed that Over-estimators exhibited a smaller upward shading in O ($11.42 - 8.91$) and a larger downward shading in U ($1.50 + 6.40$) than Under-estimators. Overall, the regression analysis confirms the differences in corrective behavior by individuals as a function of their self-assessed beliefs, in particular, between Over-estimators and Under-estimators.

5 Behavioral adaptation model

If participants were at least partially aware of their biases, why were they not correcting better for them in tasks U and O ? A possible explanation is that, even though they knew their tendency to misrepresent time in a given direction, they were unsure of the magnitude of their bias. Recall that an unbiased participant i should target t in task B , $\theta_U^{i*}(t)$ in task U and $\theta_O^{i*}(t)$ in task O . Recall also that the optimal shading should be symmetric around t : $t - \theta_U^{i*}(t) = \theta_O^{i*}(t) - t$. Suppose that participant i targeted the symmetric (but possibly suboptimal) intervals $t - \Delta^i$ and $t + \Delta^i$ in tasks U and O , and that he believed he targeted $t + \tilde{b}^i$ instead of t in task B . In his mind:

$$t + \tilde{b}^i - h_U^i = t - \Delta^i \quad \text{and} \quad t + \tilde{b}^i + h_O^i = t + \Delta^i$$

Combining both equations, we could retrieve Δ^i (the target shading around the announced time) and \tilde{b}^i (the perceived bias):

$$\Delta^i = \frac{h_U^i + h_O^i}{2} \quad \text{and} \quad \tilde{b}^i = \frac{h_U^i - h_O^i}{2}$$

In words, a participant with a perceived positive bias ($\tilde{b}^i > 0$), an Over-estimator, would shade downwards in U more than shade upwards in O ($h_U^i > h_O^i$). The opposite would be true for a participant with a perceived negative bias, an Under-estimator.

We used this behavioral model to estimate the perceived bias of each individual based on the observed shadings. We found that the estimated perceived biases of the participants were consistent with their beliefs: -3.38 for Under-estimators (lowest and significantly negative, p-value = 0.008), -1.51 for Unbiased (not significantly different from zero, p-value = 0.053), and 2.94 for Over-estimators (highest and significantly positive, p-value = 0.045). Furthermore, participants acted as if they targeted on average $\Delta = 5.67$, with no significant differences across groups (mean between 5.4 and 5.9, t-tests all p-values > 0.4). This number was in the order of magnitude of our calibrated optimal shading $t - \theta_U^*(t) = \theta_O^*(t) - t \equiv 4.72$. Overall, even though we observed reports as represented in

Figure 6 (left), participants acted as if they believed that their behavior was consistent with the perceived reports represented in Figure 6 (right). This behavioral model provided an as if representation of behavior that was both consistent with self-assessed biases and empirically observed biases.

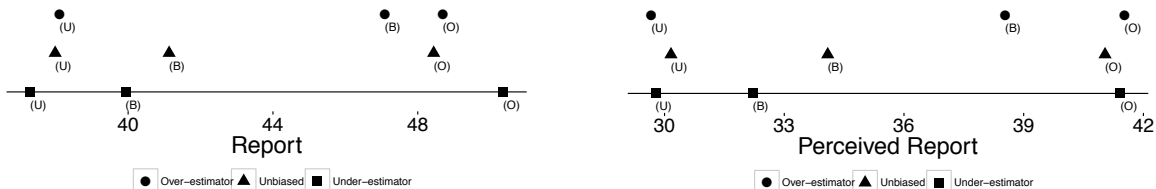


Figure 6: Average reports (left) and average perceived reports (right) of self-perceived Under-estimators, Unbiased and Over-estimators. Reports across groups are different in task *B*, but beliefs are correct and participants shade according to them, so reports across groups are not statistically different in tasks *U* and *O*.

6 Discussion

It has been proposed that perceived time synchronizes with the ticking of an internal clock (Treisman (1963)), an idea mathematically represented by the scalar timing model (Gibbon (1977)). This internal clock consists of a pacemaker that emits pulses (ticking) stored in an accumulator. Accuracy is obtained when the pacemaker is perfectly synchronized with objective time, and distortions occur when it speeds up or slows down. In our framework, when a time interval is announced, the participant chooses a target as a function of the reward (decision stage), and this target is implemented subject to noise and biases (time production stage). This time production stage corresponds to what an accumulator model would focus on. Instead, our main focus is on the decision stage.

