Goodhart's law in the labor market: signal structure, strategies, and productivity[*]

Eldar Dadon Ro'i Zultan[†]

April 18, 2022

Abstract

Managers observe worker production to inform compensation decisions. Intuitively, more observability allows managers to improve the compensation schemes to incentivize workers better. Nonetheless, incorporating quantitative production measures into the observed signal incentivizes workers to adapt their effort allocation to enhance the signal at the expense of efficiency. Thus, a richer signal can be associated with lower productivity. For example, working from home entails lower observability compared to office work. Consequently, office workers may prioritize time-consuming tasks to increase their observed work hours. We test this hypothesis in a controlled laboratory experiment. We find that removing quantitative information from the signal leads to more efficient strategies and higher productivity. Enriching the signal, however, does not have the opposite effect.

Keywords: incentives, multitasking, experiment JEL classification: C92, D86, J31

[*]We thank Or Avishay-Rizi, Fabio Galleoti, Tomer Ifergane, Marie Claire Villeval, Hadas Stiassnie, Adam Zylbersztejn, and seminar and conference 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 David Lagziel for his contribution to the model. We thank Yael Cohen, Tom Shefer and Reut Goldman, Rony Natan, Gali Goldberg, and Dor Ahady for assistance in conducting the experiments. 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

Designing optimal incentives for workers poses a significant challenge for managers. The problem is mitigated if an observable and contractible output is closely related to effort. While this may be feasible in some tasks and industries, such as simple production lines, in others, the manager's targets deviate from the observable worker behavior. Worker productivity may be difficult to measure directly or be realized only in the long run. Furthermore, observed productivity often depends on myriad factors, making the worker's marginal contribution impossible to identify. Overall production is a function of both quantity and quality. Usually, quality is hard to assess, and managers turn to form performance evaluations based on readily available quantitative indicators. For example, it is easy to measure the number of analyses submitted by a data analyst. However, this quantitative measure is not a good proxy for the marginal value of the work for the firm. In such cases, remuneration based on the observable quantitative measures may be suboptimal as it distorts incentives resulting in an overall reduction in productivity (Baker, 1992, 2002; Gibbons, 2005; Holmstrom and Milgrom, 1991).

As formulated by Hoskin (1996), Goodhart's law states that when a measure becomes a target, it ceases to be a good measure. Work hours provide a natural example for a quantitative measure that is easy to monitor and correlates positively with productivity. However, basing remuneration on work hours incentivizes workers to prioritize time-consuming tasks, potentially lowering overall productivity. Thus, when production is not easily defined and measured, managers' assessment of worker productivity is based on aggregate impressions, in which quantitative indicators may loom large. 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, Sassaman, and Zhu, 2020) or performance assessment (Moore et al., 2010). Even if managers are aware that quantitative measures such as work hours are invalid, they may be unable to ignore them.

This line of reasoning suggests a somewhat counter-intuitive implication. It is possible that providing managers with *richer* information leads to *lower* productivity, as the additional information draws attention from more valid—though less reliable—indicators of productivity. For example, when workers work in the office, managers can observe many aspects of work. Some of these aspects, such as the work hours discussed above, are highly salient in the office but are not directly linked to productivity. In contrast, managers observe fewer indicators when workers work from home, but these indicators are more closely related to productivity. 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, Foroughi, and Larson, 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, Bloom, and Davis, 2021; DeFilippis et al., 2020).

In this paper, we explore this argument to show that the structure of the signal observed by the manager matters. We assume that the manager observes a noisy signal of the worker's effort. This signal is composed of quality and quantity elements. We hypothesize that if we shut down the quantity element—thereby providing less information in the signal—effort will be more focused on quality, and overall productivity will increase. We illustrate this argument in a simplified model and test it in a controlled laboratory experiment.

Laboratory experiments are a natural venue of investigation when considering moral hazard and signaling. To test hypotheses relating to observability, the researcher should be able to measure effort and productivity independently of the signal that the manager observes. In a natural work environment, this is generally not feasible. In our laboratory experiment, in contrast, we can simultaneously observe real productivity and control and manipulate the characteristics of the signal observed by the manager.

We implement an incomplete contract environment, where a manager and a worker interact repeatedly. In each period, the manager observes a noisy signal of the worker's productivity and chooses a bonus to pay the worker. We manipulate whether the noise element in the signal is sensitive to the work quantity or is purely random. These settings aim to capture environmental differences, e.g., between working at the office—where quantitative measures such as work hours factor into the manager's impression of the worker—compared to working from home—where managers observe only the final output. When the signal incorporates quantitative information, workers provide an artificially high quantity. Removing the quantitative information decreases quantity and increases quality, confirming our hypothesis that shifting resources to quantity potentially comes at the expense of quality. We analyze workers' strategies using a finite mixture model and task difficulty estimations based on machine learning methods. In the control 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.

