Missing the forest for the trees: when monitoring quantitative measures distorts task prioritization*

Eldar Dadon Ro'i Zultan†

September 14, 2023

Abstract

Managers' use of remote monitoring software increased following the transition to working from home during the Covid-19 pandemic to compensate for reduced observability. Higher observability entails quantitative measures not directly related to productivity, potentially incentivizing workers to prioritize quantity over quality. For example, office workers may increase observable work hours by directing effort inefficiently. Observing the number of completed tasks incentivizes workers to perform many meaningless tasks rather than prioritize productive ones. We design an experiment where workers can allocate effort based on perceived task difficulty and manipulate the structure of the signal to the manager. We show theoretically that quantitative information in the signal distorts incentives. In equilibrium, workers prioritize productive tasks less, reducing overall productivity. Our results confirm that removing quantitative information from the signal increases productivity by shifting workers' strategies. Enriching the signal with quantitative information, however, does not have the opposite effect.

Keywords: incentives, multitasking, working from home, laboratory experiment
JEL classification: C92, D86, J30

*We thank Or Avishay-Rizi, Fabio Galleoti, Tomer Ifergane, David Lagziel, Marie Claire Villeval, Jonathan Rosenblatt, Hadas Stiassnie, Adam Zylbersztejn, and participants at the 2019 Economic Science Association European meeting in Dijon, the 5th Coller Conference on Behavioral Economics in Tel Aviv, and GATE-Lyon for helpful discussion and suggestions. We thank Yael Cohen, Tom Shefer and Reut Goldman, Rony Natan, Gali Goldberg, and Dor Ahady for assistance in conducting the experiments. The study was approved by the Ben-Gurion University of the Negev ethics committee. Dadon thanks the Chateaubriand fellowship and the National Insurance Institute of Israel for financial assistance. Zultan acknowledges financial assistance from the Israeli Science Foundation, grant No. 397/21.

†Dadon: Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel. E-mail: Eldarda@post.bgu.ac.il, Tel: +972(0)548131619. Zultan: Ben-Gurion University of the Negev, P.O.B. 653, Beer-Sheva 84105, Israel. E-mail: zultan@bgu.ac.il, Tel: +972(0)86472306.

1. Introduction

The design of optimal incentives for workers poses a key challenge for managers. While individual productivity is readily measurable in some tasks, measuring workers' effort and output directly and in real time is typically unfeasible. Furthermore, observed productivity often depends on myriad factors, making the worker's marginal contribution impossible to identify.[1] Consequently, managers face difficulties designing compensation schemes that properly align workers' incentives with the managerial objectives.

Overall production is a function of both quantity and quality. Often, quality is hard to assess, and managers turn to form performance evaluations based on readily available quantitative indicators such as number of tasks performed. Remuneration based on observable quantitative measures may, however, distort incentives. Workers aiming to complete many tasks will invest less effort per task, thereby reducing quality. For example, software developers judged by the number of completed tasks may make shortcuts that technically close tickets but compromise code quality, requiring rework down the road. Generally, relying on measurable aspects can divert effort away from other important but unobservable objectives and result in an overall reduction in productivity (Baker, 1992, 2002; Gibbons, 2005; Holmstrom and Milgrom, 1991). Even if managers are aware that quantitative measures such as the number of completed tasks are invalid, they may be unable to ignore them. Research in cognitive psychology shows that people find it challenging to ignore irrelevant ("non-diagnostic") information, in particular in labor hiring decisions (Carr et al., 2017; Dalal et al., 2020) or performance assessment (Moore et al., 2010).

Thus, it is possible that providing managers with *richer* information leads to *lower* productivity, as the additional information draws attention from more valid—though difficult to measure—indicators of productivity. For example, when workers work in the office, managers can observe many aspects of work. Some quantitative aspects, such as the work hours, are highly salient in the office but are not directly linked to productivity. In contrast, when workers work from home, managers are less able to observe direct measures of worker performance, and rely more on "bottom line" productivity outcomes.[2] Some empirical studies corroborate this conjecture, showing a positive effect of working from home on employees' productivity (Angelici and Profeta, 2020; Bloom et al., 2015; Choudhury et al., 2021). Working from home became the reality of a substantial part of the labor market following the outbreak of the Covid-19 pandemic at the end of 2019 (Gartner, 2020), with recent surveys indicating that productivity increased as a result (Barrero et al., 2021; DeFilippis et al., 2022). Nonetheless, many managers utilize employee monitoring software to compensate for the reduced observability (Kalischko and Riedl, 2021; Trivedi and Patel, 2021).

In this paper, we test the role of incorporating quantitative information into the signal

[1]We use the term "productivity" to refer to the worker's output in terms of its *value to the manager*, as opposed to some quantitative measure of output.

[2]In-office work offers additional advantages beyond the scope of this analysis, including the facilitation of knowledge sharing among workers and the nurturing of mentorship relationships between senior and junior employees.

observed by the manager. We assume that the manager observes a noisy signal composed of qualitative and quantitative factors. We hypothesize that shutting down the quantitative element—thereby providing less information in the signal—will focus effort on quality, resulting in an increase in overall productivity.

We focus on effort allocation decisions when tasks arrive sequentially. For each task, the worker decides whether to exert effort and complete it successfully, or to skip it by exerting the minimal effort required to mark the task as technically completed and move to the next task. To maximize productivity, the worker should prioritize tasks offering higher returns per time invested and skip less efficient tasks. We hypothesize that when the manager has access to quantitative information—in this case, the *number* of completed tasks—the worker will not only skip more tasks, but will also be less selective in which tasks to direct effort to, leading to higher quantity but overall lower productivity.

We present a simple theoretical model formalizing this argument and test its predictions in a controlled laboratory experiment. Laboratory experiments are uniquely suited to study productivity under moral hazard and signaling. Testing our hypotheses requires measuring effort and productivity independent of the signal seen by managers. This is typically infeasible in natural work environments, where researchers and managers face the same limitations in monitoring workers. Our experimental environment overcomes this obstacle by explicitly manipulating the signal available to the managers while fully tracking the worker's performance.

Our laboratory environment precludes various features of real-world work environments, such as interactions between workers. By intentionally abstracting from certain real-world elements, our focus remains on the incentivizing effects of observability and signal structure. This methodological approach helps minimize potential confounding factors that could otherwise obscure our findings. On the other hand, caution should be exercised when generalizing our conclusions to natural work settings, fully taking into account all aspects absent from the experimental environment, such as distinctions between office and home-based work.

We implement a principal-agent setting with repeated interactions, noisy monitoring, and incomplete contracts. In each period, the manager observes an imperfect signal of the worker's performance and provides discretionary wages. We manipulate whether the signal incorporates quantitative metrics. When the signal depends on quantity, workers artificially inflate the number of tasks completed. Removing the quantitative information from the signal decreases the volume and increases productivity, confirming that shifting effort to quantity comes at the expense of quality. When we add quantitative content to the signal, however, volume increases but productivity does not decrease.

We analyze workers' strategies using a finite mixture model and machine learning-based task difficulty estimates. In the baseline treatment, strategies align well with the manager's interests. However, when quantitative information feeds into the signal, workers' strategies shift away from productivity-maximizing strategies and towards strategies that favor quantity.

The findings contribute to the literature on incentive design and optimal task prioritization when principals can only observe or reward partial or imperfect metrics of agent effort. We assume that managers form a holistic impression of workers. When managers observe quantitative production indicators, workers adapt by shifting resources to increasing quantity. This shift comes at the expense of quality and leads to reduced efficiency. Less monitoring, as happens when workers work from home, improves incentives and increases productivity. This insight is essential in considering changes in labor and management, especially in the aftermath of the Covid-19 pandemic.

2. Related literature

The multitasking literature, beginning with Holmstrom and Milgrom (1991) and Baker (1992), consider situations in which a worker can invest independently in quality and in quantity. Holmstrom and Milgrom (1991) analyzed settings where workers engage in different activities, not all of which are observable and contractible. In characterizing the optimal contract, Holmstrom and Milgrom (1991) showed that incentivizing the observable actions can be counterproductive, as it draws effort from other activities and reduces overall productivity. For example, production workers may be responsible for both increasing output and maintaining equipment. If the manager can monitor quantity, paying by piece rate will lead employees to produce more output while overusing and damaging machines. Paying a global salary is thus superior to per-performance pay. While the theoretical literature is concerned with optimal contracts, our focus is on the information available to the manager. Nonetheless, applying a multitasking model to the question of signal structure confirms that increasing the observability of quantitative information may reduce overall productivity.[3]

Al-Ubaydli et al. (2015) tested the predictions of the Holmstrom and Milgrom (1991) model in a series of laboratory and field experiments where workers were hired to perform a short-term job. They find that incentivizing quantity reduces quality—unless there is ambiguity regarding the ability of the manager to assess quantity, piece rate can increase quality. The field experiment by Hong et al. (2018) provided further evidence with factory workers performing routine tasks. The workers knew that the quantity they produced was regularly measured, while quality was not individually recorded and could not be connected to a specific worker. Introducing piece-rate compensation for each unit above a fixed threshold resulted in an increase in productivity and a decrease in quality.

A few studies applied the multitasking framework in laboratory experiments. Fehr and Schmidt (2004) studied a principal-agent framework where the worker chooses effort levels in two tasks. The two tasks are complementary in the production function, but only the "quantity" task is contractible. The research question was whether—as quality is non-contractible—an implicit contract can outperform a piece-rate contract and whether managers prefer the implicit contract. The results confirmed the hypotheses. The managers mostly opted for the implicit bonus contract, which, in turn, yielded more efficient effort allocation.

Rubin et al. (2018) used a real effort task in which participants had to sum up five 2-digit numbers. Submitting a solution granted the participants with a given amount of money, regardless of the correctness of the solution. However, submitting a correct solution could gen-

[3]More details are available upon request.

erate extra payoffs for the participant based on the treatment group. The results showed that an increase in the payment for the correct solution increases the number of correct solutions at the expense of the total number of solutions submitted. In another experiment, Oosterbeek et al. (2011) showed that incentivizing an unproductive investment results in an increase in the unproductive investment and a decrease in the productive investment.