The results obtained in our baseline task *B* were comparable to previous studies of production of similar time intervals (Matell and Meck (2000), Wittmann et al. (2007)). Instead of asking directly participants to be as accurate as possible, we rewarded them for accuracy. As expected, this variant in design did not produce inconsistent results. On average, mean accuracy was violated as participants tended to overestimate time. However, the distributions of reports were symmetric, and the scalar property of interval timing was satisfied (see Appendix A5 for details). The observed tendency to overestimate was also consistent with findings in Brocas et al. (2016), where we used the same experimental procedure. In that study, we observed a form of Vierordt’s law over the second to minute range and intervals around 30s were, on average, over-estimated.

A key result of our analysis is that biases in time perception carry over across different incentives schemes while decision-making is optimal *conditional on biases*. This result has two implications. First, it suggests that time-keeping mechanisms are dissociated from decision-making mechanisms. Reports that do not agree with theoretical predictions in U and O result from errors in perception, not from the inability to optimally shade. Second, optimality of decision-making indicates that the decision system is tuned to optimize behavior. These two implications together provide support for modeling the brain as an organization of systems acting optimally within their range of action, in the tradition of earlier brain-based dual process theories (Brocas and Carrillo, 2008). For instance, decision-making that requires time considerations (e.g. dynamic choices or choices under time constraints) could be modeled as an interplay between a time keeping system that processes the perception of time and a decision-making system that receives (possibly biased) time related information and uses it to formulate a decision.

Another key result is that participants had correct beliefs about their choices in B and took correct decisions given those beliefs in tasks U and O (even though beliefs were elicited after completing all three tasks). This suggests that awareness of their own time distortions determined how they made decisions regarding time estimates. However, this knowledge was imperfect. Indeed, participants were aware of their biases but (i) they were not able to anticipate them and correct for them in task B , and (ii) they missed the magnitude of the bias to fully adjust their behavior in tasks U and O . Decisions in the latter tasks suggest that participants possessed at that point specific knowledge about their performance in the previous task B , knowledge that helped them find their best course of action in subsequent tasks. Even in the absence of feedback, participants were able to access introspectively their experience of time and to measure it accurately.

The fact that biases were not anticipated beforehand, and not entirely corrected for afterwards is intriguing. It indicates in particular a limitation of metacognition. Investigating the relationship between intrinsic limitations of metacognition in time perception and behavioral anomalies involving time management, such as chronic lateness, seems to be a promising alley to investigate.

Last, as discussed in the introduction, we expected that the increase in cortisol levels generated by the CPT would have an (unobservable) effect on dopamine release resulting in an (observable) increase in time perception. However, although cortisol levels were significantly increased, time estimation was not significantly different between the two populations. This absence of effect paralleled an absence of differences in self-assessments of stress levels. A possible interpretation of these unexpected findings is that different types of stress may affect information processing differentially, as suggested in Kogler et al. (2015). Such hypothesis would need further investigation.

References

1. Akdogan B. and F. Balci (2016). The effects of payoff manipulations on temporal bisection performance. *Acta Psychologica*, 170, 74-83.
2. Akdogan B. and F. Balci (2017). Are you early or late? Temporal error monitoring. *Journal of Experimental Psychology: General*, in press.
3. Balci F. and P. Simen (2014). Decision processes in temporal discrimination. *Acta Psychologica*, 149, 157-168.
4. Baudouin A., S. Vanneste, V. Pouthas, M. Isingrini (2006). Age-related changes in duration reproduction: Involvement of working memory processes. *Brain and Cognition* 62(1), 17-23.
5. Block R.A. and D. Zakay (1997). Prospective and retrospective duration judgments: a meta-analytic review. *Psychonomic Bulletin and Review*, 4, 184-197.
6. Bos, Ruud Van Den, M. Harteveld, and H. Stoop (2009) "Stress and decision-making in humans: performance is related to cortisol reactivity, albeit differently in men and women", *Psychoneuroendocrinology*, 34(10), 1449-1458.
7. Brocas I. and J.D. Carrillo (2008). The brain as a hierarchical organization. *American Economic Review*, 98(4), 1312-1346.
8. Brocas, I. , J.D. Carrillo and R. Kendall (2017) "Stress Induces Contextual Blindness in Lotteries and Coordination Games" *Frontiers in Behavioral Neuroscience*, 11 Dec. 2017, doi: 10.3389/fnbeh.2017.00236.
9. Brocas I., J.D. Carrillo and J. Tarraso (2016). How long is a minute? *USC, working paper*.
10. Brown S.W., Newcomb D.C. and K.G. Kahr (1995). Temporal-signal detection and individual differences in timing. *Perception*, 24(5), 525-38.
11. Buhusi, C.V. and Meck, W.H. (2006). Interval timing with gaps and distracters: evaluation of the ambiguity, switch, and time-sharing hypotheses. *Journal of Experimental Psychology: Animal Behavior Processes*, 32, 329-338.
12. Capra M. and Park J.Y. (2016), "Anomalous Time Preferences? The Role of Subjective Time Perception", working paper, Claremont Graduate University.
13. Çoşkun F., Sayali Z.C., Gürbüz E. and F. Balci (2015), Optimal time discrimination. *The Quarterly Journal of Experimental Psychology*, 68(2), 381-401.

