Hidden inefficiency: strategic inflation of project schedules

Lorko, Matej and Servátka, Maroš and Zhang, Le

Macquarie University, Sydney, Australia, University of Economics in Bratislava, Bratislava, Slavakia

21 December 2022

Hidden inefficiency: Strategic inflation of project schedules

Matej Lorko
MQBS Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia
University of Economics in Bratislava, Slovakia
matej.lorko@gmail.com
Maroš Servátka
MQBS Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia
University of Economics in Bratislava, Slovakia
maros.servatka@mgsm.edu.au
Le Zhang
MQBS Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia
lyla.zhang@mq.edu.au

October 15, 2022

Abstract: Establishing realistic project plans and completing the resulting business projects on schedule is crucial for organizations striving to effectively utilize their resources. However, incentivizing on-time project delivery may result in moral hazard, as people could respond to estimation accuracy incentives by strategically inflating duration estimates and subsequently prolonging project execution. While the project is delivered on time, the resources are underutilized. We conjecture that the possibility of moral hazard can be mitigated by introducing speed incentives in addition to the schedule accuracy incentives. We conduct a diagnostic test of the effect of accuracy and speed incentives on the process of project estimation and delivery. Our study presents direct empirical evidence that the incentive structure rewarding solely the estimation accuracy can result in hidden inefficiency due to inflated estimates and deliberately slower project execution. However, when speed incentives are implemented alongside estimation accuracy incentives, the estimates are significantly lower and the project is completed more quickly, without compromising the schedule accuracy or output quality. Aligning the objectives of a project owner with those of planners, by incentivizing the planners for both estimation accuracy and quick project completion, fosters more compressed but still accurate and reliable project schedules, and accelerated project delivery.

Keywords: project management, project planning, time management, duration estimation, moral hazard

JEL codes: C91, D82, D83, O21, O22

1. Introduction

Accurate schedules are central to ensuring effective utilization of resources in a business project. Underestimating the time necessary to complete a project leads to schedule overruns, often associated with cost overruns and customer dissatisfaction, while overestimating it gives rise to opportunity costs stemming from misallocation and/or underutilization of resources. Accurate project schedules are especially crucial when managing a project portfolio, in which resources are assigned to individual projects temporarily. Arguably, organizations prefer to have projects delivered not only on time, but also as quickly as possible, to promptly collect returns on investments from the current project(s) and engage in new ones (Grushka-Cockayne et al., 2018). Widely used project management methodologies (IPMA, 2015; Project Management Institute, 2013) therefore assume that project planners contribute towards the operational efficiency by proposing schedules that are both realistic and compressed in length.

However, operational efficiency may not always be the primary concern of project planners. At times, circumstances may lead them to intentionally underestimate the project duration, for example to seemingly fit the project into a constrained timeframe and increase the chances of being awarded a contract (akin to deliberate cost underestimation reported in Flyvbjerg, Holm, & Buhl, 2002).[1] Conversely, strategic considerations may lead planners to deliberately overestimate the project duration to gain more time for execution. Thus, instead of the planner's true belief about the amount of time required to deliver the project, the estimated duration may capture the maximum amount believed to be acceptable by the management or customer.

In this paper, we provide scientific evidence of deliberate overestimation induced by incentives for accurate schedules. We argue that this type of overestimation commonly occurs when the planner both estimates the project duration and executes the project, as forecasting own future performance allows for moral hazard. Such scenarios typically arise in fixed-price contracts, where the contractor first submits a proposal that includes the project schedule and if the proposal is accepted, the contractor is bound by the contract to deliver the project on time (Bajari & Tadelis, 2001; von Branconi & Loch, 2004). A similar process, albeit less formal, occurs within organizations where employee(s) responsible for project estimation and delivery are strongly encouraged or explicitly incentivized to deliver the project in accordance with the proposed schedule.

[1] A prominent recent example of manipulation and misinforming about the schedule and budget estimates is the failed multibillion project to build two nuclear reactors in South Carolina.

We conjecture that when planners are incentivized solely for the accuracy of their schedules, they deliberately inflate them and then strategically prolong the execution to match the project completion date with the scheduled deadline. We further conjecture that the morally hazardous behavior can be mitigated by adding incentives to complete the project quickly (henceforth speed incentives), therefore aligning the objectives of the planner with those of an organization (or external customer).

To illustrate, consider a scenario within an organization where an employee responsible for project planning and execution typically knows more than his executives about the involved tasks and their likely duration (Chao, Lichtendahl, & Grushka-Cockayne, 2014). This information asymmetry makes it challenging for the executives to assess whether the proposed project schedule is adequate or not. If the employee is incentivized only for on-time project delivery (or if his compensation is independent of his performance on the project), he may benefit from inflating the project duration. Apart from increasing one's chances of delivering the project on time, inflated estimates can reduce stress caused by time pressure (Cahlíková, Cingl, & Levely, 2020) or secure time to allow for potential procrastination (Akerlof, 1991; Goldratt, 1997; Knowles, Servátka, Sullivan, & Genç, 2022; O'Donoghue & Rabin, 2008). However, from the perspective of the organization, inflated estimates may lead to prolonged project execution without commensurate improvement in the quality of deliverables.[2]

Strategic inflation of project schedules and the resulting inefficiencies (e.g., unnecessary contractor compensation or wasted internal resources) are difficult to detect and measure using happenstance business data. Irrespective of the project duration being inflated deliberately or by accident, the working pace can be slowed down so that all allocated time is used up. Thus, when a project is completed on time, it is unclear whether its schedule was genuinely accurate, or the progress was adjusted to match an inflated schedule. In the latter case, the project delivered "on time" is in fact delivered later than what was feasible. Since projects delivered on time are usually considered successful, they rarely spark suspicion without which it may be difficult to uncover the morally hazardous behavior. The deliberate inflation of project schedules and subsequent prolonged execution thus present a hidden inefficiency for organizations.

[2] A similar scenario arises when a monopoly provider purposefully inflates estimates in the project schedule to gain more time (and/or funding). The first author previously worked as a project manager in a large corporation where he noticed that his contractors (especially those who had no direct competition in the local market) usually delivered the requested outcomes precisely at the agreed milestones. Given the lack of proper verifiability of the time these contractors actually spent on project tasks, it is possible that they proposed inflated schedules and overspent time (and funds) on the projects.

Due to the unobservability of our conjectured effects in the happenstance business data, we investigate our conjectures in a controlled laboratory environment that allows us to turn on and off specific incentives and create counterfactuals. Identifying whether and how individual planners respond to commonly imposed incentive structures in organizations is the first step in resolving the hidden inefficiency.[3]

In our experiment, subjects are asked to estimate how long it will take them to complete a real effort task. Upon providing a duration estimate, they execute the task. A crucial feature of our experimental design is the ability to unambiguously detect whether individuals deliberately decelerate their pace (or outright wait) towards the end of the project if their current pace would result in finishing ahead of their estimate, effectively making our study a diagnostic test. An additional benefit of a stylized experiment is that it allows us to analyze duration estimation in isolation and thus to eliminate potential confounds related to project cost, risks, and unforeseen events that could affect the behavior of planners in the field in an uncontrolled manner. We also control for the scope (quality) of the work by holding the output of all subjects identical.

The experiment consists of two main treatments and two controls. In both of our main treatments, subjects are incentivized for their estimation accuracy, meaning that the more accurate their estimates are, the more money they earn. We then vary whether speed incentives are also present or not. The control treatments are designed to isolate the effect of speed incentives and generate a baseline with the shortest task duration. One control treatment thus features only speed incentives and the other a flat payment (and thus no speed or accuracy incentives). To parallel business practice, in the experiment we establish an environment in which the conjectured strategic behavior (manipulating the task progress to match the elicited estimate) is feasible. We do so by providing subjects with a time measuring tool that enables them to monitor the time already spent on the task. The tool effectively allows subjects to control when exactly to complete the task and enables us to collect data on whether and how often subjects check the elapsed time.

In line with our conjectures, our data show when only the estimation accuracy is incentivized (i.e., in the absence of speed incentives) subjects provide higher estimates and prolong the task execution so that their actual task duration matches their estimate. When both speed and accuracy incentives are in place, subjects provide lower duration estimates and work faster, while preserving the estimation

[3] Conditional on detecting morally hazardous behavior of planners and thus providing a proof of concept, the next step would be to explore the trade-off between increased remuneration costs due to speed incentives and the costs of prolonged project execution.

accuracy and quality of work. Although we observe inflation of estimates and strategic work pacing under both incentive structures, the behavior is more pronounced when subjects are incentivized only for the accuracy of their estimates. Finally, if only the speed is incentivized, subjects complete the task the fastest but underestimate task duration.

