Hidden inefficiency: Strategic inflation of project schedules

Lorko, Matej and Servátka, Maroš and Zhang, Le

MGSM Experimental Economics Laboratory, Macquarie Graduate School of Management, University of Economics in Bratislava

20 September 2020

Hidden inefficiency: Strategic inflation of project schedules

Matej Lorko
MGSM Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia
University of Economics in Bratislava, Slovakia
matej.lorko@gmail.com

Maroš Servátka
MGSM Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia
University of Economics in Bratislava, Slovakia
maros.servatka@mq.edu.au

Le Zhang
MGSM Experimental Economics Laboratory, Macquarie Business School, Sydney, Australia
lyla.zhang@mq.edu.au

September 19, 2020

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 introduce moral hazard, as people may respond to estimation accuracy incentives by strategically inflating project duration estimates and subsequently prolonging the project execution. While the project is delivered on time, the resources are underutilized. We conjecture that these perverse effects can be mitigated by introducing incentives to complete the project as soon as possible (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 prolonged 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. Aligning the objectives of a project owner with those of planners, by incentivizing the planners for both estimation accuracy and quick project completion, can therefore foster more compressed but still accurate and reliable project schedules and accelerated project delivery. In summary, our study contributes to the economic analysis of incentive structures in project management by identifying a hidden inefficiency that could be present in projects delivered "on time" and by pointing out a mechanism that mitigates the risk of moral hazard.

Keywords: project management, project planning, time management, duration estimation, strategic overrepresentation, 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's 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. 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). Conversely, strategic considerations may lead planners to deliberately overestimate the project duration to gain more time for the project 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.[1]

In this paper, we focus on the case of deliberate overestimation. We argue that the overestimation might occur when the planner both estimates the project duration and executes the project, as

[1] Both strategic underestimation and overestimation have been empirically documented in budget planning, which is an estimation problem similar to planning time to be spent on the project. See section 2 for a detailed discussion.

forecasting own future performance induces moral hazard.[2] Such scenarios 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. 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.

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).

For illustration, consider a scenario within an organization where an employee responsible for project planning and execution knows more than his executives about the involved tasks and their likely duration. 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 in 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. However, from the perspective of the organization,

[2] While the literature on optimal forecasting incentives (Osband, 1989) assumes that the person making an estimate has no influence over the realization of the estimated outcome, we note that in project planning and execution, people often do forecast their own future performance.

inflated estimates may lead to prolonged project execution without commensurate improvement in the quality of deliverables.[3]

Strategic inflation of project schedules and the resulting inefficiencies 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 might 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. [4]

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

---

[3] 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 might have proposed inflated schedules and overspent time (and funds) on the projects.

[4] The difficulty with detecting inflated estimates is akin to a situation encountered with credence goods (Darby & Karni, 1973; Dulleck & Kerschbamer, 2006; Dulleck et al., 2011). Compared to the research on credence goods, in the current paper we abstract from the project owners (referred to as consumers in the credence goods literature). This omission enables us to elicit incentives effects that are not confounded by considerations of whether the project will be accepted by the project owner. We elaborate on the issue of project owners in the Discussion section.

hidden inefficiency.[5] 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. An additional benefit of a stylized experiment is that it allows us to study 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. Since project management studies rarely employ experiments in which subject behavior is fully incentivized, our contribution to the project management literature could also be viewed as methodological.[6]

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. Subjects are always incentivized for their estimation accuracy, meaning that the more accurate their estimates are, the more money they earn. To parallel business practice, 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 everyday life, experience often enhances the understanding of incentives one is facing (Plott, 1987; Smith, 1962, 1991). With accumulated experience, individuals in organizations likely have a thorough knowledge of how their estimation accuracy and performance impact their compensation. In the

[5] Conditional on detecting morally hazardous behavior of planners, the next step would be look at the how much inefficiency there is and explore the trade-off between increased renumeration costs due to speed incentives and the costs of prolonged project execution.

[6] Recent large-scale interdisciplinary replication efforts document that incentivized experiments have a higher replicability rate than non-incentivized experiments (Camerer et al., 2016; Open Science Collaboration, 2015).

experiment, we therefore repeat the duration estimation and task execution for three times. To allow for causal inference, we manipulate the incentive structure as follows. In the first two rounds, each subject once faces an incentive structure with accuracy incentives only and once with both accuracy and speed incentives. The order of structures is balanced and determined randomly across subjects. In the third round, each subject faces the same incentive structure as in the first round. The implemented design enables us to explore the effect of incentive structures not only between subjects but also within subjects. Thus, we can examine whether individuals strategically adjust their behavior in reaction to different incentive structures while controlling for the individual characteristics and experience.

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 accuracy. 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.

Our experimental results yield the following managerial implications. We observe that when incentivized for on-time project delivery, individuals indeed strive to deliver the project at the "deadline". However, solely incentivizing schedule accuracy results in a non-negligible efficiency loss (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". The benchmark to reward fast performance can be based on historical information, for example the average duration of completed similar projects in the past (Lorko, Servátka, & Zhang, 2020). 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

Incentives implemented by organizations are crucial determinants of whether the initiated projects are successful or not. Should the project planners be incentivized for estimation accuracy in order for the project to be delivered on time? Lederer & Prasad (1998) report that making planners accountable for their estimates (cost estimates in particular) is the only organizational practice leading to more accurate project estimation and suggest the use of estimation accuracy as a metric in performance reviews of project planners. Jorgensen & Sjoberg (2001) warn that even though such incentives can indeed foster estimation accuracy, they might negatively affect the quality of project outcomes. On the other hand, Jensen (2003) argues that using targets, such as budget execution, in performance measurement and compensation systems induces manipulation in both estimation and spending. Our study investigates whether similar strategic behavior and the resulting inefficiencies are present also in the estimation of project duration and subsequent project execution. We contribute to the analysis of incentive structures in project management by identifying a hidden inefficiency due to inflated schedules in projects incentivizing estimation accuracy. If a project with an inflated schedule is delivered on time, it is deemed successful, often without realizing that it could have been delivered sooner.