Several studies explicitly manipulated monitoring. Manthei and Sliwka (2019) found that providing bank branch managers with objective performance measures increased profits. In small branches, the increase in sales of investment products was offset by a decrease in the sales of (more profitable) loans. The authors attribute this branch-size effect to high specialization in large branches. In small branches, sales agents in small branches necessarily serve all customers. Providing better measures of the "fringe" task of selling investment products thus draws effort from the more profitable task of selling loans, which was already previously monitored. Dickinson and Villeval (2008) conducted a comprehensive laboratory experiment studying the effect of monitoring on agents' output in different circumstances. The main results are that monitoring increases worker performance. Interpersonal relationships between the worker and manager also increased performance, conditional on a direct and continuous link between performance and manager's payoffs. In that case, adding monitoring reduced performance.

In the modern work environment, removing observable quantitative measures by allowing workers to work from home may increase productivity. Bloom et al. (2015) provided evidence supporting this reasoning. Their field experiment showed that employees who worked from home were more efficient than employees who worked from the office. The improvement was due to spending more time working during the shift. We suggest that employees who worked from home could only signal their productivity by working. In contrast, office workers allocated part of their effort to other activities that create an appearance of work without substantially increasing productivity. Similarly, Angelici and Profeta (2020) found that flexible work hours and location increased worker productivity. Choudhury et al. (2021) found that working from anywhere was even more effective than working from home, reporting a productivity increase of 4.4%.

The outburst of the Covid-19 pandemic forced many employers to abandon their conservative habits, allowing more of their employees to work from home (Barrero et al., 2021; Bick et al., 2021; Brynjolfsson et al., 2020). According to a Gartner (2020) survey of 229 Human Resources departments conducted in April 2020, 50% of surveyed organizations reported 81% or more of their employees working from home during the pandemic, with many reporting that they are planning to work from home more in the future. A March 2021 survey by Barrero et al. (2021) estimated that employees supplied about 45 percent of paid labor services from home. Furthermore, respondents reported better-than-expected working from home experiences and higher productivity at home. The Covid-19 outburst set a new trend of how the future labor market will look and that these changes might positively affect the economy.

Thus, some studies on the transition to working from home find an increase in productivity. In contrast, Gibbs et al. (2023) found decreased per-hour productivity following the pandemic-induced transition to home work in an Asian IT company. While this result appears to be at odds with our analysis, a closer look reveals a conceptual consistency. The firm closely monitored the work from home using state-of-the-art monitoring applications on the working devices, collecting detailed information on work patterns and work hours (which made the research possible). Thus, the technological advances available to the firm studied in Gibbs et al. (2023) provided workers with more, rather than fewer, channels to signal effort after the transition. Consistent with this interpretation, the data show that, When working from home, workers spent more time participating in many short meetings, resulting in short uninterrupted working spells and longer overall work hours. This study illustrates that, at least for modern tech-savvy firms, transitioning to working from home can imply higher transparency and observability that distorts incentives. Our analysis provides a unifying explanation for both increased and decreased productivity when working from home, with the key variable being the observability of quantitative indicators.

3. Experiment

The experiment included twenty rounds in two blocks of ten rounds each. Each session involved an even number of participants and last for approximately 80 minutes. Participants were randomly assigned to roles of manager and worker, which remained fixed throughout the experiment. Participants interacted in fixed pairs of manager and worker for ten rounds, switching partners for the second block such that each manager interacted with a different worker in each block (and vice versa), and the same two managers interacted with the same two workers over the two blocks.[4] In each round, workers worked on a real-effort task. Specifically, the workers had 60 seconds to solve simple problems of adding three random two-digit numbers using only pen and paper (cf. Niederle and Vesterlund, 2007). The worker submitted an answer for each problem and proceeded immediately to the next problem without feedback until the allotted time ran out. Each correct answer earned the manager 10 ECU (Experimental Currency Units).[5] After the work phase ended, the manager observed a noisy signal of the worker's productivity and chose a bonus to pay the worker between zero and 60.

The signal structure depended on the treatment. We use the labels OFFICE and HOME for the treatment where the signal reflects both quality and quantity and the treatment where the signal only reflects qualitative information, respectively.[6] In both treatments, the manager observed the true Number of Correct Answers (*NCA*) submitted by the worker with probability 0.5 and otherwise observed a random number drawn from a uniform distribution on the

[4]Thus, each matching group of two managers and two workers constitutes an independent observation. If the number of participants in the session was not a multiple of four, there was one matching group of three managers and three workers.

[5]To avoid boredom during the work phase, managers could earn additional money by repeatedly clicking on a sphere that reappeared at a random location on-screen after each click. Every ten clicks added 1 ECU to the manager's round payoff. Similar tasks were used in Corgnet et al. (2015) and invoked in Gerstenberg et al. (2023) and Zultan et al. (2012).

[6]We do not claim that our manipulation captures all differences between office and home work, and only use these labels for convenience.

integers in $[1, X]$. The treatments differed in whether the distribution of the noise, determined by the upper bound x, is sensitive to quantitative information. In the OFFICE treatment, the value of X equaled the Total Number of Answers (*TNA*), either correct or incorrect, that the worker submitted. Thus, the signal distribution in the OFFICE treatment reflects both productive (qualitative, number of correct answers) and unproductive (quantitative, total number of answers) inputs. The signal in the HOME treatment, in contrast, reflects only productive effort. To achieve this, the value of x must be independent of the actual TNA. Arguably, by increasing the expected signal, a higher value of x undermines workers' incentives to invest effort to boost the signal. We, therefore, calibrated the value of X in the HOME treatment to reflect the expected TNA under the null hypothesis of no incentive distortion. Based on pilot sessions, we set $X = 7$.[7]

The treatments varied between the two blocks, with the order counterbalanced across matching groups. Participants read the relevant instructions on-screen at the beginning of each block and were required to correctly answer control questions to ensure understanding of the experimental variation. The end-of-round feedback included the signal that the manager observed and the bonus paid to the worker. Neither the worker nor the manager received accurate feedback regarding actual productivity. At the end of the experiment, participants learned their total profits in each round.

The experimental instructions appeared on-screen and were read aloud by the experimenters (see the appendix for a translation). Participants could then ask questions privately. The experiment started after all participants confirmed that they had read and understood the instructions and answered the control questions correctly. The experiment was conducted at the Experimental Economics Laboratory at the Department of Economics of the Ben-Gurion University of the Negev. One hundred and sixty-six students from across the university were recruited from two subject pools in the Economics and Management departments using ORSEE (Greiner, 2015) and by email. No participant participated twice in the experiment.[8] The experiment was programmed using z-Tree (Fischbacher, 2007). Six rounds, three from each block, were randomly chosen for payment. Experimental earnings were converted to Israeli New Shekels (ILS) at a conversion rate of 10 ECU = 1 ILS and added to a show-up fee of 20 ILS (15 ILS in two of the sessions). Final payoffs ranged from 38 ILS to 105 ILS, with an average of 61.4 ILS (approximately 19 USD) per participant.

4. Theoretical analysis

In this section, we aim to illustrate how incorporating quantitative information into the signal—as implemented in the experiment—may lead to lower productivity. As our focus in this paper is on the worker side, we make the simplifying assumption that the manager responds to the information available to her in a monotonic and linear way and solve the resulting optimiza-

[7]The mean TNA in the experiment reported in Section 5 did not differ significantly from 7.

[8]A technical problem during one session corrupted the data of two rounds, which we consequently excluded from the analysis. Excluding the complete sessions does not affect the results qualitatively.

tion problem of the worker. We assume that the worker cannot improve his performance by increasing effort but has control over which problems to solve and which to skip.[9] As in the experiment, the worker encounters problems sequentially. The worker can evaluate the problem's difficulty, which we equate with the time required to solve it. Having assessed the problem, the worker chooses whether to solve it or skip it by submitting an arbitrary solution. The worker's optimization problem is thus reduced to choosing the highest difficulty level above which he skips a problem that maximizes the expected observed signal.

Formally, each problem k has a solving time t_k, which the worker can identify after an initial evaluation time $t_e > 0$. To simplify the analysis, we assume that the worker does not know when the manager is observing, and therefore solves the infinite-horizon problem. The worker's strategy is a function $\sigma(t_k) : [0,1] \rightarrow \{0,1\}$ indicating whether the worker solves a problem or skips it given its difficulty.[10] The strategy σ induces an expected time per solved problem $t_s(\sigma)$ and an expected response time per problem (solved or not) $t_r(\sigma)$. The manager observes performance in such time window of length T. Specifically, the manager receives a signal that is weakly increasing in the number of problems and strictly increasing in the number of solved problems. The expectation of the former is given by $q_r(\sigma) = \frac{T}{t_r(\sigma)}$ and of the latter is given by $q_s(\sigma) = \frac{T}{t_s(\sigma)}$. For simplicity, we assume that t_k is a uniform random variable distributed $t_k \sim U(0,1)$.

Lemma. *The optimal strategy $\sigma^*(t_k)$ is monotonic; $\sigma(t_k) \geq \sigma(t_l)$ iff $t_k \leq t_l$.*

Proof. See Appendix B. □

The lemma implies that, with some abuse of notation, we can replace σ^* with a cutoff value $\bar{t}$, below which the worker solves the problem, and above which the worker skips. Because of our simplifying assumptions on the distribution of t_k, the share of problems solved is $\bar{t}$, and the expected time spent solving a problem is $\frac{\bar{t}}{2}$. The worker encounters and identifies, on average, $\frac{1}{\bar{t}}$ problems for each problem solved, hence the average total time spent per solved problem is $t_s(\bar{t}) = \frac{t_e}{\bar{t}} + \frac{\bar{t}}{2}$. The average time spent on any problem is $t_r(\bar{t}) = t_e + \bar{t} \cdot \frac{\bar{t}}{2}$. The number of solved problems in T is thus $q_s = \frac{T}{\frac{t_e}{\bar{t}} + \frac{\bar{t}}{2}}$ and the total number of problems (solved or skipped) in T is $q_r = \frac{T}{t_e + \frac{\bar{t}^2}{2}}$, corresponding to *NCA* and *TNA* in the experiment, respectively.