14. Damasio AR. (1994). *Descartes error: Emotion, reason, and the human brain*. New York: Avon Books.
15. Dickerson S. S. and M. E. Kemeny (2004). Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological Bulletin*, 130(3), 355.
16. Droit-Volet S. and M. Izaute (2005). The effect of feedback on timing in children and adults: the temporal generalization task. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 58(3), 507-20.
17. Droit-Volet S. and W. H. Meck (2007). How emotions colour our perception of time. *TRENDS in Cognitive Sciences*, 11(12).
18. Droit-Volet, S. and Wearden, J. (2002). Speeding up an internal clock in children? Effects of visual flicker on subjective duration. *Quarterly Journal of Experimental Psychology. B, Comparative and psychological psychology*, 55B, 193-211.
19. Ebert J. and D. Prelec (2007) “The Fragility of Time: Time-Insensitivity and Valuation of the Near and Far Future”. *Management Science* 53(9), 1423-1438.
20. Fayolle S., Sandrine Gil, Sylvie Droit-Volet (2015). Fear and time: Fear speeds up the internal clock. *Behavioural Processes*, 120, 135-140.
21. Fraisse, P. (1984). Perception and Estimation of Time. *Annual Review of Psychology*, 35, 1-36.
22. Franssen V. and A. Vandierendonck (2002). Time estimation: does the reference memory mediate the effect of knowledge of results? *Acta Psychologica*, 109(3), 239-267.
23. Gibbon J. (1977). Scalar expectancy theory and Weber’s Law in animal timing. *Psychological Review*, 84, 279 -325.
24. Gibbon, J., Church, R. M. and W.H. Meck (1984). Scalar timing in memory. In J. Gibbon and L. Allan (Eds.), *New York Academy of Sciences*. Timing and time perception.52-77.
25. Granger, D. A., S.B. Johnson, S.L. Szanton and D. Out (2012). Incorporating salivary biomarkers into nursing research: an overview and review of best practices. *Biological Research for Nursing*, 347-356.
26. Grondin S. (2010). Timing and time perception: a review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception and Psychophysics*, 72(3), 561-582.

27. Grondin S. (2014). About the (non)scalar property for time perception'. *Advances in Experimental Medicine and Biology* 829,17-32.
28. Grondin S. and P. Killeen (2009). Tracking time with song and count: different weber functions for musicians and non musicians. *Attention, Perception and Psychophysics*, 71, 1649-1654.
29. Hines, E. A., and G. E. Brown (1936). The cold pressor test for measuring the reactivity of the blood pressure: data concerning 571 normal and hypertensive subjects. *American Heart Journal*, 1-9.
30. Hinton, S. C., and Rao, S. M. (2004). One-thousand one ... one-thousand two ...: Chronometric counting violates the scalar property in interval timing. *Psychonomic Bulletin and Review*, 11(1), 24-30.
31. Kogler L., Veronika I. Mller, Amy Chang, Simon B. Eickhoff, Peter T. Fox, Ruben C. Gur, Birgit Derntl (2015). Psychosocial versus physiological stress - Meta-analyses on deactivations and activations of the neural correlates of stress reactions, *NeuroImage*, 119, 235-251.
32. Koriat A. (2007). Metacognition and consciousness, in P.D. Zelazo, M. Moscovitch and E. Thompson (Eds) *Cambridge Handbook of Consciousness*, 289-325, Cambridge University Press
33. Lamotte M., M. Izaute and S. Droit-Volet (2012). Awareness of time distortions and its relation with time judgment: a metacognitive approach. *Consciousness and Cognition*, 21(2), 835-42.
34. Lighthall N., Michiko Sakaki, Sarinnapha Vasunilashorn, Lin Nga, Sangeetha Somayajula, Eric Y. Chen, Nicole Samii and Mara Mather (2012). Gender differences in reward-related decision processing under stress, *Social Cognitive and Affective Neuroscience*, 7(4), 476-484.
35. Matell M.S. and W.H. Meck (2000). Neuropsychological mechanisms of interval timing behavior, *Bioessays*, 22(1), 94-103.
36. Meck W. H. (1996). Neuropharmacology of timing and time perception. *Cogn. Brain Res.*, 3, 227-242.
37. Nelson T.O. (1996). Consciousness and metacognition. *American Journal of Psychology*, 51, 102-116.
38. Nichelli P. (1996). Time perception measurements in neuropsychology, in M. Pastor, J. Artieda (Eds.) *Advances in Psychology: Time, Internal Clocks and Movement*, Elsevier Amsterdam, 187-204.