Our experimental results yield the following managerial implications. Solely incentivizing on-time project delivery results in a non-negligible efficiency loss (e.g., resource underutilization) due to inflated schedules and prolonged project progress. If it is in the best interest of an organization to have the project completed both on time and quickly, managers should carefully consider incorporating speed incentives. The incentives should encourage planners to be efficient, rather than waste time (and other resources) only to deliver the project exactly "on time". A reasonably reliable benchmark to reward fast performance can be based on historical information, for example the average duration of completed similar projects in the past, as demonstrated by Lorko, Servátka, & Zhang (2021). This benchmark can effectively complement the proposed project schedule, alleviating the emergence of inflated estimates and deliberately slow project execution.

2. Relationship to the literature

In addressing our research question, we contribute to the larger empirical literature in economics, management, and psychology exploring the determinants of effective time estimation (see Halkjelsvik and Jørgensen, 2012, for a comprehensive review). Interestingly, although the existing studies usually focus on how accurately a given project or task is estimated, the experiments rarely incentivize subjects for their estimation accuracy. The lack of motivation to accurately estimate the duration creates a potential issue for establishing a causal link between the studied factor and the observed accuracy.

A notable exception, Buehler, Griffin, & MacDonald (1997) explore how monetary incentives affect optimism bias (resulting in underestimation of time) in completing a series of anagram-like word puzzles. They find that estimation accuracy incentives induce overestimation, while speed incentives lead to underestimation, and that subjects exhibit the smallest estimation bias when incentivized for both the speed and accuracy.[4] The authors argue that speed incentives increase the optimism bias in duration estimation, while the accuracy incentives reduce the bias. However, there exists an alternative explanation of the observed results, consistent with the central conjecture of our study. If subjects are incentivized only for their estimation accuracy, they have no urge to complete the task

[4] Lorko, Servátka, & Zhang (2019) also incentivize subjects simultaneously for speed and accuracy and report unbiased duration estimates in their control treatment, especially after subjects acquire task experience.

quickly. As such, they may behave strategically by deliberately inflating their estimates and then pacing their work to ensure that the estimate is accurate. In fact, subjects in Buehler et al. (1997) are fastest when incentivized only for speed and slowest (albeit non-significantly) when incentivized only for estimation accuracy. However, since subjects are not allowed to monitor time, it is difficult for them to know at which particular moment it is the most beneficial to finish the task. The study therefore cannot uncover the strategic behavior.

In our study, we directly examine such strategic behavior by providing subjects with a time measuring tool throughout the task execution. In addition, we implement a continuous incentive structure in which every second of task execution matters, rather than a simple cutoff rule.[5] Also importantly, our experimental design allows us to explore subsequent strategic behavior during the execution stage. Namely, we observe whether subjects spend more time than necessary and thus prolong the progress to ensure that the task is completed close to or right at the estimated time. Finally, to enhance subjects' understanding of the incentive structures, we implement repeated task estimation and execution. We are thus able to test for an alternative explanation of the findings by Buehler et al. (1997). In light of our conjecture, the overestimation in their accuracy-only treatment could be caused by deliberately inflating estimates (instead of reduced optimism), followed by inadequately adjusting the working pace due to limited control over the elapsed time. In fact, studies by Ariely & Wertenbroch (2002) and Buehler, Griffin, & Ross (1994) show that when people can monitor time, they adjust the work pace in accordance with the underlying incentives. In Ariely & Wertenbroch (2002) subjects self-impose deadlines that they rarely miss. In Buehler et al. (1994) subjects provide a non-binding estimate and are also given a fixed deadline to complete the task. While most subjects take longer to complete the task than they estimated, they usually do complete the task within the hard deadline.

The strategy of inflating the duration estimates is analogous to misrepresenting estimates in budgeting (see Covaleski, Evans, Luft, & Shields, 2006 for theoretical perspectives and Brown, Evans, & Moser, 2009 for a review of experimental studies). The budgeting research generally shows that a planner often creates a budgetary slack, defined by Dunk & Nouri (1998, p.73) as an "intentional overestimation of costs and resources required to complete a budgeted task," when opportunity arises, especially under information asymmetry. Another example of strategic misrepresentation in project planning is reported in Lederer et al. (1990) and Magazinius, Börjesson, & Feldt (2012) who interview software managers. Both studies find that project cost estimates are not always based purely on the outcomes of the planning process, but often driven by personal considerations, such as

[5] Buehler et al. (1997) accuracy incentives yield $2 for an estimate falling within one minute of the actual task duration, and $4 for falling within 30 seconds. The speed incentives yield $2 for finishing one minute faster and $4 for finishing two minutes faster than in the practice trial.

the fear of cost overruns (resulting in deliberate overrepresentation) or project cancellation (resulting in deliberate underrepresentation). Relatedly, Flyvbjerg, Holm, & Buhl (2002) provide evidence of strategic misrepresentation of cost estimates in public projects. Instead of realistic estimates based on a project plan, companies often produce project estimates that capture maximum thresholds that are still acceptable by project stakeholders, especially the customers.

Some evidence of strategic overrepresentation can also be found in the literature on duration estimation. For example, Shepperd, Sweeny, & Cherry (2007) investigate socially motivated incentives and find systematic overestimation of waiting time in restaurants provided by hostesses as well as in customer-care hotlines. Customers are intentionally misinformed in order to reduce or avoid their disappointment from waiting longer than expected as the costs of underestimation are believed to be larger than the costs of overestimation. Similar asymmetry of consequences is conjectured to be a driver of the frequent overestimation of computation jobs running times (Lee, Schwartzman, Hardy, & Snavely, 2005). Finally, a recent series of papers (Yimga, 2020, 2021; Zhang, Salant, & Van Mieghem, 2018) empirically documents schedule padding in airline industry.

In this study, we create a controlled environment in which we exogenously manipulate incentives while eliminating confounding factors that are often present in everyday business environment. We contribute to the time estimation and project planning literature by cleanly identifying the effects of accuracy and speed incentives on not only estimation, but also on the subsequent task (project) execution.

3. Experimental design

Our experiment is designed to test our conjectures that (1) an incentive structure aimed at increasing the estimation accuracy triggers strategic inflation of estimates and subsequent slower task execution; and that (2) adding speed incentives alongside the accuracy incentives results in lower task duration estimates and faster task execution without compromising the estimation accuracy. Furthermore, the addition of two control treatments enables us to isolate the net effect of speed incentives. In total, the experiment consists of four treatments manipulating the incentive structure. The treatments are implemented in an across-subject design.

In the main treatments (ABA and BAB) we employ two incentive structures, A and B, where A stands for **A**ccuracy-only incentives and B for **B**oth accuracy and speed incentives. In the ABA treatment, subjects execute the task under the incentive structure A in Round 1, under B in Round 2, and again under A in Round 3. In the BAB treatment, subjects execute the task under the incentive structure B in Round 1, under A in Round 2, and again under B in Round 3.[6]

The NNN control treatment involves **N**o speed or accuracy incentives as subjects are paid a flat fee. In the SSS control treatment, subjects are incentivized only for their **S**peed, meaning that the faster they complete the task, the more money they earn. The SSS treatment generates a proxy for the shortest attainable task duration, and thus allows us to measure the effect of estimation accuracy incentives in the main ABA and BAB treatments.

The task

In each of the three rounds of the experiment, subjects estimate how long it will take them to complete an individual real-effort task. After providing the estimates, subjects proceed to executing the task. In the task, subjects are shown a series of five tables, one at a time. Each table consists of 100 cells that contain either the letter "S" or the number "5". In each table, there are between 45 and 55 cells containing the letter "S", while the rest of the cells contain the number "5". A sample table is presented in Figure 1.