We assume that the worker aims to maximize the expected signal observed by the manager. In the experiment, the manager observes the NCA with probability 0.5 and a random number drawn from a uniform distribution on $[1, \bar{s}]$ otherwise. In the HOME treatment, the upper bound is $\bar{s}_H = 7$. In the OFFICE treatment, the upper bound is $\bar{s}_O = q_r(\bar{t})$.

Proposition 1. *Moving from the OFFICE environment to the HOME environment reduces skipping and increases productivity.*

[9] Performance in simple tasks carried out in limited time in the laboratory does not increase substantially with higher incentives, indicating low sensitivity to increased effort (see, e.g., Araujo et al., 2016).

[10] More generally, the strategy can be probabilistic and depend on the history. The optimal strategy is, however, necessarily stationary and deterministic (Prieto-Rumeau, 2006; Prieto-Rumeau and Hernández-Lerma, 2005; Puterman, 1974).

Proof. In the HOME environment, the worker's problem is reduced to finding the optimal cutoff that maximizes q_s:

$$\bar{t}_H^* = \underset{\bar{t}}{\operatorname{argmax}} \left[\frac{T}{\frac{t_e}{\bar{t}} + \frac{\bar{t}}{2}} \right] = \sqrt{2t_e}. \tag{1}$$

In the OFFICE treatment, the optimal cutoff that maximizes the expected signal is given by

$$\bar{t}_O^* = \underset{\bar{t}}{\operatorname{argmax}} \left[0.5 \frac{T}{\frac{t_e}{\bar{t}} + \frac{\bar{t}}{2}} + 0.5 \cdot 0.5 \left(\frac{T}{t_e + \frac{\bar{t}^2}{2}} + 1 \right) \right] = \frac{\sqrt{8t_e + 1} - 1}{2}. \tag{2}$$

It follows that $\bar{t}_H^* > \bar{t}_O^*$. That is, there is a range of problems that the worker solves in the HOME environment but skips in the OFFICE environment. Productivity is the expected number of solved problems. Substituting (1) for $\bar{t}$ in the expression for q_s yields a productivity at HOME of $\frac{T}{\sqrt{2t_e}}$. Substituting (2) for $\bar{t}$ in q_s yields a lower productivity at the OFFICE of $\frac{2T}{\sqrt{8t_e+1}}$.[11] □

4.1. Hypotheses

In line with Proposition 1, workers in the OFFICE treatment increase the expected signal by solving more problems at the cost of reduced productivity. We, therefore, predict to find more answers but fewer correct answers in OFFICE than in HOME. The first hypotheses state these predictions.

Hypothesis 1. *The TNA is higher in OFFICE than in HOME.*

Hypothesis 2. *The NCA is lower in OFFICE than in HOME.*

The analysis assumes that skipping decisions drive treatment effects in performance. To test the suggested mechanism, we aim to identify skipping decisions and explore the relation between such decisions and problem difficulty. We describe this analysis in detail in Section 5.2. Based on this direct analysis of strategies, we state the following hypotheses:

Hypothesis 3. *Workers are more likely to skip a problem as the problem difficulty increases.*

Hypothesis 4. *Workers skip more in the OFFICE treatment.*

Furthermore, we conjecture that the misaligned incentives in the OFFICE treatment may lead workers to skip regardless of problem difficulty.

Hypothesis 5. *The decision to skip a problem is more sensitive to the problem difficulty in the HOME treatment.*

5. Results

We start by analyzing the effect of the signal structure on the outcomes, as stated in Hypotheses 1 and 2. We proceed to analyze worker strategies in the two treatments following Hypotheses 3–5.

[11] This is immediately apparent from the fact that $\sqrt{8x+1} > \sqrt{8x} = 2\sqrt{2x}$.

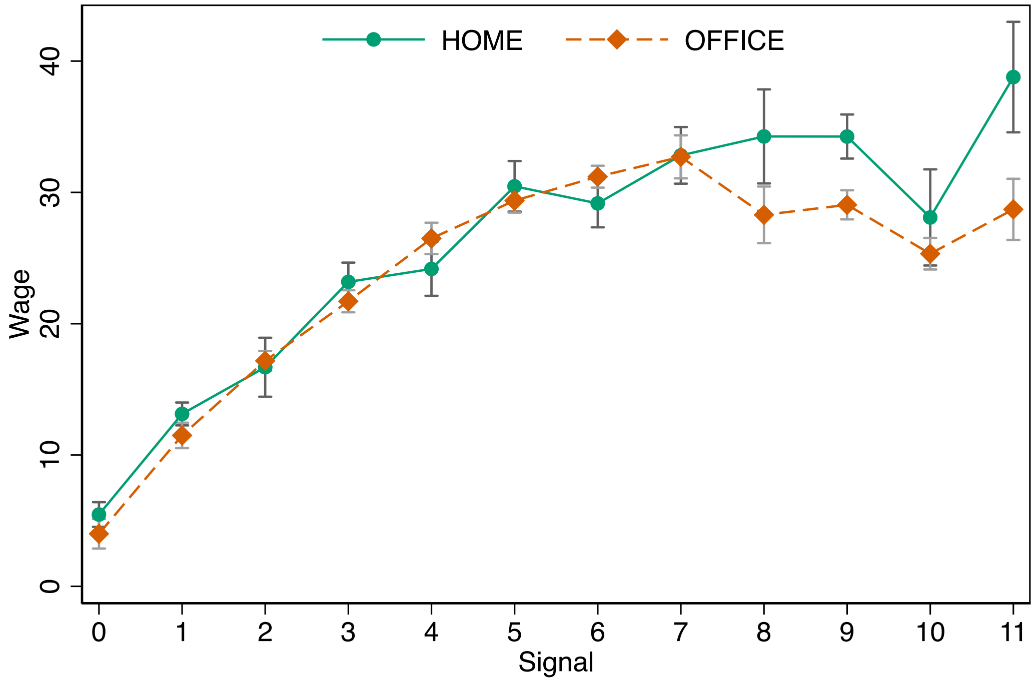

Figure 1: Wages and signals

5.1. Output and productivity

Our aim is to test the effect of the signal structure on the worker side, given that a higher signal results in higher wages. As background for the tests for our hypotheses relating to worker performance, Figure 1 depicts the mean wages as a function of the signals in the range $[0, 11]$.[12] We see that managers respond positively to the signal. In the OFFICE treatment, there is some ambiguity regarding high signals, which may come from exceptional productivity or from the noise, leading to some reduction of wages for signals above 7. Overall, managerial behavior support our assumption that workers benefit from a higher signal in both treatments, at least up to a signal of 7.

Turning to worker performance, Figure 2 shows the average total number of answers (TNA) and the number of correct answers (NCA) by treatment and within-block period. Workers increase their quantity in the OFFICE treatment, where the signal is sensitive to the TNA, in line with Hypothesis 1. In line with Hypothesis 2, this increase is accompanied by a decrease in productivity.

The regressions presented in Tables 1 and 2, controlling for the last received wage and the period and clustering standard errors on matching groups, confirm these observations. The results are robust to including fixed and random individual effects. On average, employees submitted around three answers more in the OFFICE treatment than in the HOME treatment

[12]Signals above 11 appeared only in the OFFICE treatment.

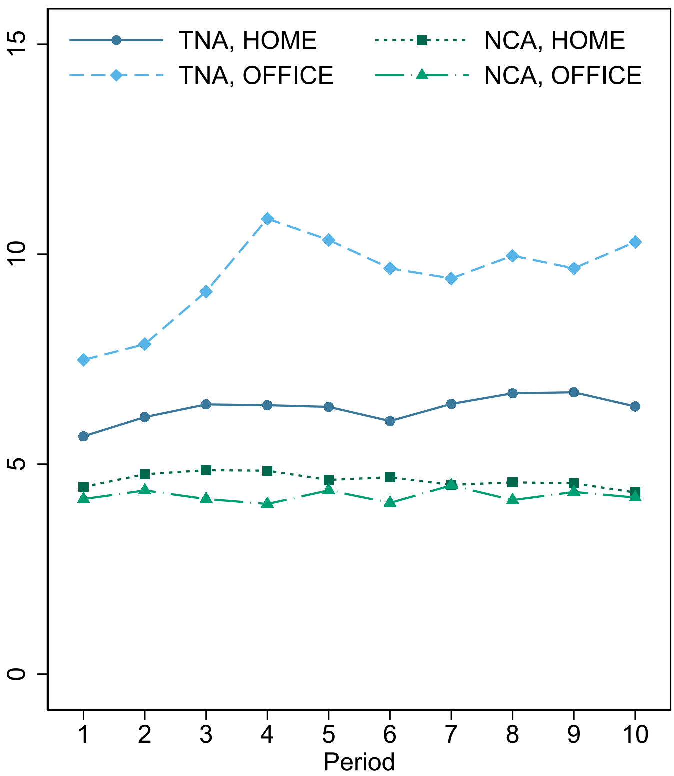

Figure 2: Total and correct answers.

Table 1: Treatment effects on TNA.

	(1)	(2)	(3)	(4)	(5)	(6)
	TNA	TNA	TNA	TNA	TNA	TNA
OFFICE	3.123***	3.022***	3.032***	3.276***	3.144***	3.156***
	(3.929)	(4.015)	(4.008)	(3.790)	(3.959)	(3.946)
Lagged wage				−0.008	0.017	0.014
				(−0.288)	(1.237)	(1.016)
Period				0.096	0.098	0.098
				(1.463)	(1.494)	(1.489)
Constant	6.321***	6.372***	6.394***	6.018***	5.466***	5.583***
	(18.952)	(16.934)	(18.478)	(8.343)	(7.238)	(12.263)
Individual effects	No	FE	RE	No	FE	RE
N	1624	1624	1624	1446	1446	1446

Notes: mixed- and fixed-effects regressions with robust standard errors clustered on matching groups. t-statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

($p < 0.01$). When the employer observes quantitative information, workers increase their non-

Table 2: Treatment effects on NCA.