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 extant 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.[7] 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 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. 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.[8] Also importantly, our task allows us to monitor the progress of each subject. 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 deliberate inflation of

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

[8] 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.

estimates (instead of reduced optimism), followed by inadequate adjustment of 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 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 works 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).

The speed incentives (in isolation) have also received some attention in extant research. In their review, Halkjelsvik and Jørgensen (2012) conclude that introducing financial incentives for finishing the assigned work fast usually leads to lower estimates of the time necessary to complete the respective task or project, but these incentives have a smaller impact on the actual performance. As a result, speed incentives usually trigger underestimation of task/project duration.[9] The study by Byram (1997) provides such evidence in a folding origami task. The incentives in Byram's study are designed in a tournament fashion, in which the subject's earnings depend on his rank within the group, sorted by the performance speed. The earnings are thus a function of both own performance and performances of others. Similar results are reported in another three studies (Henry, 1994; Henry & Sniezek, 1993; Henry and Strickland, 1994) utilizing a comparable tournament incentive structure. Out of the triad, only one study (Henry and Strickland, 1994) fails to find lower estimates under speed incentives, while none of them finds any effect of speed incentives on the actual task performance.[10]

---

[9] Underestimated project duration under speed incentives is often attributed to the concepts of motivated reasoning (people constructing seemingly reasonable justifications to arrive at the desired conclusion; Kunda, 1990), self-efficacy (decisions determined by expectations; Bandura, 1977) or strategic goal-setting (utilizing goals as a self-management technique; Latham & Locke, 1991).

[10] Subjects in these three studies do not estimate the duration of the task as in our experiment, but instead estimate how many almanac questions they will correctly answer within a given time span. Although the two elicitation formats – how quickly a task can be completed and how much can be achieved within a specific timeframe – are isomorphic in theory (if appropriately calibrated), the equivalence does not necessarily hold behaviorally (Halkjelsvik, Jørgensen, & Teigen, 2011). Since people usually have more experience with estimating how long something would take as opposed to how much can be done in a certain timeframe, we utilize the former format in our experiment, which is also advocated by Halkjelsvik and Jørgensen (2012).

We note that all above-mentioned studies on the effect of speed incentives on project estimation and execution feature either a tournament or competition against own past performance. It is conceivable that the emergence of underestimation when speed incentives are introduced is caused by these competitive features rather than by the speed incentives themselves. We contribute to this literature by cleanly identifying the effects of speed incentives by making them independent of the subject's past performance and the performance of others. In addition, by using a controlled environment that enables us to eliminate confounding factors that arise in everyday business environment and manipulate incentives, we investigate the existence of deliberate inflation in project duration estimates. Importantly, our experimental design allows us to explore subsequent strategic behavior during the task execution stage, i.e., prolonging the progress (spending more time than necessary) to ensure that the project is completed close to or right at the estimated time.

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.

Conditions

The experiment employs two incentive structures, A and B, where A stands for **a**ccuracy-only incentives and B for **b**oth accuracy and speed incentives. The incentive structures are implemented in two balanced ABA/BAB conditions. Each subject participates in one and only one condition, ABA or BAB, with a particular order of incentive structures over three rounds. In the ABA condition, 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 condition, subjects execute the task under the incentive structure B in Round 1, under A in Round 2, and again under B in Round 3. The across-subject comparison between ABA and

BAB enables us to identify the difference in behavior when facing the incentive structure A vs. B, while controlling for experience with estimation and task execution. The within-subject comparison controls for the individual differences 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). Comparing behavior in Round 1 and Round 3 in turn identifies the learning effect within the same incentive structure.

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 5 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 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. 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. 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 5 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. We therefore provide subjects with a time measuring tool on their screen. This 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 yields data of whether, when, and how often each subject checks the elapsed time, necessary to detect strategic behavior. It also 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.

Incentives

As explained above, in each round of the experiment we financially incentivize subjects either only for their estimation accuracy or both for their estimation accuracy and performance speed. Estimation accuracy earnings are determined by the absolute difference between the actual task duration and the estimate. The maximum earnings from a precise estimate are AUD 20. The accuracy earnings decrease 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). 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.[11] 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.

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

In rounds involving speed incentives, subjects receive additional earnings based on how quickly they 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

[11] 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.

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. Given the slopes of earnings functions, the accuracy incentives become stronger than speed incentives 173 seconds into the task execution. To be more precise, after 173 seconds 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 201 seconds.

Procedures

The experiment was conducted in the MGSM Vernon L. Smith Experimental Economics Laboratory at the Macquarie Graduate School of Management 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 two incentive structures 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 estimation accuracy earnings as well as performance speed earnings. 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 5, the subject was reminded that the task consisted of 5 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. Furthermore, as a manipulation check, the subjects were asked to indicate on a scale from 1 to 5 the degree to which they agreed with the statement “I estimated what I thought was the fastest time I could achieve”. Finally, subjects filled out a demographical questionnaire.