Figure 1: A sample table used in the experiment

☐ 5	☐ S	☐ S	☐ 5	☐ S	☐ 5	☐ S	☐ S	☐ 5	☐ S
☐ 5	☐ 5	☐ 5	☐ S	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5
☐ S	☐ S	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ S	☐ S	☐ S
☐ S	☐ S	☐ 5	☐ S	☐ 5	☐ S	☐ S	☐ S	☐ S	☐ 5
☐ S	☐ S	☐ S	☐ S	☐ 5	☐ 5	☐ 5	☐ S	☐ S	☐ S
☐ 5	☐ S	☐ S	☐ S	☐ 5	☐ 5	☐ 5	☐ S	☐ S	☐ S
☐ 5	☐ 5	☐ 5	☐ S	☐ S	☐ S	☐ S	☐ S	☐ 5	☐ 5
☐ 5	☐ 5	☐ 5	☐ S	☐ S	☐ S	☐ S	☐ S	☐ 5	☐ S
☐ 5	☐ S	☐ S	☐ 5	☐ 5	☐ S	☐ S	☐ S	☐ 5	☐ S
☐ 5	☐ 5	☐ S	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5	☐ 5

In order to solve each table, subjects have to check all cells containing the letter "S", while leaving the other cells unchecked, and then submit the table for verification. They cannot move to the next table

[6] The ABA/BAB design also allows for a within-subject comparison as it identifies the change in behavior *of the same person* in response to a change in the incentive structure (from Round 1 to Round 2 and from Round 2 to Round 3). Following a reviewer's suggestion to add control treatments to isolate the impact of individual incentives, which pivoted our narrative, we have removed the within-subject comparison results from the paper. They are available upon request.

unless they check all cells correctly in the current one. There is no limit on how many times each table can be submitted. The software does not uncheck the cells after an incorrect or incomplete submission.

Every subject is given the same sequence of tables, but all tables seen by the same subject are different from each other. Since the tables are accepted by the software as solved only if they are completed correctly, the quality of the output is kept constant for every subject.[7] Time to finish the task is therefore an unambiguous measure of performance.

Strategic behavior requires that subjects thoroughly understand the nature of the task. Before the first estimation, we therefore have the subjects solve one practice table that is not payoff-relevant. Incorporating the practice table into the experimental design mitigates estimating biases stemming from inadequate experience with the task and reduces the variance in duration estimates, effectively minimizing the type-I error. Although we do not provide subjects with any information regarding how much time they spent on the practice table, subjects presumably can acquire a rough idea about the duration.[8] Having solved the practice table, subjects estimate (in minutes and seconds) how long it will take them to solve the task in the first round (time to solve five payoff-relevant tables altogether) and then work through the tables, one by one. The estimation and task execution follow the same procedures in the second and third round.

Subjects cannot behave strategically if they are not able to monitor time, because the perception of actual task duration for tasks that last several minutes without any devices that measure time tends to be rather inaccurate (Lorko, Servátka, & Zhang, 2019; Roy & Christenfeld, 2007). We therefore provide subjects with a time measuring tool on their screen. The tool measures how much time a subject has already spent on the task. The information regarding the elapsed time is updated every time subject solves the current table and moves to the next table. At any time, the subject can also update the information manually by clicking on the "Update" button. The tool ensures that every subject has the same opportunity to monitor time, which might not be the case if people do not have watches or a phone on them when they come to the laboratory. We choose to provide such tool

[7] While the final output is the same for every subject because of the imposed quality standard, we still can (and do) analyze the quality of work by comparing the number of incorrect table submissions.

[8] If we allowed subjects to measure the time (e.g., using our tool) while solving the practice table, it would increase the tool's salience for the main task, possibly resulting in the experimenter demand effect (Zizzo, 2010). Importantly, the same procedure of not having the tool available when solving the practice table is being kept constant across treatments and hence it would not affect the validity of our results.

instead of a clock or a timer, because it yields data of whether, when, and how often each subject checks the elapsed time, necessary to detect strategic behavior.

Incentives

Table 1 presents the implemented incentive structures. Estimation accuracy earnings are determined by the absolute difference between the actual task duration and the estimate. The maximum earnings from a point-precise estimate are AUD 20. The accuracy earnings decrease linearly, by 6 dollars for every minute (10 cents for every second) the estimate is away from the actual task duration, as shown in Equation (1).[9] The relatively wide interval of positive estimation accuracy earnings allows all subjects to have a reasonable chance to earn money. At the same time, we implement a sharp penalty for every second of inaccuracy to motivate subjects to be as accurate in estimation as possible. The implemented accuracy incentives conservatively feature an equal penalization in both directions instead of, say, a heavier penalty for being late.[10] We do not allow for negative estimation accuracy earnings. If the difference between the actual and estimated time in either direction exceeds 200 seconds, the estimation accuracy earnings are set to zero.

$$\text{Estimation acuracy earnings } = 20 - \ 0.10 * |\text{actual time in seconds } - \text{estimated time in seconds}| \qquad (1)$$

Performance speed earnings are based on how quickly subjects finish the task. These earnings depend only on the actual duration of the task as shown in Equation (2). The shorter the duration, the higher the earnings. Based on the initial testing, we expected subjects to complete the task on average in 5 minutes (300 seconds) and earn AUD 10 for their performance speed.

$$\text{Performance speed earnings } = \frac{3000}{\text{actual time in seconds}} \qquad (2)$$

An important feature of our design is that although subjects face two sets of incentives in B, the best strategy is to focus primarily on the accuracy of estimates (just as in A). It is because the speed earnings decline exponentially, while the estimation accuracy earnings are linear. From the earnings functions of accuracy and speed, one can work out analytically that the accuracy incentives become stronger than speed incentives 173 seconds into the task execution. To be more precise, after 173 seconds

[9] Although the linear scoring rule might not be the most incentive compatible one, it has an advantage of easier comprehension (Woods & Servátka, 2016). We find it more practical to implement in an experimental environment than more complex scoring rules (e.g., quadratic or logarithmic).

[10] A heavier penalty for being late than for being early (as often implemented in business practice) could increase the propensity to inflate estimates in order to ensure that one is able to finish within the estimated time. Our symmetrical penalization is designed to pick up the lower bound of the effect of accuracy incentives on inflating duration estimates.

from the start of the task, finishing the task one second later yields less than 10 cents decrease in speed earnings. On the other hand, the estimation accuracy earnings change by 10 cents per second whenever they are positive. Thus, if a subject cannot finish the task within 173 seconds, he is best off by maximizing his accuracy earnings and collecting residual speed earnings. Based on the results from a pilot session, we had not expected any subject to finish the task in 173 seconds or less, which was indeed the case, as the fastest recorded round in the entire experiment was 198 seconds. Since we expect subjects to maximize their estimation accuracy earnings (AUD 20) and the expected average speed earnings are AUD 10, we set flat fees in the SSS treatment (AUD 10) and the NNN (AUD 20) treatment in a way that the expected earnings from those treatments are equal to expected estimation accuracy earnings.

Table 1: Incentive structures across treatments and rounds

	Incentives			
	Round 1	Round 2	Round 3	Flat fee
Treatment ABA	Accuracy (A)	Accuracy + Speed (B)	Accuracy (A)	None
Treatment BAB	Accuracy + Speed (B)	Accuracy (A)	Accuracy + Speed (B)	None
Treatment SSS	Speed (S)	Speed (S)	Speed (S)	AUD 10
Treatment NNN	None (N)	None (N)	None (N)	AUD 20

Procedures

The study was conducted in the MGSM Vernon L. Smith Experimental Economics Laboratory at Macquarie Business School in Sydney. Subjects (mostly undergraduate business majors and MBAs with no prior experience with laboratory experiments on duration estimation) were recruited using the online database system ORSEE (Greiner, 2015). The experimental software was programmed in zTree (Fischbacher, 2007).

Subjects, seated in individual cubicles, were given the instructions (provided in the appendix) that described the experimental task, the duration estimation, the incentive structure and the time measuring tool. After reading the instructions, subjects were given a few minutes to privately ask questions regarding the experiment. Once all questions were answered by the experimenter (also privately), the experiment proceeded with the practice table and the decision-making part.

To ensure a thorough understanding of the incentive structures, before each round subjects were asked seven control questions related to how much they could earn from estimation accuracy earnings and/or performance speed earnings (except for the NNN treatment). Subjects were not allowed to proceed to estimating until they answered all questions correctly. After subjects submitted their estimates, we asked them how many tables their estimates were referring to. If the answer was anything other than five, the subject was reminded that the task consisted of five tables and prompted to re-estimate, after which we asked this control question again. This procedure was implemented to mitigate possible errors from not paying attention to the instructions and estimating, say, how long it would take to complete one table instead of five.

Upon completing each round, subjects received feedback reminding them of their estimate and how much time they actually spent on the task. They were also informed about their earnings for the round. After completing all three rounds, subjects participated in an incentivized risk assessment (Holt & Laury, 2002) and incentivized three-item cognitive reflection test (CRT; Frederick, 2005) in which they could earn AUD 0.50 for every correct answer. The cognitive reflection test enables us to verify whether subjects with a higher CRT score are more responsive to incentive structures, especially in the ABA and BAB treatments. Finally, subjects filled out a demographical questionnaire.