39. Palfrey, T. and S.W. Wang (2009) “On eliciting beliefs in strategic games”, *Journal of Economic Behavior & Organization* 71, 98-109.
40. Phelps E.A., Lempert K.M., Sokol-Hessner P. (2014). Emotion and decision making: multiple modulatory neural circuits. *Annual Review Neuroscience*, 37, 263-87.
41. Porcelli A.J. and M.R. Delgado (2009). Acute Stress Modulates Risk Taking in Financial Decision Making. *Psychological Science*, 20(3), 278-283.
42. Preston, SD, TW Buchanan, RB Stansfield, and A. Bechara (2007) “Effects of anticipatory stress on decision making in a gambling task,? *Behavioral neuroscience*, 121(2), 257.
43. Pruessner J.C., F. Champagne, M. J. Meaney, and A. Dagher (2004). Dopamine Release in Response to a Psychological Stress in Humans and Its Relationship to Early Life Maternal Care: A Positron Emission Tomography Study Using [11C]Raclopride, *The Journal of Neuroscience*, 24(11), 2825 - 2831.
44. Ray, D. and P. Bossaerts (2011) “Positive temporal dependence of the biological clock implies hyperbolic discounting”, *Frontiers in Neuroscience* 5(2), 1-5.
45. Sackett A.M., T. Meyvis, L. D. Nelson, B. A. Converse and A. L. Sackett (2010). You’re Having Fun When Time Flies: The Hedonic Consequences of Subjective Time Progression. *Psychological Science*, 21(1), 111-117.
46. Schotter A. and I. Trevino (2014). Belief Elicitation in the Laboratory. *Annual Review of Economics*, 6(1), 103-128.
47. Schwabe L., L. Haddad and H. Schachinger (2008). HPA axis activation by a socially evaluated cold-pressor test?. *Psychoneuroendocrinology*, 890-895.
48. Schwartz N. (2000). Emotion, cognition, and decision making. *Cognition and Emotion*, 14 (4), 433 - 440.
49. Treisman M. (1963). Temporal discrimination and the indifference interval: implications for a model of the internal clock’. *Psychological Monographs*, 77.
50. Wearden J.H. (2003). Applying the scalar timing model to human time psychology: progress and challenges. In H. Helfrich (Ed), *Time and mind II*, 21-39.
51. Wearden J.H., L. Denovan, M. Fakhti and R. Haworth (1997). Scalar timing in temporal generalization in humans with longer stimulus durations. *Journal of Experimental Psychology: Animal Behavior Processes*, 23, 502-511.

52. Wearden J.H. and R. Grindrod (2003). Manipulating decision processes in the human scalar timing system. *Behavioural Processes*, 61(1-2), 47-56.
53. Wearden J.H. and H. Lejeune (2008). Scalar properties in human timing: conformity and violations. *The Quarterly Journal of Experimental Psychology*, 61(4), 569-87.
54. Wittmann M., David S. Lelanda, Jan Churand, Martin P. Paulus (2007). Impaired time perception and motor timing in stimulant-dependent subjects. *Drug and Alcohol Dependence*, 90, 183-192.

Appendix A

Appendix A1. Physiological stress and the Cold Pressure Task (CPT).