At the end of the experiment, subjects privately and individually received their experimental earnings in cash in the control room at the back of the laboratory. 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

The incentive structure A 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 for missing one or more letters “S” and having to systematically go through the entire table again. The incentive structure B also motivates accurate estimation, but at the same time incentivizes subjects to finish the task fast. We thus hypothesize that subjects will strategically provide higher estimates under incentive structure A than under structure B. We further hypothesize that the estimates will not be significantly different in the within-subject comparison of Round 1 and Round 3 in which subjects face the same incentive structure. However, we note there could exist two types of learning effects. On one hand, subjects may find ways to solve the tables more efficiently and opt to use lower estimates in Round 3. On the other hand, subjects may realize that it is beneficial to inflate their estimates more, in which case they may opt to use higher estimates in Round 3.

- *Hypothesis 1: The estimates under the incentive structure A (accuracy incentives only) will be higher than the estimates under the incentive structure B (both accuracy and speed incentives).*

An analogous logic applies to our prediction regarding the actual task duration. Since subjects can decide when exactly to finish the task and the imposed accuracy incentives are stronger relative to the speed incentives, we expect subjects to adjust their work pace to finish the task close to their estimates. Thus, the hypothesis related to the actual task duration parallels the hypothesis for duration estimates.

- *Hypothesis 2: Subjects will take longer to finish the task under the incentive structure A than under the incentive structure B.*

Our final hypothesis relates specifically to strategic pacing of work to match the actual task duration with the estimate, which should be especially pronounced in the last (fifth) table. In particular, we hypothesize that subjects will (on purpose) take longer to solve the fifth table than to solve any of the first four tables, on average. We further hypothesize that spending more time than necessary to complete the task will be detectable in all rounds of both conditions, but especially under the incentive structure A, where there is no trade-off between the performance speed and estimation accuracy.

- *Hypothesis 3: In each round, subjects will take longer to solve the fifth table than to solve an average table from the first four tables.*

5. Results

A total of 119 subjects participated in the experiment. Data of three subjects were excluded from the analysis due their misunderstanding of the instructions, limiting our sample to 116 subjects (45 females) with a mean age of 23.8 and a standard deviation of 6.1 years.[12] Of these 116 subjects, 59 were randomly assigned into the ABA condition (four sessions) and 57 into the BAB condition (four sessions). On average, an experimental session lasted around 60 minutes including the initial instructional period and payment of subjects. The subjects earned AUD 22.90 on average. Summary statistics by conditions and rounds are presented in Table 1.

[12] These three subjects repeatedly provided an answer different from 5 when asked about the completion time of how many tables they were estimating.

Table 1: Summary statistics by conditions and rounds

	Summary statistics – means (SD)					
	Condition ABA (N=59)			Condition BAB (N=57)		
	Round 1 (A)	Round 2 (B)	Round 3 (A)	Round 1 (B)	Round 2 (A)	Round 3 (B)
Estimates	421 (401)	315 (85)	329 (98)	294 (185)	347 (126)	318 (104)
Actual duration – all tables	381 (199)	318 (85)	343 (95)	351 (98)	336 (89)	312 (77)
Actual duration – the 5th table	105 (119)	83 (53)	99 (61)	74 (38)	91 (55)	80 (48)
Inaccuracy (Estimation errors)	155 (301)	29 (36)	25 (34)	114 (116)	39 (84)	42 (80)
Clicks on "Update" button	6 (11)	3 (7)	5 (9)	2 (4)	5 (11)	2 (4)
Time spent estimating	50 (42)	22 (16)	14 (10)	49 (35)	18 (12)	19 (17)
Number of incorrect answers	3 (3)	2 (3)	2 (3)	3 (3)	2 (2)	2 (2)

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

Consistent with our hypotheses, subjects in both conditions systematically provide higher estimates under the incentive structure A (when being incentivized only for estimation accuracy) than under the incentive structure B (when being incentivized for both estimation accuracy and performance speed). With the exception of Round 1 in BAB, subjects also take longer to complete the task under the incentive structure A. Since the same trend – subjects being slowest overall in Round 1 – holds for the ABA condition too, the exception can be attributed to the learning effect.

To examine the effects of incentive structures without the confounding learning effects, we conduct linear fixed-effect regressions (presented in Table 2) on estimates, actual duration and estimation (in)accuracy. The regression confirms that the incentive structure A results in significantly higher estimates and significantly longer actual task duration.

- Result 1: Subjects provide higher estimates and complete the task slower when the accuracy is incentivized but the performance speed is not.