Once everybody finished the experiment (subjects were not allowed to leave the laboratory earlier), subjects privately and individually received their experimental earnings in cash in the control room at the back of the laboratory. In ABA, BAB, and SSS treatments, one out of three rounds was randomly (with the same probability) selected for payment, independently for every subject. This payoff protocol, announced in the instructions, avoids cross-contamination of incentives across rounds and controls for the wealth and portfolio effects. Thus, by paying one round randomly, we effectively make subjects consider every round as an independent chance to earn money. The implemented isolation of decisions allows for a crisper identification of our conjectured effects.

4. Hypotheses

In this section, we derive theoretical predictions for estimation and performance speed in each treatment. The incentive structure rewarding only speed motivates subjects to work as quickly as possible. On the other hand, the incentive structure rewarding estimation accuracy without incentivizing speed is slack-inducing, as it motivates subjects to finish the task exactly at their estimate, which is arguably easier if the estimate is inflated. Inflating the estimate mitigates the risk of not being able to finish on time, for example because of missing one or more letters “S” and having to systematically go through the entire table again. The incentive structure combining both accuracy and speed incentives motivates accurate estimation, but at the same time incentivizes subjects to finish the task fast.

The formalized intuition is as follows. Across all treatments, it is costly for subjects to exert performance speed effort $e_s \in [0, +\infty]$ and estimation effort $e_e \in [0, +\infty]$, respectively. We assume that the cost functions are convex and the performance functions (speed f and accuracy g) are concave in effort e_s and e_e. Subjects can choose to delay d (i.e., slow down) their performance speed if necessary, $d \in [0, d_{max}]$. We further assume that participation constraints are satisfied, otherwise subjects would not participate in the experiment. F denotes the fixed payment.

1. No speed or accuracy incentives (N)

Subjects choose $e_{e,}\, e_s$ to $\max F - c(e_e) - c(e_s)$. Given $\frac{dc(e_e)}{de_e} > 0$ and $\frac{dc(e_s)}{de_s} > 0$, subjects will choose $e_e = 0$ and $e_s = 0$.

2. Only Speed incentives (S)

Subjects choose $e_{e,}\, e_s, d$ to $\max F + f(e_s, d) - c(e_e) - c(e_s)$. Given $\frac{dc(e_e)}{de_e} > 0$ and $\frac{dc(e_s)}{de_s} > 0$, $\frac{\partial^2 f(e_s,d)}{\partial e_s{}^2} - \frac{d^2 c(e_s)}{de_s{}^2} < 0$, subjects will choose $e_e = 0$ and $e_s > 0$, so that $\frac{\partial f(e_s,d)}{\partial e_s} - \frac{dc(e_s)}{de_s} = 0$. Given $\frac{\partial f(e_s+d)}{\partial d} < 0$, they will choose $d = 0$.

3. Only Accuracy incentives (A)

Subjects choose $e_{e,}\, e_s, d$ to $\max F + g(e_e, d) - c(e_e) - c(e_s)$. To maximize their profit, given $\frac{dc(e_e)}{de_e} > 0$, $\frac{dc(e_s)}{de_s} > 0$, and $\frac{\partial^2 g(e_e,d)}{\partial e_e{}^2} - \frac{d^2 c(e_e)}{de_e{}^2} < 0$, subjects will choose $e_s = 0$ and $e_e > 0$, so that $\frac{\partial g(e_e,d)}{\partial e_e} = \frac{dc(e_e)}{de_e} > 0$.

The possibility of delay makes the estimates more accurate by deliberately prolonging the task duration, hence $\frac{\partial g(e_e,d)}{\partial d} > 0$. Hence, subjects will choose $d = d_{max} > 0$.

4. Both speed and accuracy incentives (B)

Subjects choose $e_{e,}\, e_s, d$ to $\max F + g(e_e, d) + f(e_s, d) - c(e_e) - c(e_s)$. To maximize their profit, given $\frac{dc(e_e)}{de_e} > 0$, $\frac{dc(e_s)}{de_s} > 0$, $\frac{\partial^2 f(e_s,d)}{\partial e_s{}^2} - \frac{d^2 c(e_s)}{de_s{}^2} < 0$, $\frac{\partial^2 g(e_e,d)}{\partial e_e{}^2} - \frac{d^2 c(e_e)}{de_e{}^2} < 0$, and $\frac{\partial^2 g(e_e,d)}{\partial d^2} + \frac{\partial^2 f(e_s,d)}{\partial d^2} < 0$, subjects will choose

$e_s > 0$ so that $\frac{\partial f(e_s+d)}{\partial e_s} = \frac{dc(e_s)}{de_s} > 0$

$e_e > 0$ so that $\frac{\partial g(e_e,d)}{\partial e_e} = \frac{dc(e_e)}{de_e} > 0$

$d_{max} > d > 0$ so that $\frac{\partial g(e_e,d)}{\partial d} = -\frac{\partial f(e_s,d)}{\partial d} > 0$

For the ease of notation, we summarize the theoretical predictions in the form of treatment comparisons.

- Performance speed effort without considering delay e_s: B=S>A=N
- Accuracy (estimation effort) e_e: A=B>S=N
- Deliberate delay d: A>B>S=N=0

Based on the above theoretical predictions and in line with our motivating conjecture, we derive the following hypotheses:

- *Hypothesis 1: Speed incentives result in faster task completion.*
- *Hypothesis 2: Accuracy incentives induce deliberately inflated estimates and slower performance so that the actual task duration matches the estimate.*
- *Hypothesis 3: The incentive structure rewarding both speed and accuracy results in faster performance compared to accuracy-only incentives, while the estimation accuracy does not deteriorate.*

5. Results

A total of 206 subjects participated in the experiment. Data of eight subjects were excluded from the analysis due their misunderstanding of the instructions, limiting our sample to 198 subjects (77 females) with a mean age of 24.1 and a standard deviation of 6.1 years.[11] Of these 198 subjects, 59 were randomly assigned into the ABA treatment, 57 into the BAB treatment, 41 into the NNN treatment, and 41 into the SSS treatment.[12] On average, an experimental session lasted around 60 minutes including the initial instructional period and payment of subjects. The subjects earned AUD 23.10 on average. Summary statistics by treatments and rounds are presented in Table 2.

[11] These eight subjects repeatedly provided an answer different from five when asked about the completion time of how many tables they were estimating.

[12] Buehler, Griffin, & MacDonald (1997) find that subjects exhibit an optimistic bias under speed-only incentives (M=-1.37, t=1.96) and a pessimistic bias under accuracy-only incentives (M=1.11, t=1.61). They do not find any bias in their control treatment (M=-0.27). Using their effect sizes, we computed that the required sample size to achieve a significance level of 5% and statistical power of 90% is 12 and 8 to detect the effect of speed and accuracy incentives, respectively. (G*power software; Faul, Erdfelder, Lang, & Buchner, 2007). The sample sizes in our main and control treatments are significantly larger than that.

Table 2: Summary statistics by treatment and round

		Summary statistics – means (SD)			
		ABA (N=59)	BAB (N=57)	SSS (N=41)	NNN (N=41)
Estimates	Round 1	421 (401)	294 (185)	268 (197)	523 (340)
	Round 2	315 (85)	347 (126)	291 (98)	411 (171)
	Round 3	329 (98)	318 (104)	277 (97)	373 (121)
Actual duration (all tables)	Round 1	381 (199)	351 (98)	309 (79)	355 (111)
	Round 2	318 (85)	336 (89)	314 (98)	339 (82)
	Round 3	343 (95)	312 (78)	286 (77)	319 (62)
Actual duration – (5^{th} table)	Round 1	105 (119)	74 (38)	66 (27)	77 (33)
	Round 2	83 (53)	91 (55)	65 (27)	65 (24)
	Round 3	99 (61)	80 (48)	63 (26)	70 (28)
Inaccuracy (Absolute estimation errors)	Round 1	155 (301)	114 (116)	133 (169)	234 (245)
	Round 2	29 (36)	39 (84)	66 (82)	102 (110)
	Round 3	25 (34)	42 (80)	58 (85)	71 (80)
Clicks on the "Update" button	Round 1	6 (11)	2 (4)	0 (1)	1 (1)
	Round 2	3 (7)	5 (11)	0 (0)	0 (1)
	Round 3	5 (9)	2 (4)	0 (0)	0 (1)
Time taken to provide the estimate	Round 1	50 (42)	49 (35)	29 (18)	21 (9)
	Round 2	22 (16)	18 (12)	16 (13)	12 (6)
	Round 3	14 (10)	19 (17)	12 (9)	9 (6)
Number of incorrect answers	Round 1	3 (3)	3 (3)	3 (3)	2 (2)
	Round 2	2 (3)	2 (2)	3 (3)	2 (2)
	Round 3	2 (3)	2 (2)	2 (2)	2 (2)

Notes: The table presents the means and standard deviations. Estimates, actual duration, inaccuracy, and estimation time are measured in seconds.