	(1) NCA	(2) NCA	(3) NCA	(4) NCA	(5) NCA	(6) NCA
OFFICE	-0.373^{**}	-0.353^{**}	-0.354^{**}	-0.347^{**}	-0.325^{**}	-0.326^{**}
	(−2.625)	(−2.539)	(−2.545)	(−2.407)	(−2.322)	(−2.331)
Lagged wage				0.033^{***}	0.029^{***}	0.030^{***}
				(3.295)	(5.223)	(5.332)
Period				−0.019	−0.018	−0.018
				(−1.033)	(−0.989)	(−0.991)
Constant	4.615^{***}	4.605^{***}	4.590^{***}	3.929^{***}	3.989^{***}	3.960^{***}
	(18.982)	(66.183)	(18.831)	(10.150)	(24.649)	(13.636)
Individual effects	No	FE	RE	No	FE	RE
N	1624	1624	1624	1446	1446	1446

Notes: mixed- and fixed-effects regressions with robust standard errors clustered on matching groups. t-statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

productive effort to influence the signal in their favor. Thus, the results support Hypothesis 1.

Result 1. *The TNA in the OFFICE treatment is higher than in the HOME treatment.*

Furthermore, we observe a significant difference in productivity between the OFFICE and HOME treatments. On average, workers solved correctly 0.3 more problems in the HOME treatment compared to the OFFICE treatment ($p < 0.05$). These findings support our hypothesis, suggesting that the increase in task quantity comes at the cost of task quality.

To validate our proposed mechanism, we must address potential alternative explanations. Specifically, we investigate the unique signaling dynamics within the HOME treatment. Unlike the OFFICE treatment, any signal higher than 7 necessarily corresponds to the true number of correct answers. As a result, workers in the HOME treatment are aware that the manager will recognize signals of 8 or higher as true representations of their performance. This heightened signaling value for more than seven correct answers may incentivize workers to aim for more than seven correct answers in the HOME treatment only.

This alternative explanation assumes that workers can increase their productivity in response to the implied incentives. Contrary to this assumption, prior research, such as Araujo et al., 2016, demonstrates that workers in simple laboratory tasks typically perform to the best of their abilities, and additional effort has a negligible effect on their performance. Thus, we attribute any observed treatment differences to potential skipping strategies rather than variations in exerted effort, effectively undermining this alternative explanation. Nevertheless, to ensure a comprehensive analysis, we carefully examine the data for any evidence that could support this alternative hypothesis.

To this end, we scrutinize the distribution of correct answers. If workers in the HOME treatment indeed respond to the signaling incentives by exceeding seven correct answers, we would expect to observe fewer instances where the number of correct answers (NCA) is 7 or slightly below in the HOME treatment, accompanied by more instances where the NCA is 8 or slightly above. Figure 3 shows the NCA distribution by treatments. Whereas more workers solve eight problems correctly in the HOME treatment, this is also true for six or seven problems. The most prominent effect apparent in the figure is that it is more common in the OFFICE treatment that a worker does not solve any problem, in line with our hypothesis that shifting focus to quantity comes at the expense of quality.

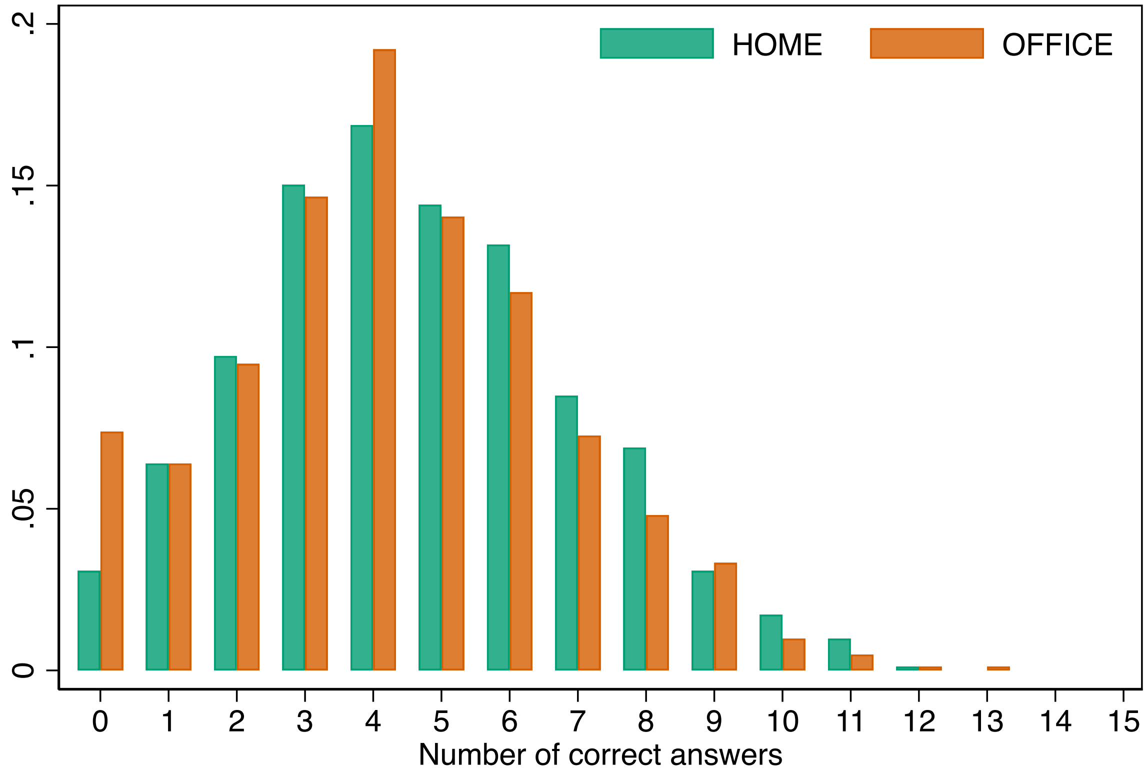

Figure 3: Number of correct answers by treatment.

To rigorously test the alternative explanation, we examine whether the treatment effect is driven by a shift in the distribution towards values above 7 in the HOME treatment. Table 3 reports the relevant regression results. Column (1) repeats the basic regression reported in Column (1) of Tables 1 and 2. Column (2) repeats the analysis dropping the observations in which the signal exceeds 7. Naturally, this adjustment reduces the effect size irrespective of the underlying explanation. To account for workers pushing the NCA beyond 7 without artificially suppressing the overall treatment difference, we report in Columns (3)–(4) tobit regressions censoring the dependent variable at 7, 6, and 5, respectively. The results remain consistent with the main specification reported in Column (1). Consequently, we can attribute the majority of the treatment effect to the role of quantitative information in the signal within the OFFICE treatment, given the reasonable assumption that any extra effort only impacts values close to the cutoff point of seven correct answers. The analysis of strategies to be presented in Section 5.2 below further supports our interpretation of the results. To sum, the experimental evidence establishes that the increase in quantity comes at the expense of quality, supporting Hypothesis 2.

Result 2. *The NCA in the OFFICE treatment is lower than in the HOME treatment.*

Table 3: Regressions on restricted NCA.

	(1) Unrestricted	(2) 7-capped	(3) 7-censored	(4) 6-censored	(5) 5-censored
OFFICE	-0.347^{**}	-0.231^{*}	-0.365^{**}	-0.372^{**}	-0.368^{**}
	(-2.407)	(-1.724)	(-2.331)	(-2.178)	(-2.082)
Lagged wage	0.033^{***}	0.030^{***}	0.035^{***}	0.037^{***}	0.041^{***}
	(3.295)	(4.256)	(3.434)	(3.682)	(4.121)
Period	-0.019	-0.031^{**}	-0.022	-0.033	-0.037^{*}
	(-1.033)	(-2.032)	(-1.101)	(-1.523)	(-1.714)
Constant	3.929^{***}	3.455^{***}	3.943^{***}	4.004^{***}	3.958^{***}
	(10.150)	(14.614)	(10.649)	(10.947)	(10.651)
N	1446	1280	1446	1446	1446

Notes: OLS and tobit regressions for number of correct answers with robust standard errors clustered on matching groups. t-statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

5.1.1. Order effects

Further investigation reveals an asymmetric effect with respect to the order of the treatments. Table 4 presents mixed-effects regressions with individual random effects and robust standard errors clustered on matching groups testing the order effect. Although the interaction of the treatment and the order on the NCA does not reach significance, the treatment effect appears only when switching from OFFICE to HOME. The effect in the opposite order is weaker and non-significant.

There is also a main effect for order, reflecting higher quantity and quality in the HOME treatment when following the OFFICE treatment. This difference is not likely to reflect a pure learning effect, as the effect of the period within blocks is negligible and negative. A possible explanation for this order effect is that when workers switch to HOME, they increase their productivity to compensate for the reduction in their influence on the signal in the OFFICE treatment. We expect that when workers transition from HOME to OFFICE, they increase quantity to boost the signal at the expense of productivity. However, although we see an

Table 4: Order effects.