The Cold Pressure Task (CPT) is a well-established method to induce physical stress and has been safely used on adults and children since the 1930s (Hines and Brown, 1936). Each subject had an elastic band placed on their left arm slightly above the protrusion where the ulna connects to the wrist. Subjects were then required to spread their fingers and place their non-dominant hand in ice water at 4 to 6 degrees Celsius to a depth that reached the elastic band for a certain period of time (in our experiment, 3 minutes). To ensure that CPT was a salient stressor, we used a slight variant called the Socially Evaluated Cold Pressor Task (Schwabe, et al. 2008). In this variant, we instructed participants to look directly into a camera that video recorded their facial expression during the hand submersion process. To increase stress further, an experimenter stood behind the participant during the process. All our participants complied with the task. Since stress responses widely vary across individuals, we followed most of the literature on stress (Preston et al. (2007), Van den Bos et al. (2009), Brocas et al. (2017)) and implemented a between-subjects design, with Control (non-CPT) and Stress (CPT) subjects. This method also avoids learning and endowment effects.

The level of physiological stress was assessed by collecting three saliva samples per participant (Base, Peak and End of experiment) in order to measure changes in cortisol levels. We measured cortisol levels of all 170 participants (CPT and non-CPT) using the “passive drool” method, where subjects spit directly into a test tube to a level of at least 2.5 mL per sample. Immediately after collection, samples were stored in a freezer below -20 degrees Celsius. Within seven days of collection, samples were shipped in frozen shipping containers to ZRT laboratories for testing (see <http://zrtlab.com> for details). Results were sent back within one week.

Appendix A2. Interfering task.

	hades	zeus	poseidon	athena	apollo	atlas
physics						
biology						
chemistry						
geology						

Figure 7: In this example, the instructions read: “Please click the cell where the column to the right of the column labeled athena intersects the row above the biology row.” The correct answer required a click on the cell where the 1st row and 5th column intersect.

Appendix A3. Theory.

The payment $\Pi_k(r_k^i)$ for participant i in task k as a function of the report r_k^i is:

$$\begin{aligned}\Pi_B(r_B^i) &= \begin{cases} \max\left\{0, \frac{r_B^i - (1-\alpha)t}{\alpha t} G\right\} & \text{if } r_B^i \leq t \\ \max\left\{0, \frac{(1+\alpha)t - r_B^i}{\alpha t} G\right\} & \text{if } r_B^i \geq t \end{cases} \\ \Pi_U(r_U^i) &= \begin{cases} \max\left\{0, \frac{r_U^i - (1-\alpha)t}{\alpha t} G\right\} & \text{if } r_U^i \leq t \\ 0 & \text{if } r_U^i \geq t \end{cases} \\ \Pi_O(r_O^i) &= \begin{cases} 0 & \text{if } r_O^i \leq t \\ \max\left\{0, \frac{(1+\alpha)t - r_O^i}{\alpha t} G\right\} & \text{if } r_O^i \geq t \end{cases}\end{aligned}$$

as graphically depicted in Figure 1. In the experiment, $G = \$20$ and $\alpha = 0.4$. Needless to say, these functional forms were not presented to subjects. Instead, we explained that they would earn $\$20$ if the report coincided with t and lose 50¢ for each 1% that their report was above or below the announced interval in task B (with the corresponding instructions for tasks U and O).

Proposition 1 (i) $\theta_B^*(t) = t$; (ii) $\theta_U^*(t) \in ((1-\alpha)t, t)$; (iii) $\theta_O^*(t) \in (t, (1+\alpha)t)$; and (iv) $\theta_B^*(t) - \theta_U^*(t) = \theta_O^*(t) - \theta_B^*(t)$.

Proof. For expositional convenience, we drop the announced interval t from the target function θ_k . We also drop the participant's superscript i . We assume risk-neutral profit maximizing individuals, although the results carry over to risk-averse individuals. Let $a \equiv b + \varepsilon$. Given a target θ_k in task k , we know from (4) that the report is:

$$r_k = \theta_k + a \quad \text{where } a \sim F(a)$$

(i) Baseline (B). In task B , the expected payoff of the participant, V_B , is:

$$\begin{aligned}V_B &= \frac{G}{\alpha t} \left[\int_{a=(1-\alpha)t-\theta_B}^{t-\theta_B} \left(a + \theta_B - (1-\alpha)t \right) f(a) da + \int_{a=t-\theta_B}^{(1+\alpha)t-\theta_B} \left((1+\alpha)t - (a + \theta_B) \right) f(a) da \right] \\ \frac{\alpha t}{G} V_B &= \left((a + \theta_B) - (1-\alpha)t \right) F(a) \Big|_{(1-\alpha)t-\theta_B}^{t-\theta_B} - \int_{(1-\alpha)t-\theta_B}^{t-\theta_B} F(a) da \\ &\quad + \left((1+\alpha)t - (a + \theta_B) \right) F(a) \Big|_{t-\theta_B}^{(1+\alpha)t-\theta_B} + \int_{t-\theta_B}^{(1+\alpha)t-\theta_B} F(a) da \\ &= \int_{t-\theta_B}^{(1+\alpha)t-\theta_B} F(a) da - \int_{(1-\alpha)t-\theta_B}^{t-\theta_B} F(a) da\end{aligned}$$