To test the effect of our treatment manipulations, we conduct an OLS regression presented in Table 3. Consistent with our hypotheses, speed and accuracy incentives operate in the opposite direction. In particular, we find that speed incentives result in a significantly lower actual task duration (as well as estimates), while the accuracy incentives result in significantly higher task duration.

Result 1: Subjects complete the task faster when incentivized for speed. When incentivized for accuracy, subjects work more slowly.

Table 3: Regression analysis (OLS model)

	(1) Estimate	(2) Actual duration (all tables)	(3) Actual duration (5th table)	(4) Inaccuracy (Absolute est. errors)	(5) Underestimation (Estimate – Duration < 0)	(6) Overestimation (Estimate – Duration < 0)	(7) Clicks on "Update" button	(8) Time to provide estimate	(9) Number of incorrect answers
Accuracy	-15.60	24.78***	22.25***	-40.82***	18.89**	-33.41	3.73***	12.50***	0.22
incentives	(16.78)	(8.11)	(4.29)	(11.95)	(8.55)	(23.65)	(0.48)	(1.55)	(0.20)
Speed	-97.45***	-29.81***	-14.13***	-25.29**	-17.61**	-56.64**	-1.87***	3.65**	0.43**
incentives	(16.21)	(7.98)	(4.40)	(11.69)	(7.92)	(21.98)	(0.50)	(1.70)	(0.20)
Round 2	-34.15	-24.87**	-4.24	-99.95***	68.61***	-144.99***	0.21	-21.59***	-0.32
	(22.82)	(10.95)	(5.88)	(16.16)	(9.53)	(35.75)	(0.67)	(2.41)	(0.26)
Round 3	-49.86**	-34.49***	-1.87	-108.74***	73.82***	-145.16***	-0.13	-25.12***	-0.53**
	(22.56)	(10.63)	(5.90)	(16.02)	(10.07)	(32.56)	(0.60)	(2.39)	(0.25)
Age	2.15	3.27***	1.00**	0.25	-0.94	-0.81	0.01	0.40	0.01
	(2.00)	(1.22)	(0.46)	(1.59)	(1.11)	(3.47)	(0.05)	(0.24)	(0.03)
Female	-2.78	-1.69	-5.19	-10.11	-0.48	-23.65	0.16	2.81	0.01
	(19.15)	(8.17)	(4.23)	(14.16)	(9.46)	(25.69)	(0.61)	(1.84)	(0.24)
Degree	0.63	-6.83	-3.31	-0.34	2.53	-0.08	-0.23	0.13	-0.14
	(8.03)	(4.59)	(2.35)	(5.87)	(3.90)	(10.29)	(0.21)	(0.84)	(0.08)
Playing video	-16.54**	-6.09	0.22	-14.26***	2.20	-27.12***	-0.19	-0.92	0.07
games often	(7.32)	(3.81)	(2.30)	(5.21)	(4.03)	(9.93)	(0.27)	(0.99)	(0.09)
Cognitive	-19.97***	-7.34**	2.95*	-26.30***	14.73***	-35.74***	0.43*	0.32	-0.17*
reflection	(6.63)	(2.99)	(1.56)	(4.85)	(3.30)	(9.54)	(0.24)	(0.72)	(0.09)
Practice table	0.44**	0.65***	0.18***	0.12	-0.14*	0.12	-0.01	0.02	0.00
duration	(0.20)	(0.12)	(0.06)	(0.11)	(0.08)	(0.28)	(0.00)	(0.02)	(0.00)
Risk attitudes	3.15	0.28	0.51	-0.44	2.73	1.93	-0.03	0.63	-0.02
	(4.46)	(1.87)	(0.95)	(3.66)	(1.85)	(7.14)	(0.10)	(0.41)	(0.05)
Constant	396.15***	265.79***	41.57***	258.67***	-137.05***	378.04***	1.91	15.37**	2.34***
	(55.38)	(28.92)	(14.51)	(40.22)	(27.91)	(83.13)	(1.80)	(6.34)	(0.72)
N	594	594	594	594	284	282	594	594	594
R^2	0.11	0.15	0.08	0.19	0.31	0.22	0.11	0.29	0.03

Note: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

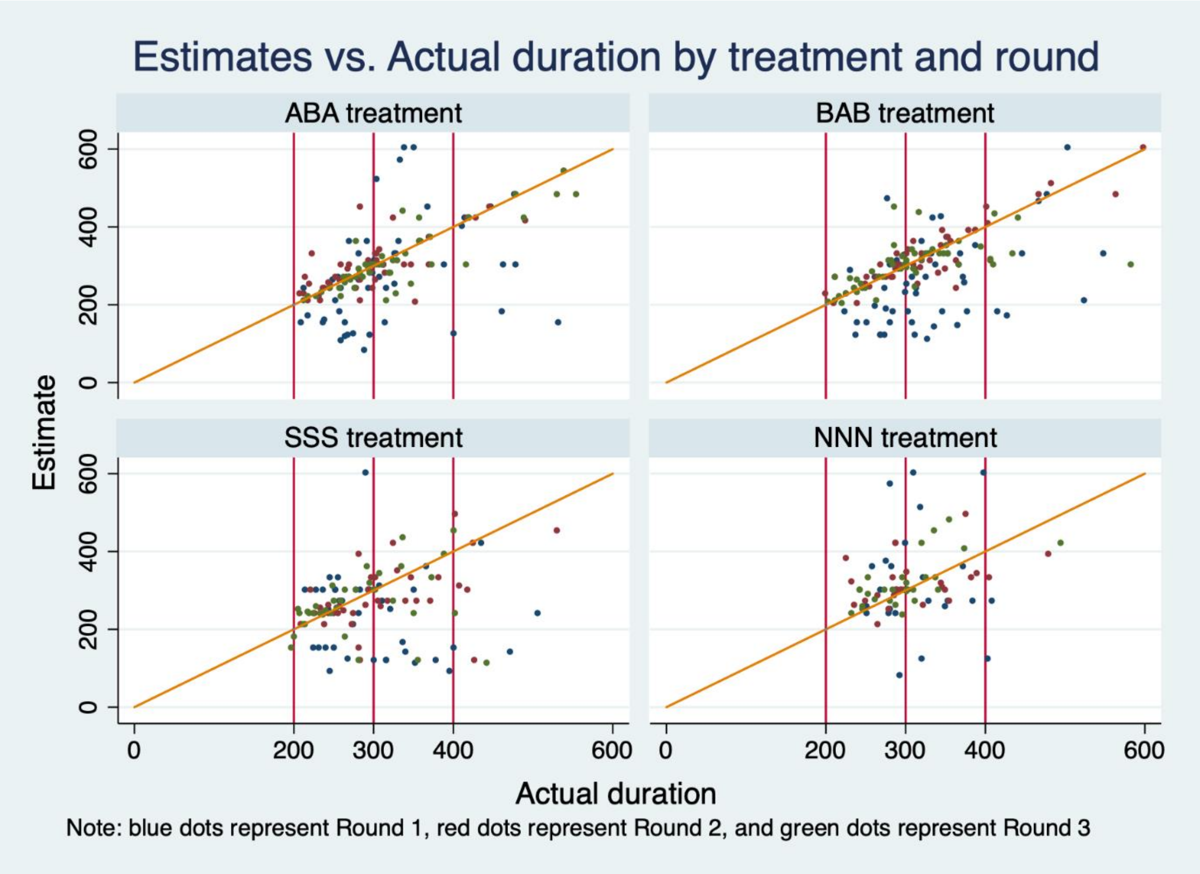
Estimation bias and (in)accuracy

Figure 2 shows the relationship between estimates and the actual duration in more detail. The figure displays scatter plots of individual-level estimates on the vertical axis and the actual duration on the horizontal axis by treatment and round. Precise estimates lie on the orange 45-degree line. A dot above the orange line indicates overestimation, while a dot below the orange line indicates underestimation.[13] As apparent from the figure, overestimation is the most prevalent in the NNN treatment, while underestimation in the SSS treatment. Evidence of speed incentives resulting in underestimation of task duration can also be found in the regression analysis (Table 3, Model 5) and is in line with extant literature (Halkjelsvik and Jørgensen, 2012).