	(1) TNA, All	(2) TNA, OFFICE first	(3) TNA, HOME first	(4) NCA, All	(5) NCA, OFFICE first	(6) NCA, HOME first
OFFICE	2.565***	2.582***	3.752***	−0.555***	−0.558**	−0.130
	(2.680)	(2.643)	(2.886)	(−2.594)	(−2.572)	(−0.767)
OFFICE × HOME first	1.137			0.439		
	(0.718)			(1.584)		
HOME first	−2.200***			−1.054**		
	(−3.200)			(−2.203)		
Lagged wage	0.016	0.003	0.030	0.031***	0.035***	0.027***
	(1.160)	(0.174)	(1.329)	(5.519)	(4.541)	(3.314)
Period	0.097	0.176	0.024	−0.018	−0.015	−0.021
	(1.483)	(1.470)	(0.417)	(−0.995)	(−0.688)	(−0.722)
Constant	6.665***	6.466***	4.510***	4.483***	4.377***	3.569***
	(12.180)	(13.497)	(5.405)	(11.576)	(11.933)	(8.217)
N	1446	696	750	1446	696	750

Notes: mixed-effects regressions with robust standard errors clustered on matching groups. t-statistics in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

increase in the signal when switching to OFFICE, the lack of decrease in productivity can be explained by the characteristic of the task. Increasing the signal (by submitting random answers) is relatively easy and might not dramatically affect the workers' productivity. We summarize in the following result.

Result 3. *Reducing the information incorporated in the signal (transition to "working from home") increases productivity. Enriching the signal (transition to "working in the office") does not decrease productivity significantly. Productivity "at home" is higher if following a transition from "in the office."*

5.2. Worker strategy

We now turn to analyze the workers' strategies and how these strategies differ across treatments. We begin with identifying the decision to skip a problem, followed by analyzing problem difficulty. Section 5.2.3 combines the two variables to estimate strategies.

5.2.1. Estimating the probability of skipping

We assume that the time spent on any given problem results from a two-step process. First, the worker quickly estimates the problem difficulty and decides whether to skip it. If the worker decides to skip, the solving time follows a lognormal distribution with a low mean and is independent of the problem difficulty. If the worker attempts to solve the problem,

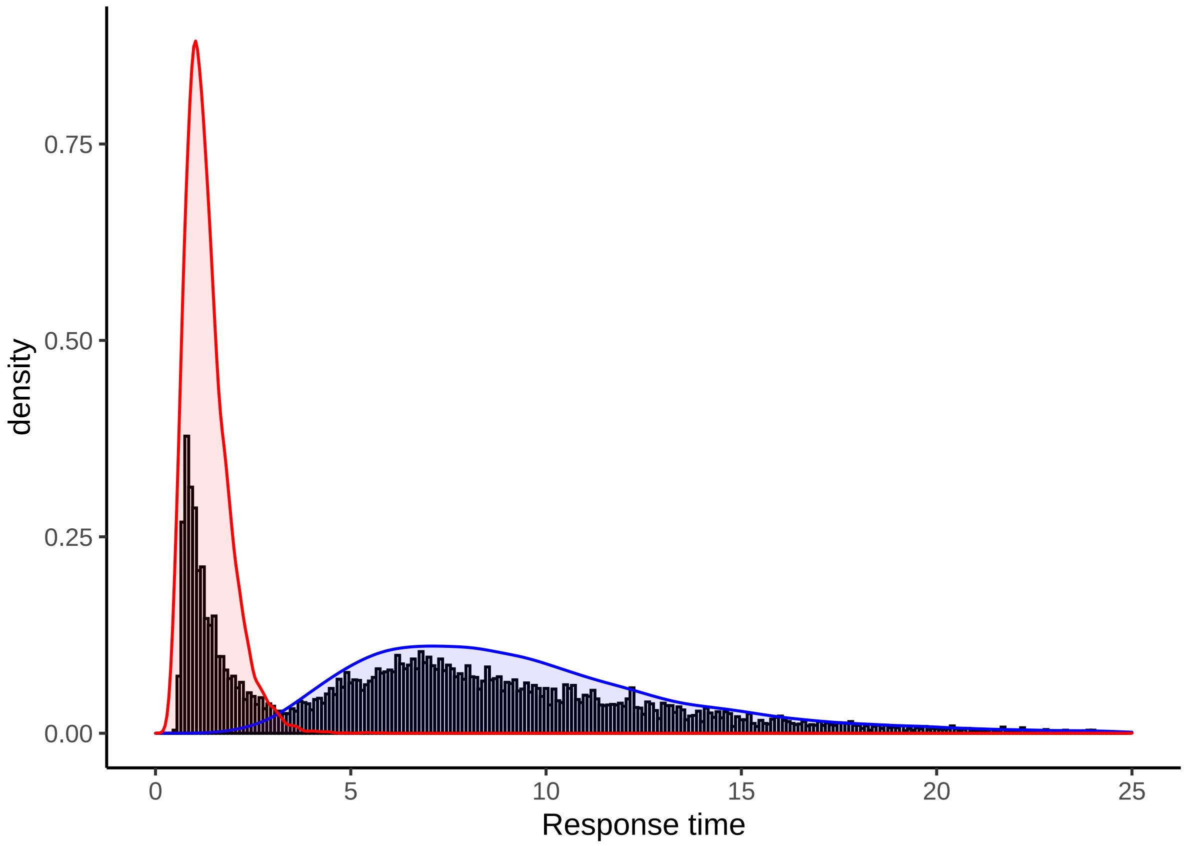

Figure 4: Empirical and modelled distributions of solving time.

the solving time follows a lognormal distribution with a higher mean, and depends on the problem difficulty.[13] The histogram in Figure 4 presents the distribution of solving times across all problems solved in the experiment. Consistent with our assumption, the distribution is bimodal, with a low mode at around two seconds and a high mode at about seven seconds.

We fitted a finite mixture model to estimate the mean and standard deviation of each of the two underlying lognormal distributions and the share of each distribution. The two distributions generated by the model are overlaid over the histogram in Figure 4. The model generates for each observation a posterior probability that the participant skipped the problem (i.e., that the solving time for the problem comes from the low distribution). We define a problem as skipped if and only if the posterior probability is higher than 50%,[14] and regardless of the correctness of the solution.[15] See C.1 for more details.

[13]The assumption that response time distributions are lognormal is a standard assumption in psychology and economics (Linden, 2006; Moffatt, 2005; Thissen, 1983)

[14]The probability of skipping is lower than 40% or higher than 60% for over 99% of the observations. The skipping probability is lower than 1% for over two-thirds of the observations, and higher than 99% for over 20%.

[15]Among 3,649 skipped problems, only 7 were solved correctly.

5.2.2. Estimating difficulty level

A productivity-maximizing worker should skip a problem if the time required to answer the problem is higher than the expected solving time of a new random problem.[16] Accordingly, we aim to create a measure of problem difficulty based on solving time, subject to individual fixed effects such as mathematical ability. Because the problems in the experiment were randomized, it is not feasible to estimate the solving time for each problem individually. We, therefore, used supervised machine learning to predict the solving time based on the problem characteristics in an independent data set.

To generate this independent data set, we recruited five participants who did not participate in the experiment.[17] The participants solved randomized problems for 30 minutes, earning 0.5 NIS for each correct answer. Participants had to submit correct answers before proceeding, and skipping was not possible. After submitting a correct answer, participants could rest as the clock paused while the computer screen presented the time elapsed and problems solved up to that point. Overall, the participants solved a total of 794 problems. Thirty-six problems with a solving time of more than 25 seconds (indicating loss of concentration) were removed, leaving 769 problems to comprise the training data set.

We trained our model to predict the standardized solving time based on fifteen problem characteristics using lasso regressions.[18] Applying the prediction model to the experimental problem set and standardizing the results generates a difficulty score for each problem our participants faced. The full details are in C.2.

5.2.3. Skipping and difficulty

The posited data-generation process described above predicts that the solving time depends on the problem difficulty only if the participant does not skip the problem. We use this prediction to validate our measures of skipping and problem difficulty. Regressions of solving time on the problem difficulty interacted with whether our skipping measure marks the problem as skipped, with robust standard errors clustered on participants, confirm the prediction. Solving time is significantly correlated with the problem difficulty for solved problems ($\beta = 0.51, t(82) = 3.09, p = .003$) but not for skipped problems ($\beta = -0.39, t(82) = -1.20, p = .235$).

Figure 5 shows the relation between difficulty and skipping by treatment. Panel A presents the share of skipped problems in the two treatments based on the problem difficulty. The most evident effect is that workers are more likely to skip a problem in OFFICE than in HOME.[19] Within treatments, more difficult problems are more likely to be skipped in the

[16] The solving time includes the skipping decision time and the answer time.

[17] To minimize the variance in mathematical ability, all five participants were Industrial Engineering and Management graduate students.

[18] Problem characteristics include, for example, “two of the three numbers sum to a round number” and “the sum is less than 100”.

[19] This difference is consistent throughout the 60-second duration. Additionally, in the OFFICE treatment we observe a steep increase in the propensity to skip in the last ten seconds that is almost completely absent in the HOME treatment.

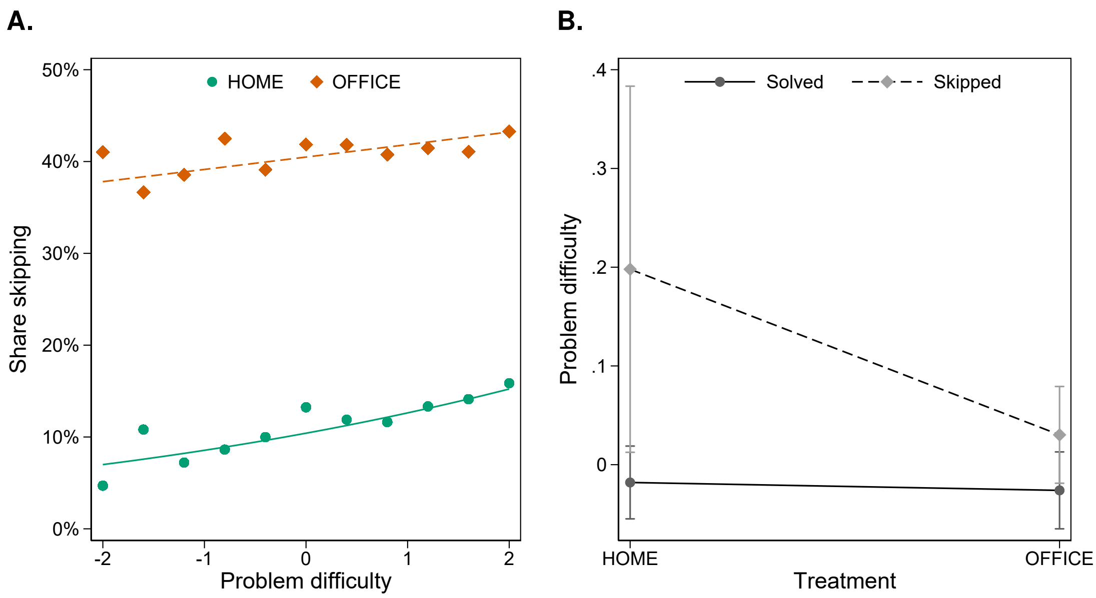

Figure 5: Problem difficulty and skipping.