Optimizing over the target θ_B , we can write the first-order condition as:

$$\begin{aligned}\frac{\partial V_B}{\partial \theta_B} \Big|_{\theta_B^*} = 0 &\Leftrightarrow \left[F(t - \theta_B^*) - F((1-\alpha)t - \theta_B^*) \right] - \left[F((1+\alpha)t - \theta_B^*) - F(t - \theta_B^*) \right] = 0 \\ &\Leftrightarrow \int_{(1-\alpha)t-\theta_B^*}^{t-\theta_B^*} f(a) da = \int_{t-\theta_B^*}^{(1+\alpha)t-\theta_B^*} f(a) da\end{aligned}\tag{5}$$

The second-order condition can be written as:

$$\begin{aligned}\frac{\alpha t}{G} \frac{\partial^2 V_B}{\partial (\theta_B)^2} &= - \left[f(t - \theta_B) - f((1 - \alpha)t - \theta_B) \right] + \left[f((1 + \alpha)t - \theta_B) - f(t - \theta_B) \right] \\ &= - \int_{(1 - \alpha)t - \theta_B}^{t - \theta_B} f'(a) da + \int_{t - \theta_B}^{(1 + \alpha)t - \theta_B} f'(a) da\end{aligned}$$

Log-concavity of $f(a)$ implies that $\frac{f'(a)}{f(a)} \leq \frac{f'(t - \theta_B^*)}{f(t - \theta_B^*)}$ for all $a \geq t - \theta_B^*$. Therefore,

$$\begin{aligned}\frac{\alpha t}{G} \frac{\partial^2 V_B}{\partial (\theta_B)^2} \Big|_{\theta_B^*} &< - \int_{(1 - \alpha)t - \theta_B^*}^{t - \theta_B^*} f(a) \frac{f'(t - \theta_B^*)}{f(t - \theta_B^*)} da + \int_{t - \theta_B^*}^{(1 + \alpha)t - \theta_B^*} f(a) \frac{f'(t - \theta_B^*)}{f(t - \theta_B^*)} da \\ &< \frac{f'(t - \theta_B^*)}{f(t - \theta_B^*)} \left[- \int_{(1 - \alpha)t - \theta_B^*}^{t - \theta_B^*} f(a) da + \int_{t - \theta_B^*}^{(1 + \alpha)t - \theta_B^*} f(a) da \right] = 0\end{aligned}$$

so $\theta_B^*(t)$ is unique and a maximum. Finally, from (5) it is trivial to check that:

$$\theta_B^*(t) = t$$

since by the symmetry of $f(a)$, we have:

$$\int_{-\alpha t}^0 f(a) da = \int_0^{\alpha t} f(a) da$$

(ii) Under-report (U). In task U , the expected payoff of the participant, V_U , is:

$$\begin{aligned}\frac{\alpha t}{G} V_U &= \int_{a=(1 - \alpha)t - \theta_U}^{t - \theta_U} \left(a + \theta_U - (1 - \alpha)t \right) f(a) da \\ &= \left((a + \theta_U) - (1 - \alpha)t \right) F(a) \Big|_{(1 - \alpha)t - \theta_U}^{t - \theta_U} - \int_{(1 - \alpha)t - \theta_U}^{t - \theta_U} F(a) da \\ &= \alpha t F(t - \theta_U) - \int_{(1 - \alpha)t - \theta_U}^{t - \theta_U} F(a) da\end{aligned}$$

Optimizing over the target θ_U , we can write the first-order condition as:

$$\begin{aligned}\frac{\partial V_U}{\partial \theta_U} \Big|_{\theta_U^*(t)} = 0 &\Leftrightarrow F(t - \theta_U^*(t)) - F((1 - \alpha)t - \theta_U^*(t)) - \alpha t f(t - \theta_U^*(t)) = 0 \quad (6) \\ &\Leftrightarrow \int_{(1 - \alpha)t - \theta_U^*}^{t - \theta_U^*} [f(a) - f(t - \theta_U^*)] da = 0\end{aligned}$$