Table 2: Fixed-effects OLS models

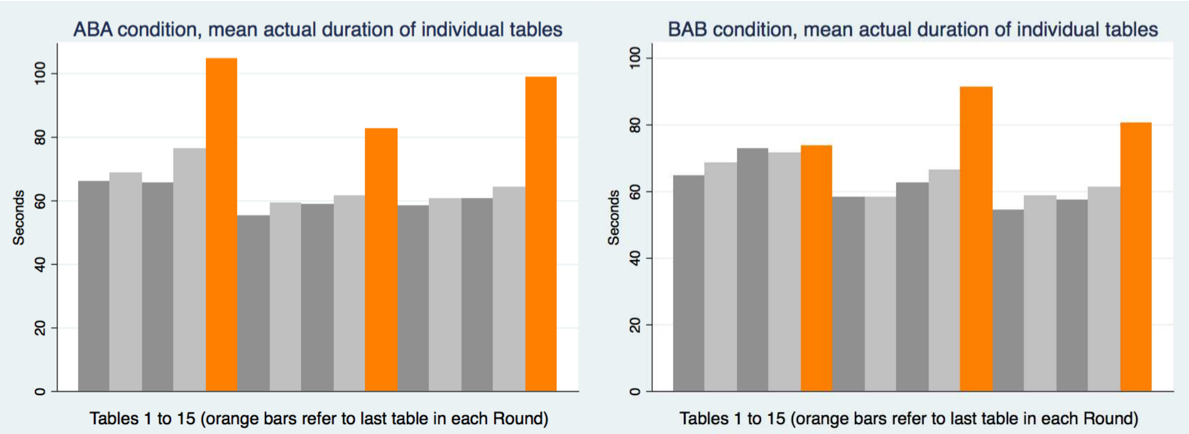
Dependent Variables	(1) Estimates	(2) Actual duration	(3) Accuracy (Estimation error)	(4) Time spent on the 5th table	(5) "Update" button clicks (the 5th table)
Incentive structure A	50.77***	24.32***	10.76	14.98***	2.28***
	(13.08)	(7.07)	(10.65)	(5.00)	(0.75)
Round 2	-27.64	-38.80***	-101.09***	2.95	0.61
	(25.56)	(10.70)	(20.30)	(6.53)	(0.84)
Round 3	-35.35	-38.28***	-101.59***	6.45	0.61
	(26.51)	(11.21)	(20.51)	(6.16)	(0.56)
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*
	(13.31)	(5.52)	(9.64)	(54.36)	(0.56)
N	348	348	348	348	348

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

The regression model (3) in Table 2 also shows that the coefficient on Incentive structure A is not significant, meaning that subjects are similarly accurate in their estimation under both incentive structures. Do subjects achieve accuracy by deliberately prolonging the task execution? After receiving feedback regarding their actual task duration in Round 1, subjects in both the ABA and BAB conditions significantly improve their estimation accuracy in the following two rounds. The median estimation error across all subjects in Rounds 2 and 3 is only 14 seconds, which is approximately 4% off the estimate. Low estimation errors are indeed reached by taking more time than necessary to complete the task, which is most prominent in the last table in each round. With one exception, subjects in all rounds of both conditions take significantly longer to solve the last (fifth) table compared to the average time spent on solving one table out of the first four (Wilcoxon matched-pairs signed-rank test for the ABA condition yields a p-value = 0.01 in Round 1, <0.01 in Round 2 and <0.01 in Round 3, for the BAB condition yields a p-value = 0.45 in Round 1, <0.01 in Round 2 and <0.01 in Round 3; see also Figure 2 for the mean duration for each table across rounds).[13]

- Result 2: Subjects 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 2: Mean actual duration of individual tables



Notes: Figure 2 displays bar charts of mean duration of all 15 tables in chronological order in the ABA condition (left panel) and in the BAB condition (right panel). The duration of the last table of each round (i.e., the 5th, 10th and 15th table) is highlighted in orange.

[13] The only exception is Round 1 in the BAB condition, in which majority of subjects try to be as fast as possible. This is because subjects in the BAB condition often underestimate the task duration in Round 1, while overestimation is more prevalent in Round 1 of the ABA condition. Such difference in bias direction (which is statistically significant, Mann-Whitney test p-value = 0.03) is consistent with earlier studies on speed incentives reviewed in Section 2. However, after receiving the Round 1 feedback, any systematic tendencies to underestimate or overestimate the task duration disappear.

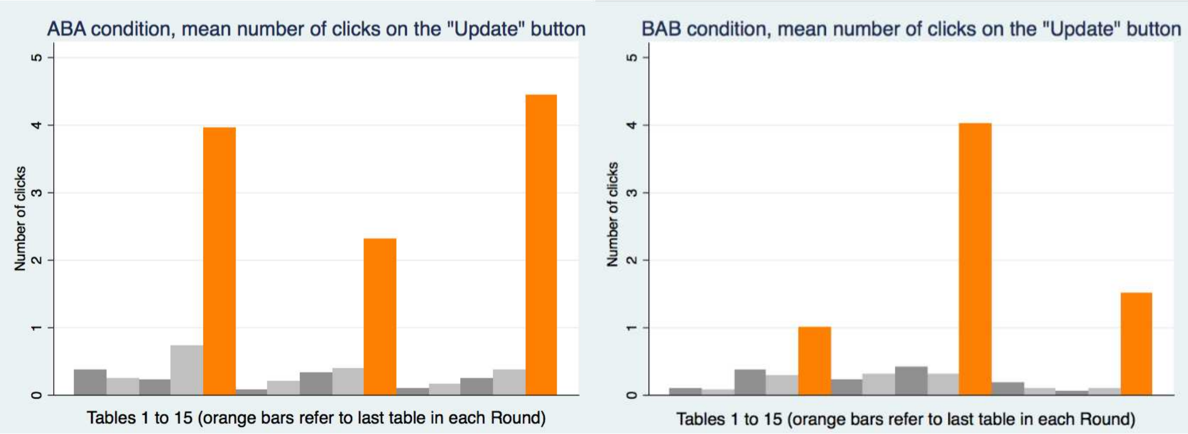
Auxiliary analysis

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. For the time spent on estimation, there are no differences across conditions in Round 1 (Mann-Whitney test, p-value = 0.81). However, in the following two rounds it takes subjects who face the incentive structure B significantly longer to provide their estimates than those who face the incentive structure A (Mann-Whitney test, p-value = 0.03 in Round 2; and 0.03 in Round 3). In the "Update" button clicks analysis, we find that subjects facing the incentive structure A consistently check time more often, with the differences being statistically significant across conditions in two out of three rounds (Mann-Whitney test, p-value = 0.03 in Round 1; 0.89 in Round 2; and 0.02 in Round 3). Number of clicks on the "Update" button for all 15 tables, by conditions, is displayed in Figure 3.