In the treatments with incentives for estimation accuracy (ABA and BAB treatments) we observe the actual duration being closely matched with the estimates more often. This is the case especially in Round 2 of the BAB treatment (Figure 2, red dots) and Round 3 of ABA treatment (green dots), in which estimation accuracy is the only incentive (while accuracy is the only incentive also in Round 1 of ABA, the estimation errors are larger due to the lack of experience with estimation).

[13] For presentational clarity purposes, we have removed from the figure estimates and actual durations that exceeded 600 seconds.

Figure 2: Individual-level estimates vs. actual task duration by treatment and round



Evidence of strategic behavior

Do subjects in the ABA and BAB treatments achieve accuracy by inflating estimates and deliberately prolonging the task execution so that their performance matches their estimate? If they wanted to do so, they would likely slow down on the last table of the round, since delaying the execution on the last table bears less risk than delaying any other table.

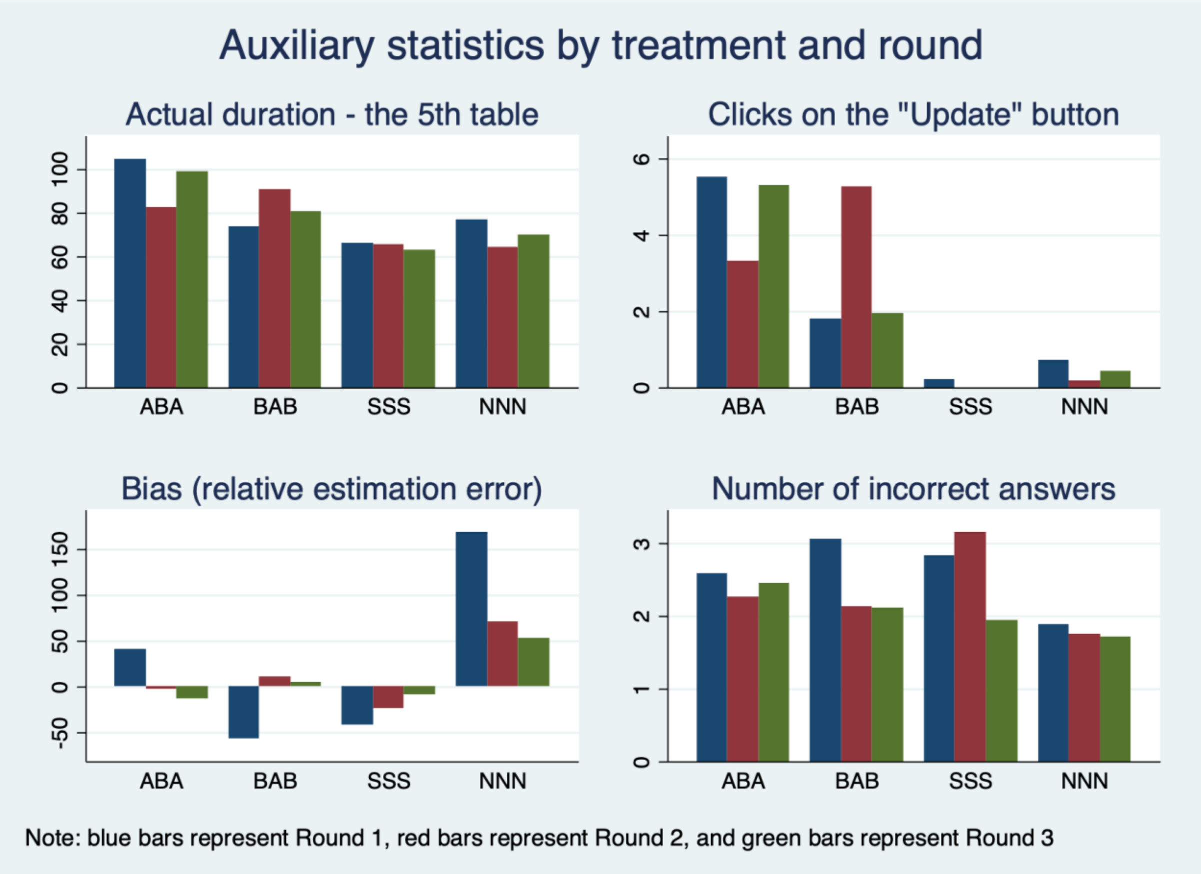
In the regression analysis (Table 3, model 3) we indeed find that accuracy incentives lead to significantly longer time to complete the fifth (i.e., last) tables (see also Figure 3). In the subsequent within-subject analysis we find that subjects with accuracy incentives take significantly longer to solve the fifth table compared to the average time spent on solving one table out of the first four (the Wilcoxon matched-pairs signed-rank test yields a p-value <0.01 for the pooled three rounds under the incentive structure A as well as B). Thus, the observed low estimation errors are indeed reached by taking more time than necessary to complete the task.

Recall that our design allows us not only to observe how long subjects take to complete each table, but also how long they take to provide their estimates in each round and how often they use the "Update" button to check the elapsed time during task execution. A longer time spent on estimation and more clicks on the "Update" button could be signs of strategic behavior. We find that subjects facing accuracy incentives take significantly longer to provide their estimates (Table 3, model 8) and check time significantly more often (Table 3, model 7; see also Figure 3).

While the regression in Table 3 shows a non-significant (and negative) effect of accuracy incentives on estimates, we note that the result is driven mostly by the baseline in the regression being the NNN treatment in which subjects are not rewarded for their estimation effort. In the subsequent analysis in which only ABA and BAB treatments are included (see Table 4), we indeed find support for deliberative inflation of estimates when only accuracy is incentivized.

Result 2: Subjects incentivized for estimation accuracy deliberately inflate their estimates and subsequently take longer to finish the task by strategically pacing themselves so that their actual task duration matches their estimate.

Figure 3: Auxiliary statistics by treatment and round



ABA vs. BAB

Next, we test Hypotheses 3 stating that adding speed incentives to accuracy incentives results in more effective task delivery. We focus solely on the ABA and BAB treatments and conduct linear fixed-effect regressions, presented in Table 4. We find that estimation accuracy incentives (incentive structure A) result in significantly higher estimates and significantly longer actual task duration than joint incentives for estimation accuracy and performance speed (incentive structure B). We also find that the longer task duration under the incentive structure A is not accompanied by higher quality. This is because the tables are accepted by the software only if checked correctly and because there is no significant effect of the incentive structures on the number of incorrect table submissions. The regression in Table 4 also shows that the two incentive structures do not result in significantly different estimation errors, meaning that subjects are similarly accurate in their estimation under both A and B.

Result 3: Implementing speed incentives alongside accuracy incentives improves the efficiency, as it results in faster task completion, while not compromising the estimation accuracy or the quality of work.

Table 4: Fixed-effects OLS models

Dependent Variables	(1) Estimates	(2) Actual duration	(3) Inaccuracy (Absolute est. errors)	(4) Actual duration (5th table)	(5) Clicks on "Update" (5th table)	(6) Number of Incorrect answers
Incentive structure A	50.77***	24.32***	10.76	14.98***	2.28***	-0.09
	(13.08)	(7.07)	(10.65)	(5.00)	(0.75)	(0.21)
Round 2	-27.64	-38.80***	-101.09***	2.95	0.61	-0.61**
	(25.56)	(10.70)	(20.30)	(6.53)	(0.84)	(0.25)
Round 3	-35.35	-38.28***	-101.59***	6.45	0.61	-0.53*
	(26.51)	(11.21)	(20.51)	(6.16)	(0.56)	(0.27)
Time spent on the 1st table				0.16		
				(0.42)		
Time spent on the 2nd table				-0.00		
				(0.23)		
Time spent on the 3rd table				0.22		
				(0.20)		
Time spent on the 4th table				0.24		
				(0.19)		
Update button clicks (1st table)					-0.31	
					(0.54)	
Update button clicks (2nd table)					0.26	
					(0.61)	
Update button clicks (3rd table)					0.70*	
					(0.41)	
Update button clicks (4th table)					0.11	
					(0.19)	
Constant	333.06***	353.90***	129.63***	39.13	1.11*	2.86***
	(13.31)	(5.52)	(9.64)	(54.36)	(0.56)	(0.18)
N	348	348	348	348	348	348

Note: Standard errors are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1%-level, respectively.

Control treatments (NNN and SSS)

Finally, we briefly present the results of our two control treatments. In the NNN treatment, subjects take (significantly) the shortest time to provide their estimates, suggesting they likely put less effort into estimation than in all other treatments. The provided estimates are inaccurate and (mostly) overstate the actual task duration. Interestingly, while subjects in the NNN treatment take significantly more time to complete the task than subjects in the SSS treatment, their performance speed is similar to the speed observed in the ABA and BAB treatments. On the other hand, subjects in the SSS treatment complete the task significantly faster than in any other treatment. However, they also underestimate duration the most across all treatments.