The second-order condition can be written as:

$$\begin{aligned}\frac{\alpha t}{G} \frac{\partial^2 V_U}{\partial (\theta_U)^2} &= -f(t - \theta_U) + f((1 - \alpha)t - \theta_U) + \alpha t f'(t - \theta_U) \\ &= \int_{(1 - \alpha)t - \theta_U}^{t - \theta_U} [f'(t - \theta_U) - f'(a)] da\end{aligned}$$

Again, log-concavity of $f(a)$ implies that $\frac{f'(a)}{f(a)} > \frac{f'(t-\theta_U^*)}{f(t-\theta_U^*)}$ for all $a < t - \theta_U^*$. Therefore,

$$\begin{aligned} \frac{\alpha t}{G} \frac{\partial^2 V_U}{\partial(\theta_U)^2} \Big|_{\theta_U^*} &< \int_{(1-\alpha)t-\theta_U^*}^{t-\theta_U^*} \left[f'(t-\theta_U^*) - \frac{f'(t-\theta_U^*)}{f(t-\theta_U^*)} f(a) \right] da \\ &< \frac{f'(t-\theta_U^*)}{f(t-\theta_U^*)} \int_{(1-\alpha)t-\theta_U^*}^{t-\theta_U^*} [f(t-\theta_U^*) - f(a)] da = 0 \end{aligned}$$

Therefore $\theta_U^*(t)$ is unique and a maximum. Finally, from (6) it is trivial to check that:

$$\theta_U^*(t) \in ((1-\alpha)t, t)$$

since:

$$\frac{\alpha t}{G} \frac{\partial V_U}{\partial \theta_U} \Big|_{(1-\alpha)t} = \int_0^{\alpha t} [f(a) - f(\alpha t)] da > 0 \quad \text{and} \quad \frac{\alpha t}{G} \frac{\partial V_U}{\partial \theta_U} \Big|_t = \int_{-\alpha t}^0 [f(a) - f(0)] da < 0$$

(iii) Over-report (O). In task O , the expected payoff of the participant, V_O , is:

$$\frac{\alpha t}{G} V_O = \int_{a=t-\theta_O}^{(1+\alpha)t-\theta_O} F(a) da - \alpha t F(t-\theta_O)$$

Optimizing over the target θ_O , we can write the first-order condition as:

$$\begin{aligned} \frac{\partial V_O}{\partial \theta_O} \Big|_{\theta_O^*(t)} = 0 &\Leftrightarrow -F((1+\alpha)t - \theta_O^*(t)) + F(t - \theta_O^*(t)) + \alpha t f(t - \theta_O^*(t)) = 0 \quad (7) \\ &\Leftrightarrow \int_{t-\theta_O^*}^{(1+\alpha)t-\theta_O^*} [f(t-\theta_O^*) - f(a)] da = 0 \end{aligned}$$

The second-order condition can be written as:

$$\begin{aligned} \frac{\alpha t}{G} \frac{\partial^2 V_O}{\partial(\theta_O)^2} &= f((1+\alpha)t - \theta_O) - f(t - \theta_O) - \alpha t f'(t - \theta_O) \\ &= \int_{t-\theta_O}^{(1+\alpha)t-\theta_O} [f'(a) - f'(t - \theta_O)] da \end{aligned}$$

Log-concavity of $f(a)$ implies that $\frac{f'(a)}{f(a)} < \frac{f'(t-\theta_O^*)}{f(t-\theta_O^*)}$ for all $a > t - \theta_O^*$. Therefore,

$$\begin{aligned} \frac{\alpha t}{G} \frac{\partial^2 V_O}{\partial(\theta_O)^2} \Big|_{\theta_O^*} &< \int_{t-\theta_O^*}^{(1+\alpha)t-\theta_O^*} \left[\frac{f'(t-\theta_O^*)}{f(t-\theta_O^*)} f(a) - f'(t-\theta_O^*) \right] da \\ &< \frac{f'(t-\theta_O^*)}{f(t-\theta_O^*)} \int_{t-\theta_O^*}^{(1+\alpha)t-\theta_O^*} [f(a) - f(t-\theta_O^*)] da = 0 \end{aligned}$$