Finally, we analyze responses to the question in which subjects indicated to what degree they agree with the statement "I estimated what I thought was the fastest time I could achieve". We find that subjects in the ABA condition (that includes only one round in which the speed is incentivized and ends with the incentive structure A) agree significantly less than subjects in the BAB condition (Mann-Whitney test, p-value = 0.02). This auxiliary analysis provides further evidence that subjects understand the incentives they are facing and act in their best interest.

- Result 3: Subjects take longer to figure out their estimates when facing the two types of incentives simultaneously. Once the task is underway, subjects check the elapsed time more frequently when being rewarded only for estimation accuracy.

Figure 3: Mean number of clicks on the "Update" button



Notes: Figure 3 displays bar charts of mean number of clicks on the "Update" button for all 15 tables in chronological order in the ABA condition (left panel) and in the BAB condition (right panel). Clicks on the last table of each round (i.e., the 5th, 10th and 15th table) are highlighted in orange.

Within-subject analysis

Using the Wilcoxon matched-pairs signed-rank test, we analyze how subjects respond to a change in the incentive structure from one round to another (see Table 3). Consistent with the earlier analysis and also with our hypotheses, we find that subjects increase their estimates when facing the incentive structure A, and decrease the estimates when facing the incentive structure B. However, only the changes in the BAB condition are statistically significant. The subjects in the ABA condition seem to behave more conservatively. We speculate that the differences in estimates of the same subjects across rounds are more pronounced in the BAB condition because most of the subjects in that condition completed the task in Round 1 as fast as they could (as apparent from Figure 2) and thus received a good signal about their (maximum) potential speed. On the other hand, subjects in the ABA condition had no urge to work quickly in Round 1 and thus many of them did not learn how quickly they could complete the task in subsequent rounds.

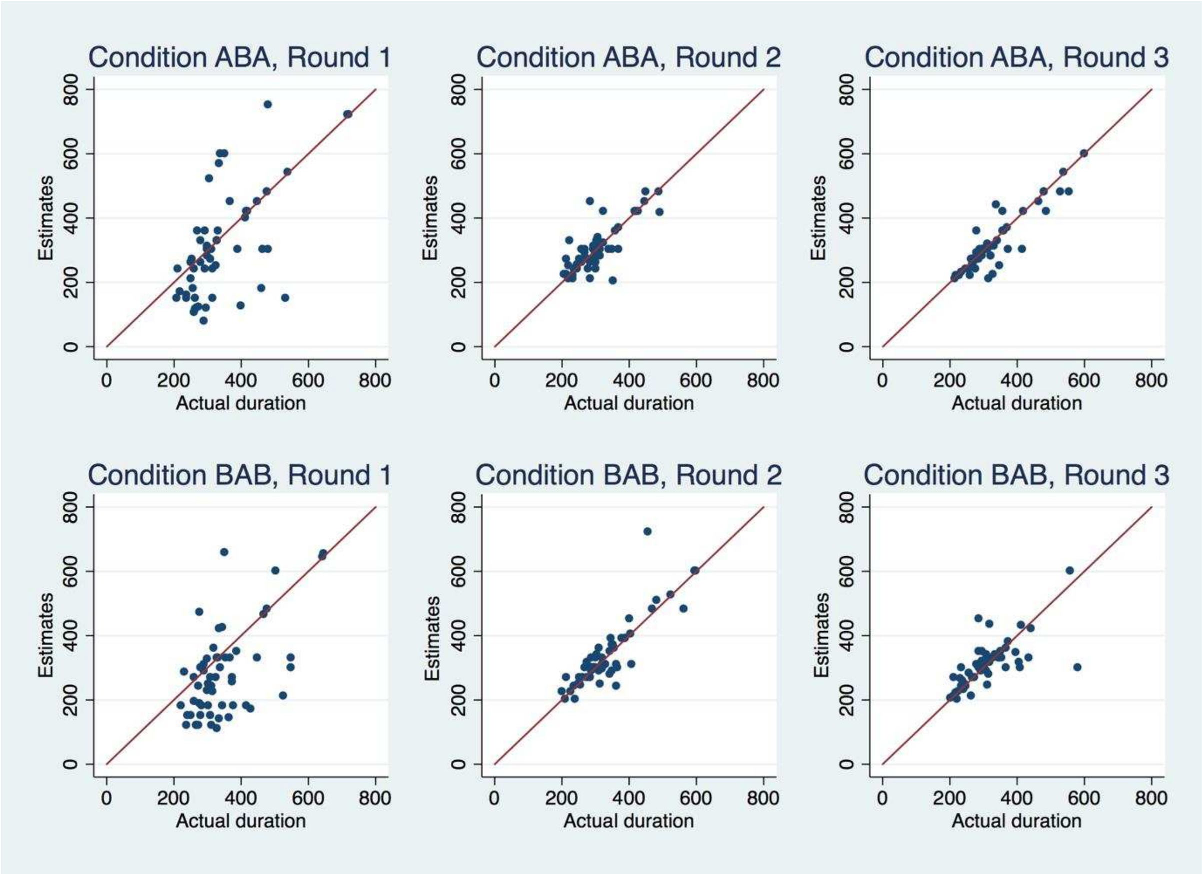
Table 3: Within-subject non-parametric tests