In summary, in this paper, we study a special case of Goodhart's law. 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 following the COVID-19 pandemic.

2. Related literature

In standard principal-agent models, higher per-performance remuneration leads to higher effort. This result follows from the assumption that effort is fully observable and easy to measure, which is usually not the case. Under realistic circumstances, the principal can and does monitor indicators directly and indirectly associated with effort. 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, but only when the link between performance and the manager's payoffs was direct and continuous. However, in that case, monitoring harmed performance. Monitoring is also a function of the output demanded by the manager. When demand is reasonable, little monitoring is needed. If it is too high so that workers know they cannot hit the target, they engage in alternative activities (Engel, 2011).

Our analysis of qualitative and quantitative measures mirrors models developed in the multitasking literature, beginning with Holmstrom and Milgrom (1991) and Baker (1992). Holmstrom and Milgrom (1991) analyzed settings where the 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 production and machine maintenance. 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.

A few studies applied the multitasking framework in laboratory experiments. Fehr, Klein, and Schmidt (2001) 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.[1] 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

[1] The published version omits this experiment (Fehr, Klein, and Schmidt, 2007).

turn, yielded more efficient effort allocation.

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 increase worker productivity. Choudhury, Foroughi, and Larson (2021) found that working from anywhere is 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, Bloom, and Davis, 2021; Bick, Blandin, and Mertens, 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, Bloom, and Davis (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, Mengel, and Siemroth (2022) 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, Mengel, and Siemroth (2022) 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 suggests that Goodhart's Law can provide a unifying explanation for both increased and decreased productivity when working from home, with the key variable being the observability of quantitative indicators.

3. Model

In this section we analyze a simplified and tractable setting, where the manager can commit to the bonus scheme, and quality and quantity are symmetric with regard to effort costs and signaling properties (in Scenario B, see below). The model is closely related to the existing models in the multitasking literature (e.g., Baker, 1992, 2002; Holmstrom and Milgrom, 1991). Our focus is not on the characteristics of the optimal remuneration scheme but the effects of the signal structure on effort and productivity, specifically on the distortion effects of enriching the signal. Unlike most treatments, we assume risk neutrality on the part of the worker and a deterministic environment.

The experiment described in the next section deviates from the model assumptions in that it implements a more natural setting, where expectations of bonuses arise implicitly from norms and repeated interactions, and—as we use a real effort task—costs and impression formation are not quantifiable. Thus, the role of the theoretical analysis is to illustrate how higher observability of non-productive effort stands to undermine productivity and illuminate the underlying mechanism. This role is consistent with the applications commonly discussed in the multitasking literature.[2]

We consider a principal-agent with a single firm (manager) and a single worker. The worker chooses how much effort to invest in each of two activities, increasing either quality or quantity. Increased quantity does not affect the overall productivity of the firm. The manager only observes a combined measure of quantity and quality. Thus, the worker can choose different combinations of quality and quantity to provide the same observable outcome.

Formally, the worker can choose two effort levels $e_1 \geq 0$ and $e_2 \geq 0$ at a personal cost of $C(e_1, e_2)$, where $C : \mathbb{R}^2_+ \to \mathbb{R}_+$ is a standard cost function (i.e., strictly increasing

[2]A prominent example often raised is teacher contracts. If students' exam performance forms the basis of teacher remuneration, teachers will "teach to the test," potentially undermining the true objectives of the school system. This environment deviates from the models' assumptions, yet we understand that the theoretical insights shed light on the practical issues. In this sense, the link between the model and the experiment parallels the link between any theoretical model and the target natural environment. All models aim to teach us something about a "real-world" environment in which the model assumptions do not hold. The experiment serves as an interim setting, testing the robustness and generalizability of the insight gained from the theoretical analysis when the assumptions are relaxed to increase ecological validity.

and strictly convex in both arguments) with the following properties: (i) $C(0,0) = 0$; and (ii) all derivatives up to the third order are strictly positive. For tractability, we also assume that $C(x, y)$ is symmetric with respect to x and y, i.e., $C(x, y) = C(y, x)$ for every $x, y \geq 0$.[3] Effort e_1 is the *productive* (quality) effort, and generates a profit $b(e_1)$ to the firm, where $b : \mathbb{R}_+ \to \mathbb{R}_+$ is a strictly increasing and strictly concave payoff function with a normalization of $b(0) = 0$. Effort e_2 is the *nonproductive* (quantity) effort and has no direct effect on the firm's profits.