HOME treatment. The relation between problem difficulty and skipping is weaker in OFFICE and appears to be flat for problems that are not easy (difficulty of one standard deviation below the mean or higher). The difference between the treatments is more apparent in Panel B, which flips the relation to show the mean difficulty of solved and skipped problems by treatment. The skipped problems are significantly more difficult than the solved problems in HOME ($t(82) = 2.14, p = .035$, based on an OLS regression of difficulty on treatment interacted with skipping and robust standard errors clustered on subjects). In contrast, the difference is much smaller and only weakly significant in OFFICE ($t(82) = 1.67, p = .099$). As we saw in Section 5.1, the effect of the signal structure is strongest when transiting from OFFICE to HOME. Indeed, adding the order and its interactions to the model reveals that the difference is only significant when HOME follows the OFFICE treatment ($t(82) = 2.20$, $p = .031$; $p > .110$ for the other three comparisons). The interaction of treatment and skipping is significant in the OFFICE-first order ($F(1, 82) = 4.87, p = .030$) but not in the HOME-first order ($F(1, 82) = 0.53, p = .469$).

Table 5 presents a series of logistic regressions supporting these conclusions. In the HOME treatment, the probability of skipping is significantly higher for more difficult problems. This relationship is substantially weaker in the OFFICE treatment. The interaction, however, does not reach significance. The interaction is stronger and significant when excluding the 10 percent easiest problems in Column (2) and when participants transition from the OFFICE to the HOME treatment in Column (3). Somewhat unexpectedly, problem difficulty does not significantly affect the skipping probability in the opposite order.

Thus, the analysis of worker strategy confirms Hypothesis 4. Support for Hypotheses 3 and 5 is strongest when workers transition from OFFICE to HOME. The following result summarizes the strategy analysis.

Table 5: Problem skipping.

	(1) All	(2) All	(3) OFFICE first	(4) HOME first
OFFICE × Difficulty	−0.161	−0.183*	−0.237**	0.0495
	(−1.63)	(−1.90)	(−2.16)	(0.75)
OFFICE	1.767***	1.782***	1.411***	2.510***
	(4.88)	(4.95)	(3.31)	(3.32)
Difficulty	0.218**	0.214**	0.265**	0.0375
	(2.12)	(2.12)	(2.15)	(0.61)
Constant	−2.152***	−2.155***	−1.746***	−2.948***
	(−6.39)	(−6.38)	(−4.58)	(−4.21)
Exclude easy	NO	YES	NO	NO
N	12802	11521	6727	6075

Notes: t-statistics based on robust standard errors clustered on subjects in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Result 4. *Workers skip problems more in the OFFICE treatment. Problem difficulty level increases the probability of skipping the problem in the HOME treatment but less so in the OFFICE treatment.*

6. Conclusion

The literature on multitasking primarily analyses the optimal incentive scheme given the correspondence between the production technology and the signal structure (Baker, 1992, 2002; Gibbons, 2005; Holmstrom and Milgrom, 1991). This paper focuses on how workers respond to different signal structures and the implications for productivity. Workers necessarily have some autonomy in allocating effort between various tasks or between different aspects of a task. To make the best of a worker's limited time and effort, optimal effort allocation requires focusing on tasks that maximize the production rate. We refer to this aspect of the task as quality. When the observed signal is sensitive to quantitative aspects of the job, the worker is incentivized to reallocate effort to less efficient tasks. As a result, the quality of the output decreases. This insight is relevant for the design of work environments. Specifically, it suggests the counter-intuitive conclusion that reducing observability—as when switching from office work to working from home—may reduce the transparency of quantitative indicators, thereby improving incentives and increasing productivity.

A natural response to this argument is that managers can ignore irrelevant indicators. Much practical effort is indeed given to identifying the best procedures and measures to assess worker performance and productivity. However, this is not only often practically impossible, as various dimensions simultaneously affect the observed indicators—but is psychologically difficult.

Research in psychology shows that people are unable to ignore information, even if the information is irrelevant or "non-diagnostic" to the task at hand (Nisbett et al., 1981; Waller and Zimbelman, 2003; Zukier, 1982). This so-called *Dilution effect* extends to performance evaluations (Humphrey, 1997).

Our experimental results are in line with the reasoning presented above. When quantity is (noisily) observable, workers artificially inflate the number of tasks they work on; and become less efficient in their effort allocation between tasks. The experiment reveals an unexpected order effect. We find significant support for our hypotheses when workers transition from high observability ("office") to low observability ("home"). In contrast, the treatment effects diminish and disappear when workers start in the limited observability treatment. Productivity increases but does not decrease with a shift in the signal structure. A possible explanation is that workers only learn to estimate problem difficulty with experience. There are, accordingly, two conditions for the implementation of efficient strategies that consider problem difficulty. Workers must be both experienced and subject to undistorted incentives. As a result, we only observe a strong effect for problem difficulty on skipping decisions in the late HOME treatment. The effect is considerably weaker in the late OFFICE treatment and does not exist in the first block regardless of the treatment. Does such asymmetry exist in actual firms? The literature on working from home reviewed in Section 2 only looked at the transition from office to home work. We could not find any field study testing the opposite direction. Future work is required to understand the reasons for this asymmetric effect better and to what extent it generalizes to field conditions.

These findings have important implications for performance evaluations. Even in an environment of incomplete contracts, workers allocate attention and effort to tasks that maximize the observable measures, which may harm efficiency. Workers who understand that working long hours improves their employer's (explicit or implicit) evaluations will prioritize their tasks to extend their work hours while reducing their total productivity. This aspect of transparency may provide part of the explanation for recent research pointing at the benefits of working from home (Angelici and Profeta, 2020; Bloom et al., 2015; Choudhury et al., 2021).

Finally, this study focuses on the effect of the signal structure on worker behavior. Several questions regarding the employer side remain open. Do employers understand the potential downside of more transparency? Can employers ignore non-diagnostic information? Future research is needed to provide a more complete understanding of this phenomenon.

Disclaimer

During the preparation of this work the authors used Claude 2 in order to make slight grammatical and style changes. After using this tool/service, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

7. References

Angelici, Marta and Paola Profeta (2020). Smart-working: work flexibility without constraints. *Management Science*. Forthcoming.

Araujo, Felipe A, Erin Carbone, Lynn Conell-Price, Marli W Dunietz, Ania Jaroszewicz, Rachel Landsman, Diego Lamé, Lise Vesterlund, Stephanie W Wang, and Alistair J Wilson (2016). The slider task: an example of restricted inference on incentive effects. *Journal of the Economic Science Association* 2(1), pp. 1–12.

Baker, George (1992). Incentive contracts and performance measurement. *Journal of Political Economy* 100 (3), pp. 598–614.

— (2002). Distortion and risk in optimal incentive contracts. *Journal of human resources*, pp. 728–751.

Barrero, Jose Maria, Nicholas Bloom, and Steven J. Davis (2021). Why working from home will stick. *NBER working paper* 28731.

Bick, Alexander, Adam Blandin, and Karel Mertens (2021). Work from home before and after the covid-19 outbreak. *American Economic Journal: Macroeconomics*. Forthcoming.

Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying (2015). Does working from home work? evidence from a chinese experiment. *The Quarterly Journal of Economics* 130(1), pp. 165–218.

Brynjolfsson, Erik, John J Horton, Adam Ozimek, Daniel Rock, Garima Sharma, and Hong-Yi TuYe (2020). Covid-19 and remote work: an early look at us data. *NBER working paper* 27344.

Carr, Caleb T., Robert D. Hall, Adam J. Mason, and Eric J. Varney (2017). Cueing employability in the gig economy: effects of task-relevant and task-irrelevant information on fiverr. *Management Communication Quarterly* 31(3), pp. 409–428.

Choudhury, Prithwiraj, Cirrus Foroughi, and Barbara Larson (2021). Work-from-anywhere: the productivity effects of geographic flexibility. *Strategic Management Journal* 42(4), pp. 655–683.

Corgnet, Brice, Roberto Hernán-González, and Eric Schniter (2015). Why real leisure really matters: incentive effects on real effort in the laboratory. *Experimental Economics* 18, pp. 284–301.

Dalal, Dev K., Levi Sassaman, and Xiaoyuan Susan Zhu (2020). The impact of nondiagnostic information on selection decision making: a cautionary note and mitigation strategies. *Personnel Assessment and Decisions* 6(2), pp. 54–64.

DeFilippis, Evan, Stephen Michael Impink, Madison Singell, Jeffrey T. Polzer, and Raffaella Sadun (2022). The impact of covid-19 on digital communication patterns. *Nature humanities and social sciences communications* 9(180).

Dickinson, David and Marie-Claire Villeval (2008). Does monitoring decrease work effort?: the complementarity between agency and crowding-out theories. *Games and Economic behavior* 63(1), pp. 56–76.

Fehr, Ernst and Klaus M Schmidt (2004). Fairness and incentives in a multi-task principal–agent model. *The Scandinavian Journal of Economics* 106(3), pp. 453–474.

Fischbacher, Urs (2007). Z-tree: zurich toolbox for ready–made economic experiments. *Experimental Economics* 10(2), pp. 171–178.