	Wilcoxon matched-pairs signed-rank test (p-values), direction											
	Condition ABA (N=59)						Condition BAB (N=57)					
	R1 vs. R2		R2 vs. R3		R1 vs. R3		R1 vs. R2		R2 vs. R3		R1 vs. R3	
Estimates	0.26	▽	0.13	△	0.64	▽	0.00	▲	0.00	▼	0.06	△
Actual duration – all tables	0.00	▼	0.00	▲	0.08	▽	0.04	▼	0.01	▼	0.00	▼
Actual duration – the 5th table	0.15	▽	0.01	▲	0.46	▽	0.07	△	0.09	▽	0.35	△
Inaccuracy (Estimation errors)	0.00	▼	0.08	▽	0.00	▼	0.00	▼	0.46	△	0.00	▼
Clicks on "Update" button	0.02	▼	0.01	▲	0.88	▽	0.05	▲	0.01	▼	0.85	▽
Time spent estimating	0.00	▼	0.00	▼	0.00	▼	0.00	▼	0.60	△	0.00	▼
Number of incorrect answers	0.18	▽	0.32	△	0.95	▽	0.00	▼	0.85	▽	0.01	▼

Notes: Table 3 lists p-values of within-subject statistical tests by conditions. The directions of the changes are marked with triangles (normal triangles for increases, and inverted ones for decreases). For statistically significant changes, the triangles are solid.

Since subjects can monitor the elapsed time and thus finish the task at the desired moment, their actual task duration is often close to their estimate (the scatterplots of estimates and actual task duration at the individual level are displayed in Figure 4). The findings regarding the actual task duration parallel those for estimates, except for Round 1 in the BAB condition. We find that subjects in both conditions complete the tasks faster from Round 1 to Round 2 (probably due to them finding ways on how to be more efficient in task execution). The subjects in the BAB condition are even faster in Round 3, while the subjects in the ABA condition, having no incentives to work quickly, execute the task in Round 3 slower than in Round 2.

Figure 4: Within-condition and within-round comparison of estimates and actual duration



Notes: Figure 4 displays scatter plots of individual-level estimates (vertical axis) and actual duration (horizontal axis) by condition and round. Precise estimates are on the red 45-degree line. A dot above the red line indicates overestimation, while a dot below the red line indicates underestimation. For presentational clarity purposes, six outlier data points were removed from the figure. These outliers were recorded in Round 1 (five of them in the ABA condition and one in the BAB condition) and were driven by extremely high estimates of between 959 and 1810 seconds. Two of these estimates (959 and 1500 seconds, both in the ABA condition) were matched with similarly high actual duration. The other four subjects finished Round 1 in the time similar to the remaining subjects, whose estimates were not outliers.

Robustness

To verify the robustness of our findings, we conduct a regression analysis (presented in Table 4), controlling for risk attitudes (number of less risky choices in incentivized risk assessment), time spent completing the practice table, cognitive reflection, and demographics – age, gender, current degree of study (first year, second year, third/fourth year, honors, masters, doctoral), employment status, frequency of playing video games (from 1 - never to 5 - all the time).[14] We do not find any systematic effect other than the condition manipulation, with one exception. The estimation errors are negatively

[14] We collected data on the frequency of playing video games to control for potential differences in intrinsic ability to complete the task quickly. Our task required to click on the screen approximately 250 times within a couple of minutes in each round, which could have been easier for those who frequently play video games.

associated with cognitive reflection, meaning that those who scored higher in the cognitive reflection test also estimated more accurately in all three rounds (statistically significantly in Round 1 and Round 3). Thus, in line with our conjecture, subjects with a higher CRT score behave more strategically in the task estimation and execution.