The manager observes an additive signal $x = f_1(e_1) + f_2(e_2)$, where $f_i : \mathbb{R}_+ \to \mathbb{R}_+$ for $i = 1, 2$, are two signaling functions. Since we later vary between the properties of f_2, let us now define a *generic signaling function* f and relate its properties to f_1 and f_2 when needed. Specifically, the generic signaling function $f : \mathbb{R}_+ \to \mathbb{R}_+$ is bounded, strictly increasing, and strictly concave with a normalization of $f(0) = 0$ and $f^{(3)}(x) < 0$ for every $x \geq 0$. Based on this signal, the manager chooses a bonus scheme $\alpha x + \beta$, where $\alpha, \beta > 0$. To shorten the analysis, we generally assume that all mentioned functions are three-times continuously differentiable.

The timeline is as follows. First, the manager commits to a bonus α and β. The worker chooses effort levels e_1 and e_2. The worker receives the bonus and pays the effort costs. The manager receives the profit based on the productive effort and pays the bonus. Formally, the manager faces the following constraint-optimization problem

$$\max_{\alpha,\beta} \quad \pi = b(e_1) - (\alpha x + \beta), \tag{1}$$

subject to the worker's maximization problem

$$\begin{aligned} \max_{e_1,e_2} \quad & U = \alpha x + \beta - C(e_1, e_2), \\ & x = f_1(e_1) + f_2(e_2). \end{aligned} \tag{2}$$

3.1. Equilibrium

We compare two scenarios, varying in whether the unproductive effort enters the observable signal x. In Scenario A ("working from home"), the manager observes only a measure of the productive effort, that is, the signaling function f_2 is constant and $f_1 = f$. In scenario B, the signaling function f_2 coincides with the generic signaling function f, i.e., $f_2 = f_1 = f$. Specifically, we compare the productive effort, denoted by $e_{1,A}$, in Scenario A to the productive effort, denoted by $e_{1,B}$, in Scenario B.

The following result shows that the transition from a constant signaling function f_2 to a non-constant and symmetric signaling structure $f_2 = f_1 = f$ results in decreased productivity. In other words, the fact that workers can signal in an unproductive way by increasing quantitative indicators harms quality.

[3]Our results can be extended beyond the symmetric-cost assumption.

Proposition 1 *The productive effort in* Scenario A *is strictly higher than the productive effort in* Scenario B.

Proof. In both scenarios and for every (α, β), the worker faces the optimization problem given in (2). Taking the FOCs, we get that the optimal productive effort solves the equation $\alpha = \frac{C_1'(e_1,e_2)}{f_1'(e_1)}$. The concavity-convexity assumptions suggest that this value, e_1, is unique, positive, and monotone in α. In addition, in case $f_2 = f_1 = f$ (Scenario B), the symmetry between e_1 and e_2 (manifested through the cost and signaling function) implies that productive and unproductive efforts are equal, $e_{1,B} = e_{2,B}$. On the other hand, if f_2 is constant, the unproductive effort equals zero, $e_{2,A} = 0$.

Now we can revert to the firm's choice of an optimal linear contract. Note that β is an additive constant with no effect over incentives, so we can fix it to be zero. Starting with Scenario A and substituting the wages in (1) based on the worker's FOC, we get the following simplified optimization problem:

$$\max_{e_1} \pi = b(e_1) - \frac{C_1'(e_1, 0)}{f'(e_1)} f(e_1).$$

To compare, Scenario B yields the following optimization problem

$$\max_{e_1} \pi = b(e_1) - 2\frac{C_1'(e_1, e_1)}{f'(e_1)} f(e_1).$$

By our assumptions regarding the cost and generic signaling functions (and their derivatives up to the third order), it is straightforward to verify that $C'(e_1, e_1)$, $C'(e_1, 0)$ and $\frac{f(e_1)}{f'(e_1)}$ are all strictly increasing and convex. As the product of two strictly increasing and convex functions is increasing and convex as well, and since $b(\cdot)$ is strictly concave, we conclude that both optimization problems above have a unique solution.

Denote $H_1(e_1) = C_1'(e_1, 0)\frac{f(e_1)}{f'(e_1)}$ and $H_2(e_1) = 2C_1'(e_1, e_1)\frac{f(e_1)}{f'(e_1)}$. Note that the convexity of C, C' and $\frac{f(e_1)}{f'(e_1)}$ implies that $H_2(e_1) > H_1(e_1)$ and $H_2'(e_1) > H_1'(e_1)$ for every $e_1 \geq 0$. Thus, once we compare the FOCs of the two optimization problems

$$\begin{aligned} b'(e_{1,A}) &= \frac{dH_1(e_{1,A})}{de_1}, \\ b'(e_{1,B}) &= \frac{dH_2(e_{1,B})}{de_1}, \end{aligned}$$

it is evident that $e_{1,A} > e_{1,B}$ since the RHS of the first equation is point-wise strictly smaller than the RHS of the second equation. ∎

4. Experiment

4.1. Experimental design and procedure

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. 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 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

[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.

in the HOME treatment to reflect the expected TNA under the null hypothesis of no incentive distortion. Based on pilot sessions, we set $X = 7$.[6]

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.[7] 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.