Gartner (2020). Gartner HR survey reveals 41% of employees likely to work remotely at least some of the time post coronavirus pandemic. URL: https://www.gartner.com/en/newsroom/press-releases/2020-04-14-gartner-hr-survey-reveals-41--of-employees-likely-to- (visited on 04/14/2020).

Gerstenberg, Tobias, David A. Lagnado, and Ro'i Zultan (2023). Making a positive difference: criticality in groups. *Cognition* 238, 105499.

Gibbons, Robert (2005). Incentives between firms (and within). *Management Science* 51 (1), pp. 2–17.

Gibbs, Michael, Friederike Mengel, and Christoph Siemroth (2023). Work from home & productivity: evidence from personnel & analytics data on it professionals. *Journal of Political Economy: Microeconomics* 1(1), pp. 7–41.

Greiner, Ben (2015). Subject pool recruitment procedures: organizing experiments with orsee. *Journal of the Economic Science Association* 1(1), pp. 114–125.

Holmstrom, Bengt and Paul Milgrom (1991). Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & organization* 7, pp. 24–52.

Hong, Fuhai, Tanjim Hossain, John A List, and Migiwa Tanaka (2018). Testing the theory of multitasking: evidence from a natural field experiment in chinese factories. *International Economic Review* 59(2), pp. 511–536.

Humphrey, Ronald H. (1997). Job characteristics, prototypes, and the information dilution effect. *The Journal of psychology* 131(2), pp. 211–223.

Kalischko, Thomas and René Riedl (2021). Electronic performance monitoring in the digital workplace: conceptualization, review of effects and moderators, and future research opportunities. *Frontiers in psychology*, p. 1457.

Linden, Wim J. van der (2006). A lognormal model for response times on test items. *Journal of Educational and Behavioral Statistics* 31(2), pp. 181–204.

Manthei, Kathrin and Dirk Sliwka (2019). Multitasking and subjective performance evaluations: theory and evidence from a field experiment in a bank. *Management Science* 65(12), pp. 5861–5883.

Moffatt, Peter G. (2005). Stochastic choice and the allocation of cognitive effort. *Experimental Economics* 8(4), pp. 369–388.

Moore, Don A., Samuel A. Swift, Zachariah S. Sharek, and Francesca Gino (2010). Correspondence bias in performance evaluation: why grade inflation works. *Personality and Social Psychology Bulletin* 36(6), pp. 843–852.

Niederle, Muriel and Lise Vesterlund (2007). Do women shy away from competition? do men compete too much? *The Quarterly Journal of Economics* 122(3), pp. 1067–1101.

Nisbett, Richard E., Henry Zukier, and Ronald E. Lemley (1981). The dilution effect: nondiagnostic information weakens the implications of diagnostic information. *Cognitive psychology* 13(2), pp. 248–277.

Oosterbeek, Hessel, Randolph Sloof, and Joep Sonnemans (2011). Rent-seeking versus productive activities in a multi-task experiment. *European Economic Review* 55(5), pp. 630–643.

Prieto-Rumeau, Tomás (2006). Blackwell optimality in the class of markov policies for continuous-time controlled markov chains. *Acta Applicandae Mathematica* 92(1), pp. 77–96.

Prieto-Rumeau, Tomás and Onésimo Hernández-Lerma (2005). The laurent series, sensitive discount and blackwell optimality for continuous-time controlled markov chains. *Mathematical Methods of Operations Research* 61(1), pp. 123–145.

Puterman, Martin L (1974). Sensitive discount optimality in controlled one-dimensional diffusions. *The Annals of Probability*, pp. 408–419.

Rubin, Jared, Anya Samek, and Roman M Sheremeta (2018). Loss aversion and the quantity–quality tradeoff. *Experimental Economics* 21, pp. 292–315.

Thissen, David (1983). Timed testing: an approach using item response theory. In: *New horizons in testing: latent trait test theory and computerized adaptive testing*. Ed. by David J. Weiss. New York: Academic Press, pp. 179–203.

Trivedi, Sandeep and Nikhil Patel (2021). Virtual employee monitoring: a review on tools, opportunities, challenges, and decision factors. *Empirical Quests for Management Essences* 1(1), pp. 86–99.

Al-Ubaydli, Omar, Steffen Andersen, Uri Gneezy, and John A List (2015). Carrots that look like sticks: toward an understanding of multitasking incentive schemes. *Southern Economic Journal* 81(3), pp. 538–561.

Waller, William S. and Mark F. Zimbelman (2003). A cognitive footprint in archival data: generalizing the dilution effect from laboratory to field settings. *Organizational Behavior and Human Decision Processes* 91(2), pp. 254–268.

Zukier, Henry (1982). The dilution effect: the role of the correlation and the dispersion of predictor variables in the use of nondiagnostic information. *Journal of Personality and Social Psychology* 43(6), pp. 1163–1174.

Zultan, Ro'i, Tobias Gerstenberg, and David A. Lagnado (2012). Finding fault: causality and counterfactuals in group attributions. *Cognition* 125, pp. 429–440.

Appendix A. Experimental instructions

Thank you for coming to participate in the experiment.

Please read the instructions carefully.

The experiment includes two parts, each consisting of 10 rounds. At the end of the experiment, the computer will randomly draw three rounds from each part, and the payment you will receive will be the total profits you earned in these six rounds.

The participants will be randomly assigned into roles of workers and employers. These roles will be fixed during the whole experiment. At the beginning of the first part of the experiment, each worker will be randomly paired with one employer, so each worker will work for one employer, and each employer will have only one worker. These pairs will be fixed during the first part of the experiment. At the beginning of the second part of the experiment, each worker will be repaired with a new and different employer than the one he was paired with for the first part. These pairs will be fixed during the second part of the experiment.

Instructions for the round

The procedure of round

During each round, the worker will work for the employer by solving simple addition problems. The worker will have 60 seconds each round to solve the problems. Each correct solution will earn the employer 10 points. After submitting a solution, the worker will be presented with a new problem, and will not be able to return to the previous problems. The worker will not know if the solution he submitted is correct or wrong. There is no limit on the number of problems during a round.

During this time the employer will be able to make extra profits by clicking on a blue ball that will appear on random locations on the screen, using the left click of the mouse. Each click will earn him 0.1 points.

At the end of the 60 seconds, the employer will receive information regarding the worker's performance and will then choose the wage he wants to pay to the worker for his work in the round. The wage can be any round number between 0 and 60 points.

The information the employer will observe will differ between the parts as follows:

- The computer will flip a virtual coin. In the case of ”heads,” the computer will present the employer the real number of problems the worker solved correctly.
- In the case of ”Tails,” the computer will replace the real number with a random number from a given range. When the employer sees the number, he will not know whether this is the real number of correct solutions or a random number. Information regarding the range will be given to you at the beginning of each part.

After the employer decides the wage he wants to pay to his worker for his work in the round, both the employer and worker will observe the following information:

- The information the employer observed before choosing the wage.
- The wage the employer chose to pay.

Then a new round will start. Please notice: the employer's profit from each round will be visible to neither of the participants until the end of the experiment.

The profits from a round

The workers' profit from each round will be the wage paid by the employer plus 30 points.

The employers' profit from each round will be 60 points, plus the number of clicks on the blue ball multiplied by 0.1, plus 10 points for each problem the worker solved correctly, minus the wage he chose to pay to the worker.

Practice

Before the start of the experiment, you will go throw a short training, during which you will solve simple addition problems similar to these in the experiment for 1 minute. The purpose of this part is to let you get familiar with the task and will not affect your payment.

End of the experiment

After the end of the experiment, you will be asked to fill out a short survey. This survey, as any other decision you make during the experiment, is anonymous. Please wait in your seats until we call you to receive your payment.

We will now read the instruction out loud. If you have any questions afterward please raise your hand and the experimenter will come to answer you privately.

A.1. On-screen instructions

The following instructions were given on-screen at the beginning of each block of the experiment.

In the OFFICE treatment:

During this part, half of the time, the number observed by the employer will be randomly drawn, and its value will be between 1 and the total number of answers submitted by the worker.

In the Home treatment:

During this part, half of the time, the number observed by the employer will be randomly drawn, and its value will be between 1 and 7.

Appendix B. Proof of the lemma

Lemma. *The optimal strategy* $\sigma^*(t_k)$ *is monotonic;* $\sigma(t_k) \geq \sigma(t_l)$ *iff* $t_k \leq t_l$.

Proof. The worker maximizes the signal observed by the manager. Recall that the signal is increasing in $q_s = \frac{T}{t_s(\sigma)}$ and weakly increasing in $q_r = \frac{T}{t_r(\sigma)}$. Because $\frac{\partial q_s}{\partial t_s}, \frac{\partial q_r}{\partial t_r} < 0$, we can represent the signal as

$$S(\sigma) = f(t_s, t_r),$$

such that $\frac{\partial f}{\partial t_s} < 0$ and $\frac{\partial f}{\partial t_r} \leq 0$. Let $I(\sigma) \subseteq [0,1]$ be the region for which the worker solves the problem, formally $t \in I(\sigma)$ iff $\sigma(t) = 1$. Denote the length and mean of $I(\sigma)$ by $|I(\sigma)|$ and $\overline{I(\sigma)}$, respectively. The expected answer time is thus

$$t_r(\sigma) = t_e + |I(\sigma)| \cdot \overline{I(\sigma)}, \tag{B1}$$

and the expected solving time t_s is

$$t_s = \frac{t_e}{|I(\sigma)|} + \overline{I(\sigma)}. \tag{B2}$$

For any strategy σ, consider the monotonic strategy σ' given by

$$\sigma'(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq |I(\sigma)|, \\ 0 & \text{if } |I(\sigma)| < t \leq 1. \end{cases} \tag{B3}$$

From the definition of σ', $t_r(\sigma') \leq t_r(\sigma)$ and $t_s(\sigma') \leq t_s(\sigma)$. Consequently, $S(\sigma') \geq S(\sigma)$, hence the optimal strategy is monotonic. □

Appendix C. Additional analyses

C.1. Skipping analysis

We fitted a finite mixture model to estimate the probability of skipping based on the solving time. The observation-level log-likelihood function is given by

$$LL_i = \ln\left[\pi\phi\left(\frac{\ln(rt_i) - \mu_1}{\sigma_1}\right) + (1-\pi)\phi\left(\frac{\ln(rt_i) - \mu_2}{\sigma_2}\right)\right],$$

where rt_i is the solving time in problem i, π is the share of skipped problems, and μ_1 and σ_1 (μ_2 and σ_2) are the mean and standard deviation of the distribution of log solving time of the skipped (solved) problems. Figure C1 shows the kernel density estimate of the log solving time. We set the initial values for the estimation by splitting the observations at the minimum density marked by the vertical line in the figure, and calculating the mean and standard deviation for the observations above and below the splitting point. The initial value for π is set as the ratio between the number of observations below the splitting point and the total number of observations.