Therefore $\theta_O^*(t)$ is unique and a maximum. Finally, from (7) it is trivial to check that:

$$\theta_O^*(t) \in (t, (1+\alpha)t)$$

since:

$$\frac{\alpha t}{G} \frac{\partial V_O}{\partial \theta_O} \Big|_t = \int_0^{\alpha t} [f(0) - f(a)] da > 0 \quad \text{and} \quad \frac{\alpha t}{G} \frac{\partial V_O}{\partial \theta_O} \Big|_{(1+\alpha)t} = \int_{-\alpha t}^0 [f(-\alpha t) - f(a)] da < 0$$

(iv) Symmetry. Immediate once we notice that if we substitute $t - \theta_U^*$ with $\theta_O^* - t$ in (6), we obtain (7). \blacksquare

Prediction 2 is a consequence of the fact that, by construction, $E[b^i] = 0$. Then, by Proposition 1(i), $E[r_B^i] = \theta_B^* = t$. Prediction 3 and Prediction 4 are immediate using similar arguments and given $\theta_O^* - \theta_B^* = \theta_B^* - \theta_U^* > 0$.

Appendix A4. Calibration.

To calibrate the optimal targets in U and O , we impose $b^i \sim N(0, \sigma^2)$. This means that:

$$a^i \equiv b^i + \varepsilon^i \sim N(0, z_i^2) \quad \text{where} \quad z_i^2 = \sigma^2 + \eta_i^2$$

Notice that $E[b^i] = 0$, $g(b^i) = \frac{1}{\sigma} \varphi\left(\frac{b^i}{\sigma}\right) = g(-b^i)$, and $f(a^i)$ is log-concave since the Normal distribution is log-concave, so the assumptions of the model are satisfied.

From Proposition 1(i) we know that $\theta_B^{i*} = t$. Replacing $F(a)$ with $\Phi\left(\frac{a}{z_i}\right)$ in (6) and (7), we obtain that θ_U^{i*} and θ_O^{i*} solve the following first-order conditions:

$$\Phi\left(\frac{t - \theta_U^{i*}}{z_i}\right) - \Phi\left(\frac{(1 - \alpha)t - \theta_U^{i*}}{z_i}\right) - \frac{\alpha t}{z_i} \phi\left(\frac{t - \theta_U^{i*}}{z_i}\right) = 0 \quad (8)$$

$$-\Phi\left(\frac{(1 + \alpha)t - \theta_O^{i*}}{z_i}\right) + \Phi\left(\frac{t - \theta_O^{i*}}{z_i}\right) + \frac{\alpha t}{z_i} \phi\left(\frac{t - \theta_O^{i*}}{z_i}\right) = 0 \quad (9)$$

From the data in task B we can estimate $\hat{\sigma}$, the standard deviation of the bias in the population. For each participant i in task B , we can also estimate $\hat{\eta}_i$, the standard deviation in the noise of his report. In our data, we have $\hat{\sigma} = 12.16$ and $\hat{\eta}_i \in [0.61, 21.14]$. Therefore, $\hat{z}_i \in [12.77, 33.3]$. Inserting each value of \hat{z}_i in (8) and (9), and setting $\alpha = 0.4$ and $\bar{t} = \sum t/10$, we get $\theta_U^{i*} \in [30.86, 30.92]$ and $\theta_O^{i*} \in [40.28, 40.34]$. Averaging over all individuals, we finally obtain $\theta_U^* = 30.88$ and $\theta_O^* = 40.32$. It is key to realize that despite the significant variability in the variance of noise across participants, optimal targets are extremely similar to each other. It is therefore with virtually no loss of generality that we can focus on the optimal average targets θ_U^* and θ_O^* .

Appendix A5. Time estimation in task B by interval.

Traditional theories of subjective time rely on two properties. The ‘‘mean accuracy’’ property postulates that people produce time intervals that are on average equal to the interval they are required to produce. The ‘‘scalar property of variance’’ requires that the sensitivity of time estimates is independent of the time to estimate (the variability to time ratio is constant). We present below the p.d.f. of reports for each announced interval t (vertical line) in task B . We computed separately the distribution for CPT (blue) and non-CPT (green) participants. These distributions were not significantly different for any interval (KS-test, p-value $\gg 0.05$). Average reports in all intervals were significantly above the announced time t (t-tests, p-value < 0.006) implying that mean accuracy was not satisfied by our data. We also tested for the scalar property of variance by

regressing the standard deviation on the mean of the reports. The R-squared of this regression was 0.55. We also computed the coefficient of variation (standard deviation / mean) for each announced time interval and regressed it on time. The slope coefficient was not significantly different from zero (p-value = 0.067). These findings indicate that the scalar property of variance was satisfied.