6. Discussion

Widely used project management methodologies (IPMA, 2015; Project Management Institute, 2013) evaluate the success of a business project by the extent to which all specified outcomes are completed and match the desired quality requirements (scope constraint), stay within the allocated budget (cost constraint), and are delivered within the approved schedule (time constraint). The three constraints are interdependent. For example, extensions to the current scope are generally associated with an extended budget and/or extended schedule. On the other hand, a more compressed schedule may require a larger budget as more resources are necessary to finish the project in shorter time, or a reduction in the number or quality of deliverables. In this paper, we experimentally explore the behavior of individuals exclusively in the time dimension, holding the scope and cost dimensions constant to eradicate the possible confounds stemming from the interdependence of the three constraints. The time dimension is arguably the most important for projects with hard deadlines, such as construction of venues for special events (e.g., the Olympic Games). Furthermore, since the project schedule is often used as a basis for cost estimation, its accuracy is particularly crucial for projects in which labor costs constitute a large portion of the overall costs (e.g., software development projects). We argue that emphasizing the schedule accuracy, for example by using monetary incentives to reward projects delivered on time or by including it in the employee performance evaluation, can induce a hidden inefficiency due to inflated schedules and slower project execution.

Since detecting the hidden inefficiency using happenstance data is challenging due to unobservables (e.g., the amount of exerted effort or the time actually used), we conduct a controlled laboratory experiment in which we manipulate the incentive structures. To investigate how the conjectured strategic behavior interacts with increasing experience with the task, we repeat the estimation and task execution three times for each subject. We find that the incentive structure that rewards the planners solely for their estimation accuracy indeed leads to inflated estimates and deliberately slower task execution. When speed incentives are implemented alongside estimation accuracy incentives, the estimates are less inflated, and the task is performed more quickly. Importantly, the estimation accuracy and the quality of the work are not compromised. Under speed incentives only, subjects complete the task the fastest, but still underestimate task duration. With no financial incentives for estimation accuracy or performance speed, subjects provide the estimates faster than in other treatments, but these estimates are inaccurate as subjects vastly overestimate task duration.

In treatments featuring estimation accuracy incentives, there is heterogeneity across subjects in terms of their strategic behavior as evidenced by the quantitative analysis. Further evidence is provided by subjects' responses to open-ended question regarding their strategies. Some subjects inflate their estimates and prolong the task execution right from the first round ("*I knew the task would take me around 6 minutes, so I gave myself 8 minutes for first and third round, as time did not matter. For second round, I gave myself 7 minutes as I wanted to have a little extra time but still finish fast.*" [sic]). Some subjects find effective strategies only after estimating and executing the task ("*For the first round, I based my estimate on how quickly I thought I could complete it, which I now realize was a mistake. I should have just put in a big estimate and finish it slowly to increase my chance of estimating correctly.*"). Some of them do not behave strategically at all and solve all tables at their own pace, without too much consideration of the resulting earnings. And finally, some subjects recognize the best strategies ("*There is no excuse for not earning the maximum accuracy earnings in rounds without speed money.*"), but do not take advantage of the environment extensively, presumably because they view the strategic behavior as dishonest ("*It seemed pretty easy to 'exploit' the tasks based solely on time estimation, was that intended?").* While previous research documents that people may utilize the existing incentives in their favor and provide biased answers (e.g., Scheubel, Schunk, & Winter, 2013), the hesitance to fully exploit the imposed incentive structure is not uncommon. For example, experiments by Fischbacher & Föllmi-Heusi (2013), Mazar, Amir, & Ariely (2008), and Scheele, Thonemann, & Slikker (2018) find that when given a chance, most people act dishonestly, but the magnitude of such behavior is far away from the maximum (see also Özer, Zheng & Chen, 2011 for a study on trust and trustworthiness in forecast information sharing). Our results are also consistent with the finding that about a half of credence goods sellers prefers to act honestly even if they have a substantial informational advantage (Dulleck, Kerschbamer, & Sutter, 2011).

We present unambiguous empirical evidence that incentivizing project planners for the accuracy of project estimates can induce moral hazard, resulting in hidden inefficiency of resource underutilization. We demonstrate the inefficiency in a stylized laboratory environment and provide a rigorous proof of concept. While we do not advocate generalizing the results of our experiment, the obtained support for our general theoretical conjectures represents a valid concern from the perspective of operational effectiveness in organizations, implying that managers should design incentives for their project planners carefully. In particular, managers should consider incentivizing fast project performance, possibly utilizing the historical information regarding similar projects in the past, which seems to generate reasonably accurate predictions (Lorko et al., 2021). Incentivizing fast project performance discourages wasting time and other resources only to deliver the project exactly "on time" and also could offset motivation to work slower in order to avoid being assigned more work for the same overall salary.

While our study demonstrates the positive effects of utilizing speed incentives in project management, such incentives inevitably result in more costly project execution. As our study focuses on the behavioral response to incentives, we do not incorporate a trade-off between additional costs (due to speed incentives) and additional revenues (due to time savings). For the speed incentives to be efficient in any application of our results, the extra expenses need to be lower than what can be gained from faster project completion. An example of appropriately used speed incentives can be found in the empirical analysis of highway construction projects in California (Lewis & Bajari, 2011). The study reveals that projects with contracts featuring speed incentives for accelerated delivery are completed faster than projects without such incentives, while maintaining the same quality. Based on the underlying parametric assumptions, the use of speed incentives increases the overall welfare as the extra remuneration is substantially lower than the commuter surplus gained from quicker construction.

By not including project owners in our design, we are able to abstract from their strategic considerations and detect an unconfounded effect of incentives on behavior of planners. However, it is important to keep in mind that an experienced manager or customer may be able to mitigate the inefficiency by rejecting project schedules that appear inflated or using different individuals for planning and project execution purposes. The hidden inefficiency can be also reduced by using critical chain scheduling or iterative approach such as Agile project management (Cohn, 2010; Project Management Institute, 2018), or by introducing competition amongst potential project contractors, although the latter option may not always be viable, especially for internal projects in organizations. On the other hand, the presence of an inexperienced project owner could also exacerbate the inefficiency if it remains undetected. Future research could enrich the understanding of incentives in project management by incorporating the above-mentioned trade-offs as well as strategic considerations of project owners while also varying the strength of incentives. Another promising direction is exploring how incentive structures affect the estimation and execution of more complex and/or creative tasks which are associated with more uncertainty in terms of scope, less predictability in terms of time, and variable quality.

Finally, while our study focuses solely on overestimation, it is important to recognize that the strategic misestimation of project schedules can also have the opposite direction. In a competitive environment, individuals and organizations often have incentives to deliberately underrepresent the necessary time (and cost) associated with a project, e.g., to secure the contract and thereby to put a foot in the door; hoping to recoup the losses from underestimation later, via contract amendments. Thus, future research could shed light also on strategic underbidding, for example in procurement or supply chain environments. It seems plausible that in such scenarios, utilizing estimation accuracy incentives could aid more honest estimation and help to mitigate inefficiencies instead of facilitating them as is the case in scenarios investigated in the current paper.

Acknowledgements: This paper is based on Matej Lorko's dissertation chapter written at the Macquarie Graduate School of Management. We wish to thank the Editor Daniela Puzzello, the associate editor, three referees as well as Barbora Baisa, Lee Benham, Lubomír Cingl, Michal Ďuriník, Guillaume Fréchette, John Nye, Tommaso Reggiani, Roman Sheremeta, Scott Wallsten, Jingjing Zhang, the audiences at the 2018 Ronald Coase Institute Workshop on Institutional Analysis, 2019 Sydney Experimental Brownbag Seminar, 2019 Economic Science Association World Meeting, and 2019 Slovak Economic Association Meeting for helpful comments and suggestions. Financial support was provided by Macquarie Graduate School of Management. Maroš Servátka thanks University of Alaska – Anchorage for their kind hospitality while working on this paper. Le Zhang thanks Australian Research Council for her DECRA fellowship while working on this paper and for financial support from Macquarie Business School.

References

Akerlof, G. A. (1991). Procrastination and obedience. *American Economic Review*, *81*(2), 1–19.