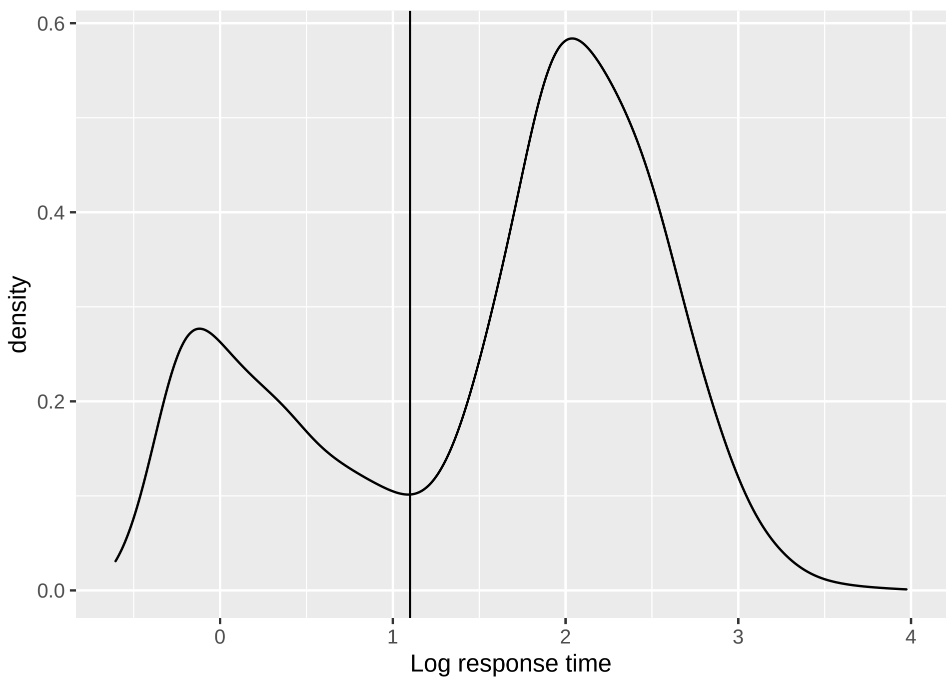

Figure C1

Next we apply Bayes' Rule to determine for each observation the posterior probability of being in the "skipped" distribution. The histogram of the resulting posteriors depicted in Figure C2 reveals that the posteriors are very informative. We therefore define a problem as skipped if the posterior is above 50%.

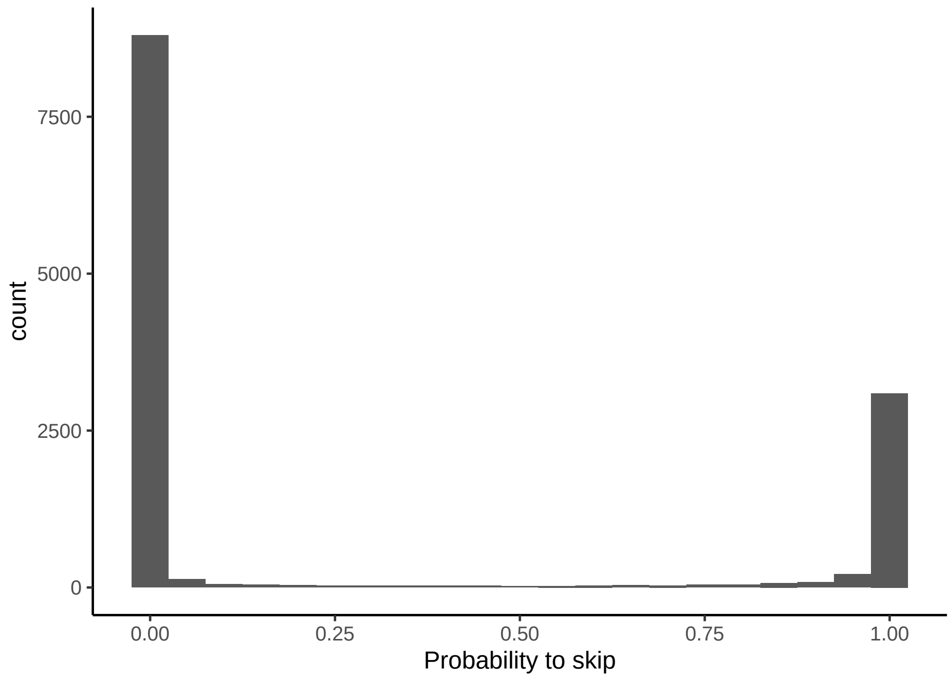
7500
5000
2500
0
count
0.00
0.25
0.50
0.75
1.00
Probability to skip

Figure C2

C.2. Problem difficulty analysis

We used supervised machine learning to estimate the difficulty level of each problem in the experiment. We constructed a model predicting the difficulty level of any given addition problem based on an independent training data set of random addition problems and solving times.

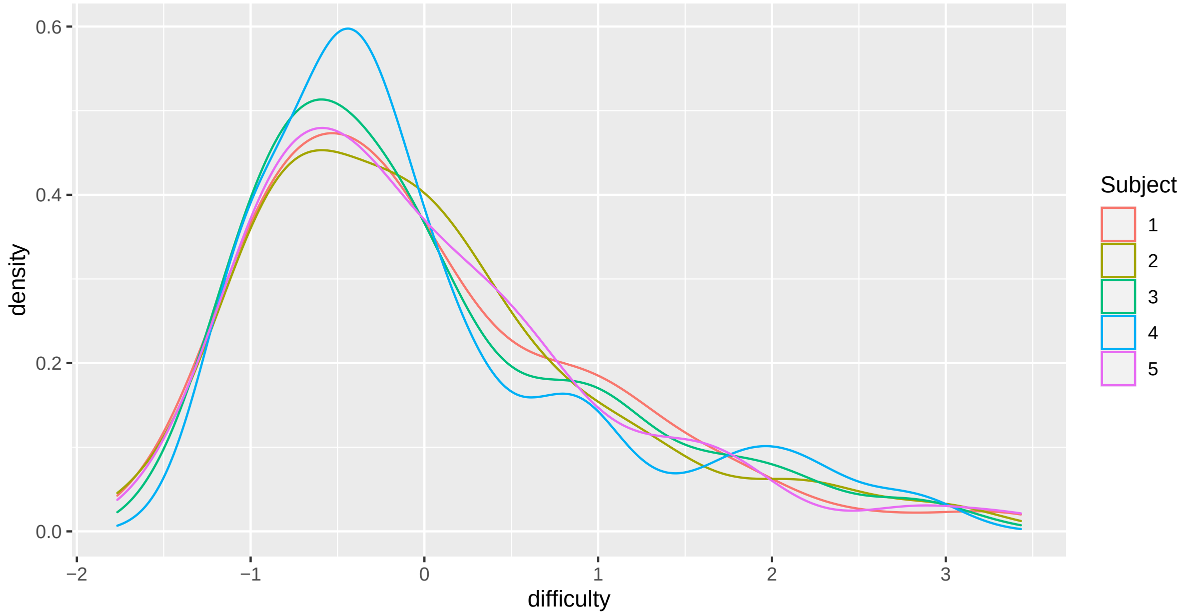

Figure C3: Distribution of problem difficulty in the training set.

To generate the independent data set, we recruited five participants who did not participate in the experiment. The participants solved randomized problems for 30 minutes, earning 0.5 NIS for each correct answer. Participants had to submit correct answers before proceeding and skipping was not possible. After submitting a correct answer, participants could rest as the clock paused while the computer screen presented the time elapsed and problems solved up to that point. Overall, the participants solved a total of 794 problems. Thirty-six problems with a solving time of more than 25 seconds (indicating loss of concentration) were removed, leaving 769 problems to comprise the training set. To wash out individual differences, the criterion used to train the prediction model was the problem's solving time, standardized separately within individuals. The resulting distribution appears in Figure C3.

As predictors, we used indicators for the features listed in Table C1. We used lasso regressions and V-fold CV, which conducts an automatic search for the optimal level of regularization.We also used the 'Lambda.min' feature which automatically chooses the lambda that brings the MSE to minimum. We consequently trained our model using only part of the independent data in order to validate the model. The procedure was as follows. We randomly split the data into a Training set, consisting of 20% of all observations and a Test set, consisting of the remaining 80% of observations. We then used this Training set for within-sample prediction, resulting in an MSE of 0.63. Next, we used the same methodology, this time training on the full independent Training data set and predicting the difficulty level in the experimental Test set.

Table C1: Predictors.

At least one unit digit is 0.
At least two unit digits are 0.
All three unit digits are 0.
At least two unit digits sum to 10.
All three unit digits sum to 10.
At least two ten digits sum to 10.
All three ten digits sum to 10.
The total sum is less than 100.
The total sum is between 100 and 200.
At least one unit digit is smaller than 7.
At least two unit digits are smaller than 7.
All three unit digits are smaller than 7.
At least one ten digit is smaller than 7.
At least two ten digits are smaller than 7.
All three ten digits are smaller than 7.