Table 4: Regression analysis (OLS model)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dependent variable	Estimate R1	Estimate R2	Estimate R3	Actual dur. R1	Actual dur. R2	Actual dur. R3	Est. error R1	Est. error R2	Est. error R3
Condition ABA	132.29**	-52.06***	24.70**	-3.65	-11.55	17.64	48.57	-11.41	-17.53
	(59.57)	(18.50)	(10.00)	(16.81)	(9.17)	(12.71)	(41.04)	(13.97)	(13.03)
Age	-0.67	-2.50*	0.59	3.37	0.28	1.61**	-2.96	-1.77*	-0.63
	(7.78)	(1.42)	(1.58)	(2.38)	(0.97)	(0.76)	(5.26)	(0.96)	(0.86)
Female	-77.26	-7.70	-21.38	12.43	6.41	-0.45	-62.88	-35.85*	-23.51
	(67.29)	(20.99)	(15.13)	(25.66)	(13.23)	(10.54)	(48.03)	(18.91)	(17.32)
Current degree of study	39.52	16.03**	-2.07	-19.48	-2.60	-1.25	23.29	11.24*	7.23
	(26.90)	(6.59)	(4.95)	(12.88)	(4.31)	(3.53)	(17.26)	(6.02)	(5.14)
Freq. of playing video games	-33.95	-12.95	-11.57*	8.00	0.15	1.37	-47.70**	-12.39	-6.33
	(30.59)	(9.39)	(6.48)	(13.84)	(5.68)	(5.01)	(20.99)	(8.79)	(7.63)
Not employed	22.14	-5.07	-0.02	1.97	5.11	1.47	28.24	-6.52	-3.41
	(32.13)	(8.89)	(5.49)	(14.17)	(5.64)	(4.74)	(23.19)	(6.77)	(6.25)
Cognitive reflection	-30.87	-5.48	0.59	14.17	2.63	-3.34	-54.99***	-5.11	-11.58**
	(24.84)	(6.51)	(3.85)	(9.92)	(3.96)	(3.93)	(18.39)	(5.80)	(5.42)
Practice table duration	-0.24	0.24	0.11	1.32***	0.23	0.07	-0.33	0.02	-0.10
	(0.72)	(0.22)	(0.12)	(0.38)	(0.15)	(0.14)	(0.42)	(0.16)	(0.13)
Risk attitudes	-11.89	3.14	0.01	-0.65	1.65	2.04	-7.45	1.11	-0.77
	(16.70)	(3.50)	(2.02)	(5.72)	(2.57)	(1.96)	(13.43)	(2.09)	(1.51)
Estimate R1		0.06	0.00	0.29**	0.02	-0.03			
		(0.05)	(0.02)	(0.11)	(0.03)	(0.02)			
Actual dur. R1		0.36***	0.15***		0.17**	0.12**			
		(0.11)	(0.06)		(0.08)	(0.05)			
Estimate R2			0.67***		0.41**	-0.30*			
			(0.18)		(0.18)	(0.16)			
Actual dur. R2			-0.12			0.32**			
			(0.20)			(0.13)			
Estimate R3						0.62***			
						(0.14)			
Constant	462.97**	236.81***	112.91*	75.40	79.39*	26.19	416.68***	139.00*	118.57
	(210.18)	(75.57)	(64.71)	(90.51)	(40.61)	(39.13)	(135.90)	(77.10)	(72.46)
N	116	116	116	116	116	116	116	116	116
R^2	0.14	0.51	0.74	0.43	0.65	0.68	0.19	0.14	0.13

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

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 might 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 performance speed incentives are implemented alongside the estimation accuracy incentives, the estimates are less inflated, and the task is performed more quickly. Importantly, the estimation accuracy is not compromised.

Naturally, there is a 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?").* The hesitance to fully exploit the imposed incentive structure is in line with the previous literature documenting that when given a chance, most people act dishonestly, but the magnitude of such behavior is far away from the maximum (e.g., Fischbacher & Föllmi-Heusi, 2013; Mazar, Amir, & Ariely, 2008). In addition, 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 nevertheless provide unambiguous empirical evidence that incentivizing project planners for the accuracy of project estimates can induce moral hazard, resulting in hidden inefficiency of resource underutilization. The inefficiency represents a valid concern from the perspective of operational effectiveness of 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, (Lorko et al., 2020) to discourage wasting time and other resources only to deliver the project exactly "on time".

While our study demonstrates the positive effects of utilizing speed incentives in project management, such incentives inevitably result in more costly project execution. One limitation of our study is that we do not incorporate a trade-off between additional costs (due to speed incentives) and additional revenues (due to time savings). In order for the speed incentives to be efficient, the extra expenses need to be lower than what can be gained from faster project completion. A demonstration 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 renumeration 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 might be able to mitigate the inefficiency by rejecting the project schedules that appear inflated. The hidden inefficiency can also be reduced by introducing competition amongst potential project contractors, although such an option might 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.

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 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 who provided 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 Macquarie University for providing financial support for this project.

References

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

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037/0033-295X.84.2.191

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., & 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

Byram, S. J. (1997). Cognitive and motivational factors influencing time prediction. *Journal of Experimental Psychology: Applied*, *3*, 216–239. https://doi.org/10.1037/1076-898X.3.3.216

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

Camerer, C. F., Dreber, A., Forsell, E., Ho, T. H., Huber, J., Johannesson, M., … Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. *Science*. https://doi.org/10.1126/science.aaf0918

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

Darby, M. R., & Karni, E. (1973). Free Competition and the Optimal Amount of Fraud. *The Journal of Law and Economics*. https://doi.org/10.1086/466756

Dulleck, U., & Kerschbamer, R. (2006). On doctors, mechanics, and computer specialists: The economics of credence goods. *Journal of Economic Literature*. https://doi.org/10.1257/002205106776162717

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

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

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

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

Halkjelsvik, T., Jørgensen, M., & Teigen, K. H. (2011). To read two pages, I need 5minutes, but give me 5minutes and I will read four: How to change productivity estimates by inverting the question. *Applied Cognitive Psychology*, *25*(2), 314–323. https://doi.org/10.1002/acp.1693

Henry, R. A. (1994). The Effects of Choice and Incentives on the Overestimation of Future Performance. *Organizational Behavior and Human Decision Processes*, *57*(2), 210–225. https://doi.org/10.1006/obhd.1994.1012

Henry, R. A., & Sniezek, J. A. (1993). Situational factors affecting judgments of future performance. *Organizational Behavior and Human Decision Processes*, *54*(1), 104–132. https://doi.org/10.1006/obhd.1993.1005

Henry, R. A., & Strickland, O. J. (1994). Performance Self-Predictions: The Impact of Expectancy Strength and Incentives. *Journal of Applied Social Psychology*, *24*(12), 1056–1069. https://doi.org/10.1111/j.1559-1816.1994.tb02373.x

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

Jensen, M. C. (2003). Paying People to Lie: The Truth about the Budgeting Process. *European Financial Management*, *9*(3), 379–406. https://doi.org/10.1111/1468-036X.00226

Jorgensen, M., & Sjoberg, D. I. K. (2001). Impact of effort estimates on software project work. *Information and Software Technology*, *43*(15), 939–948. https://doi.org/10.1016/S0950-5849(01)00203-8

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, *108*(3), 480–498. https://doi.org/10.1037/0033-2909.108.3.480

Latham, G. P., & Locke, E. A. (1991). Self Regulation through Goal Setting. *Organizational Behavior and Human Decision Processes*, *50*(April), 212–247. https://doi.org/10.5465/AMR.1991.4278976

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

Lederer, A. L., & Prasad, J. (1998). A causal model for software cost estimating error. *IEEE Transactions on Software Engineering*, *24*(2), 137–148. https://doi.org/10.1109/32.666827

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. (2020). *Improving the accuracy of project schedules*. Working paper.