Alan, S. D., Hossein, N., Dunk, A. S., & Nouri, H. (1998). Antecedents of budgetary slack: A literature review and synthesis. *Journal of Accounting Literature*, *17*, 72. Retrieved from http://proquest.umi.com/pqdlink?did=113703689&Fmt=7&clientId=36305&RQT=309&VName=PQD

Ariely, D., & Wertenbroch, K. (2002). Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, *13*(3), 219–224. https://doi.org/10.1111/1467-9280.00441

Bajari, P., & Tadelis, S. (2001). Incentives versus transaction costs: A theory of procurement contracts. *Rand Journal of Economics*, 387–407.

Brown, J. L., Evans, J. H., & Moser, D. V. (2009). Agency Theory and Participative Budgeting Experiments. *Journal of Management Accounting Research*, *21*(1), 317–345. https://doi.org/10.2308/jmar.2009.21.1.317

Buehler, R., Griffin, D., & MacDonald, H. (1997). The Role of Motivated Reasoning in Optimistic Time Predictions. *Personality and Social Psychology Bulletin*, *23*(3), 238–247.

https://doi.org/10.1177/0146167297233003
Buehler, Roger, Griffin, D., & MacDonald, H. (1997). The role of motivated reasoning in optimistic time predictions. *Personality and Social Psychology Bulletin*, *23*(3), 238–247. https://doi.org/10.1177/0146167297233003
Buehler, Roger, Griffin, D., & Ross, M. (1994). Exploring the "planning fallacy": Why people underestimate their task completion times. *Journal of Personality and Social Psychology*, *67*(3), 366–381. https://doi.org/10.1037/0022-3514.67.3.366
Cahlíková, J., Cingl, L., & Levely, I. (2020). How Stress Affects Performance and Competitiveness Across Gender. *Management Science*. https://doi.org/10.1287/mnsc.2019.3400
Chao, R. O., Lichtendahl, K. C., & Grushka-Cockayne, Y. (2014). Incentives in a stage-gate process. *Production and Operations Management*. https://doi.org/10.1111/poms.12166
Cohn, M. (2010). *Succeeding with Agile: Software Development Using Scrum*. Pearson Education.
Covaleski, M., Evans, J. H., Luft, J., & Shields, M. D. (2006). Budgeting Research: Three Theoretical Perspectives and Criteria for Selective Integration. *Handbooks of Management Accounting Research*. https://doi.org/10.1016/S1751-3243(06)02006-2
Dulleck, U., Kerschbamer, R., & Sutter, M. (2011). The economics of credence goods: An experiment on the role of liability, verifiability, reputation, and competition. *American Economic Review*. https://doi.org/10.1257/aer.101.2.526
Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3.1 manual. *Behavioral Research Methods*.
Fischbacher, U. (2007). Z-Tree: Zurich toolbox for ready-made economic experiments. *Experimental Economics*, *10*(2), 171–178. https://doi.org/10.1007/s10683-006-9159-4
Fischbacher, U., & Föllmi-Heusi, F. (2013). Lies in disguise-an experimental study on cheating. *Journal of the European Economic Association*. https://doi.org/10.1111/jeea.12014
Flyvbjerg, B., Holm, M. S., & Buhl, S. (2002). Underestimating costs in public works projects: Error or lie? *Journal of the American Planning Association*, *68*(3), 279–295. https://doi.org/10.1080/01944360208976273
Frederick, S. (2005). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*. https://doi.org/10.1257/089533005775196732
Goldratt, E. M. (1997). *Critical Chain: A Business Novel*. Great Barrington: North River Press.
Greiner, B. (2015). Subject pool recruitment procedures: organizing experiments with ORSEE. *Journal of the Economic Science Association*, *1*(1), 114–125. https://doi.org/10.1007/s40881-015-0004-4
Grushka-Cockayne, Y., Erat, S., Wooten, J., Donohue, K., Katok, E., & Leider, S. (2018). New product development and project management decisions. In *In The Handbook of Behavioral Operations* (pp. 367–392). NJ: Wiley.
Halkjelsvik, T., & Jørgensen, M. (2012). From origami to software development: A review of studies on judgment-based predictions of performance time. *Psychological Bulletin*, *138*(2), 238–271. https://doi.org/10.1037/a0025996
Holt, C. A., & Laury, S. K. (2002). Risk aversion and incentive effects. *American Economic Review*, *92*(5), 1644–1655. https://doi.org/10.1257/000282802762024700
IPMA. (2015). *IPMA Competence Baseline (ICB), Version 4.0*. *International Project Management Association*. https://doi.org/10.1002/ejoc.201200111
Knowles, S., Servátka, M., Sullivan, T., & Genç, M. (2022). Procrastination and the non-monotonic effect of deadlines on task completion. *Economic Inquiry*. https://doi.org/10.1111/ecin.13042
Lederer, A. L., Mirani, R., Neo, B. S., Pollard, C., Prasad, J., Ramamurthy, K., & Lederer, B. A. L. (1990). Information System Cost Estimating : A Management Perspective. *MIS Quarterly*, *14*(2), 159–176. https://doi.org/10.2307/248774
Lee, C. B., Schwartzman, Y., Hardy, J., & Snavely, A. (2005). Are User Runtime Estimates Inherently Inaccurate? *Job Scheduling Strategies for Parallel Processing*, (March), 253–263.
Lewis, G., & Bajari, P. (2011). Procurement contracting with time incentives: Theory and evidence.

Quarterly Journal of Economics, *126*(3), 1173–1211. https://doi.org/10.1093/qje/qjr026
Lorko, M., Servátka, M., & Zhang, L. (2019). Anchoring in project duration estimation. *Journal of Economic Behavior & Organization*, *162*, 49–65.
Lorko, M., Servátka, M., & Zhang, L. (2021). Improving the accuracy of project schedules. *Production and Operations Management*, *30*(6), 1633–1646.
Magazinius, A., Börjesson, S., & Feldt, R. (2012). Investigating intentional distortions in software cost estimation - An exploratory study. *Journal of Systems and Software*, *85*(8), 1770–1781. https://doi.org/10.1016/j.jss.2012.03.026
Mazar, N., Amir, O., & Ariely, D. (2008). The Dishonesty of Honest People: A Theory of Self-Concept Maintenance. *Journal of Marketing Research*. https://doi.org/10.1509/jmkr.45.6.633
O'Donoghue, T., & Rabin, M. (2008). Procrastination on long-term projects. *Journal of Economic Behavior & Organization*, *66*(2), 161–175.
Özer, Ö., Zheng, Y., & Chen, K. Y. (2011). Trust in forecast information sharing. *Management Science*. https://doi.org/10.1287/mnsc.1110.1334
Project Management Institute. (2013). *A guide to the project management body of knowledge (PMBOK ® guide)*. *Project Management Institute*. https://doi.org/10.1002/pmj.20125
Project Management Institute, I. (2018). Agile Practice Guide. In *Agile Practice Guide*.
Roy, M. M., & Christenfeld, N. J. S. (2007). Bias in memory predicts bias in estimation of future task duration. *Memory & Cognition*, *35*(3), 557–564. https://doi.org/10.3758/BF03193294
Scheele, L. M., Thonemann, U. W., & Slikker, M. (2018). Designing incentive systems for truthful forecast information sharing within a firm. *Management Science*. https://doi.org/10.1287/mnsc.2017.2805
Scheubel, B., Schunk, D., & Winter, J. (2013). Strategic Responses: A Survey Experiment on Opposition to Pension Reforms*. *Scandinavian Journal of Economics*. https://doi.org/10.1111/sjoe.12009
Shepperd, J. A., Sweeny, K., & Cherry, L. C. (2007). Influencing audience satisfaction by manipulating expectations. *Social Influence*, *2*(2), 98–111. https://doi.org/10.1080/15534510601095772
von Branconi, C., & Loch, C. H. (2004). Contracting for major projects: eight business levers for top management. *International Journal of Project Management*, *22*(2), 119–130.
Woods, D., & Servátka, M. (2016). Testing psychological forward induction and the updating of beliefs in the lost wallet game. *Journal of Economic Psychology*, *56*, 116–125. https://doi.org/10.1016/j.joep.2016.06.006
Yimga, J. (2020). To Pad or Not to Pad? A Note on the Schedule Padding Behavior of Airlines. *Applied Economics Letters*. https://doi.org/10.1080/13504851.2020.1866155
Yimga, J. (2021). Competition and schedule padding in the US airline industry. *Review of Network Economics*. https://doi.org/10.1515/rne-2021-0016
Zhang, D., Salant, Y., & Van Mieghem, J. A. (2018). Where Did the Time Go? On the Increase in Airline Schedule Padding Over 21 Years. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3238457
Zizzo, D. J. (2010). Experimenter demand effects in economic experiments. *Experimental Economics*, *13*(1), 75–98.