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*, *45*(6), 633–644. https://doi.org/10.1509/jmkr.45.6.633

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*. https://doi.org/10.1126/science.aac4716

Osband, K. (1989). Optimal Forecasting Incentives. *Journal of Political Economy*. https://doi.org/10.1086/261644

Plott, C. R. (1987). Dimensions of parallelism: some policy applications of experimental methods. In *Laboratory Experimentation in Economics*. Cambridge University Press.

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

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

Smith, V. L. (1962). An Experimental Study of Competitive Market Behavior. *Journal of Political Economy*. https://doi.org/10.1086/258660

Smith, V. L. (1991). Rational Choice: The Contrast between Economics and Psychology. *Journal of Political Economy*, *99*(4), 877–897. https://doi.org/10.2307/2069710

Instructions

Thank you for coming. Please note that the use of watches, mobile phones and any other devices that show time is not allowed during this experiment. Instead, a time measuring tool will be provided to you on the screen. Also, please note that from now on, until the end of the experiment, no talking or any other unauthorized communication is allowed. If you violate any of these rules, we will have to exclude you from the experiment and from all payments. If you have any questions after you finish reading the instructions, please raise your hand. The experimenter will approach you and answer your questions in private.

Please read the following instructions carefully. The instructions will explain how you can earn money in this experiment. Your decisions and earnings will not be revealed to other participants.

Three rounds of the same task

The experiment consists of three rounds. In each round, your task will be to complete 5 tables. Each table consists of 100 cells that contain either the letter “S” or the number “5”. All tables are different from each other. In each table, there are between 45 and 55 cells containing the letter “S”, while the rest of the cells contain the number “5”.

In order to complete each table, you will have to check all cells containing the letter “S”, while leaving the other cells unchecked. Then, you will click “Submit”. You will move on to the next table, only if you correctly complete the current table. The software does not uncheck the cells after an incorrect submission.

There is no limit on how many times you can submit each table. Also, there is no financial penalty for an incorrect submission.

Practice table

In order to become familiar with the task, you will first complete 1 (one) practice table. The practice table is similar to the 5 tables you will be completing in each of the three rounds. Your performance in the practice table has no effect on your earnings.

Time estimation

At the beginning of each round, you will be asked to estimate how long it will take you to complete the upcoming task. That is, you will estimate how long it will take you to correctly complete **all 5 upcoming tables together**.

Information box

During your work on the task, an information box will be displayed in the upper left corner of your screen. This box will show the following:

- The current table (out of 5)
- Your time estimate
- How much time you have already spent on the task. This will be updated every time you complete the current table and move on to the next table. At any time, you can also update it manually by clicking on the “Update” button.

Please note that the information box will not be displayed for the practice table.

Earnings

In this experiment, there are two types of earnings: earnings based on the accuracy of your time estimate and earnings based on your performance speed in the task.

Time estimation accuracy earnings

Your time estimation accuracy earnings (in AUD) will be calculated according to the following formula:

$$\text{Time estimation accuracy earnings } = 20 - \; 0.10 * |\text{actual time in seconds } - \text{time estimate in seconds}|$$

Your time estimation accuracy earnings depend on the absolute difference between the actual time it takes you to complete the 5 tables and your time estimate. Notice that the more accurate your time estimate is, the more money you earn. An exact time estimate will earn you AUD 20, but any inaccuracy will reduce your earnings. If the formula returns negative time estimation accuracy earnings, the software will set these earnings to 0. This means that if your time estimate is inaccurate by 200 seconds or more, your time estimation accuracy earnings will be zero.

Performance speed earnings

Your performance speed earnings (in AUD) will be calculated according to the following formula:

$$\text{Performance speed earnings } = \frac{3000}{\text{actual time in seconds}}$$

Your performance speed earnings depend on the actual time it takes you to complete the task. Notice that the faster you complete the task (i.e., correctly complete all 5 tables), the more money you earn.

Earnings structure changes from one round to another

In each of the three rounds, you can earn money based on either:

- both the accuracy of your time estimate and your performance speed

or

- only the accuracy of your time estimate.

At the beginning of each round, the software will inform you about the earnings structure valid for the given round. Your earnings for the round will be either the sum of your time estimation accuracy earnings and performance speed earnings or only your time estimation accuracy earnings.

At the end of the experiment, the software will randomly select one of the three rounds for payment. Each round has the same probability to be selected. You will be paid based on your earnings achieved in the selected round only.

When you finish

After you complete the third round, you will be asked to answer a few questions about the experiment. The final screen will display the summary of your earnings. When you finish the experiment, please stay quietly seated in your cubicle until the experimenter calls your cubicle number. You will then go to the room at the back of the laboratory to privately collect your experimental earnings in cash. You need to complete the entire experiment in order to get paid. If you have any questions, please raise your hand.