4.2. Hypotheses

In the OFFICE treatment, workers can increase the expected value of the signal by solving more problems. We, therefore, expect that the employees submit more answers in OFFICE than in HOME. Moreover, we hypothesize that shifting effort to increase the total number of answers comes at the expense of productivity. That is, the workers make more mistakes compared to HOME and submit overall fewer correct answers. The first hypotheses summarize these predictions.

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

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

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

[7]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.

We assume that workers can (roughly) estimate the difficulty of an addition problem. Given the task structure, a worker who aims to maximize productivity should skip problems that appear difficult and time-consuming (by submitting random answers) and focus on easy problems, which yield a high return on time invested. A worker who aims to increase the TNA, in contrast, should skip problems regardless of their difficulty. Thus, we can identify a shift in workers' strategies if we can identify a 'skipping decision' and model the decision to skip based on problem difficulty. We describe this analysis further in Section 5.2. Here we state the following hypotheses:

Hypothesis 3 *Workers skip more in the OFFICE treatment.*

Hypothesis 4 *Workers are more likely to skip a problem as the problem difficulty increases.*

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.

5.1. Outcomes

Figure 1 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 Table 1, 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 the HOME treatment ($p < 0.01$). When the employer observes quantitative information, workers increase their non-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.*

Moreover, workers correctly solved, on average, 0.3 fewer problems in the OFFICE than in the HOME treatment ($p < 0.05$). These results establish that the increase in quantity comes at the expense of quality, supporting Hypothesis 2.

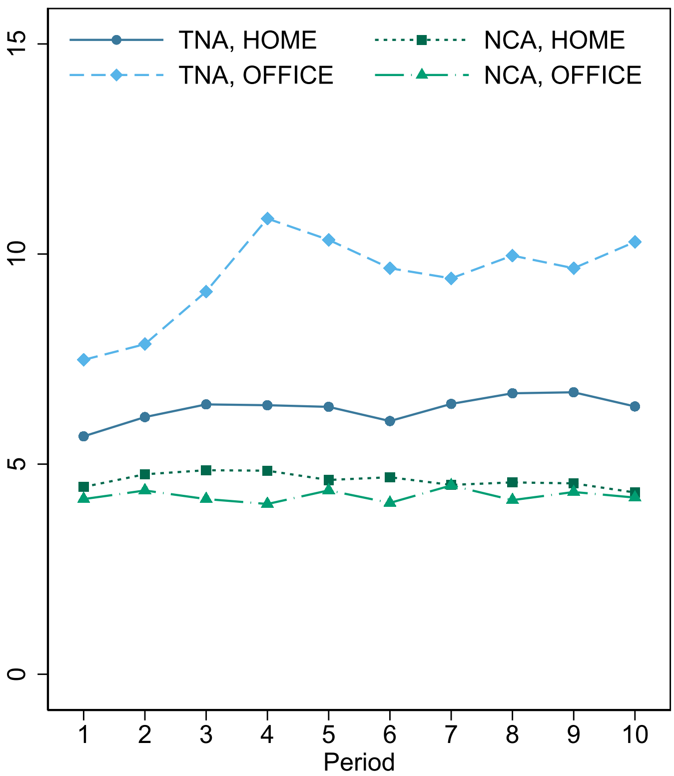


Figure 1: Total and correct answers.

Table 1: Treatment effects.

	(1) TNA	(2) TNA	(3) TNA	(4) NCA	(5) NCA	(6) NCA
OFFICE	3.276***	3.144***	3.156***	−0.347**	−0.325**	−0.326**
	(3.790)	(3.959)	(3.946)	(−2.407)	(−2.322)	(−2.331)
Lagged wage	−0.008	0.017	0.014	0.033***	0.029***	0.030***
	(−0.288)	(1.237)	(1.016)	(3.295)	(5.223)	(5.332)
Period	0.096	0.098	0.098	−0.019	−0.018	−0.018
	(1.463)	(1.494)	(1.489)	(−1.033)	(−0.989)	(−0.991)
Constant	6.018***	5.466***	5.583***	3.929***	3.989***	3.960***
	(8.343)	(7.238)	(12.263)	(10.150)	(24.649)	(13.636)
Individual effects	No	FE	RE	No	FE	RE
N	1446	1446	1446	1446	1446	1446

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

Table 2: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

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

Further investigation reveals an asymmetric effect with respect to the order of the treatments. Table 2 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 doesn't 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 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 turn now 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, the solving time follows a lognormal distribution with a higher mean, and depends on the problem difficulty.[8] The histogram in Figure 2 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 overlayed over the histogram in Figure 2. 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%.[9] See Appendix A for more details.

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

[9]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%.

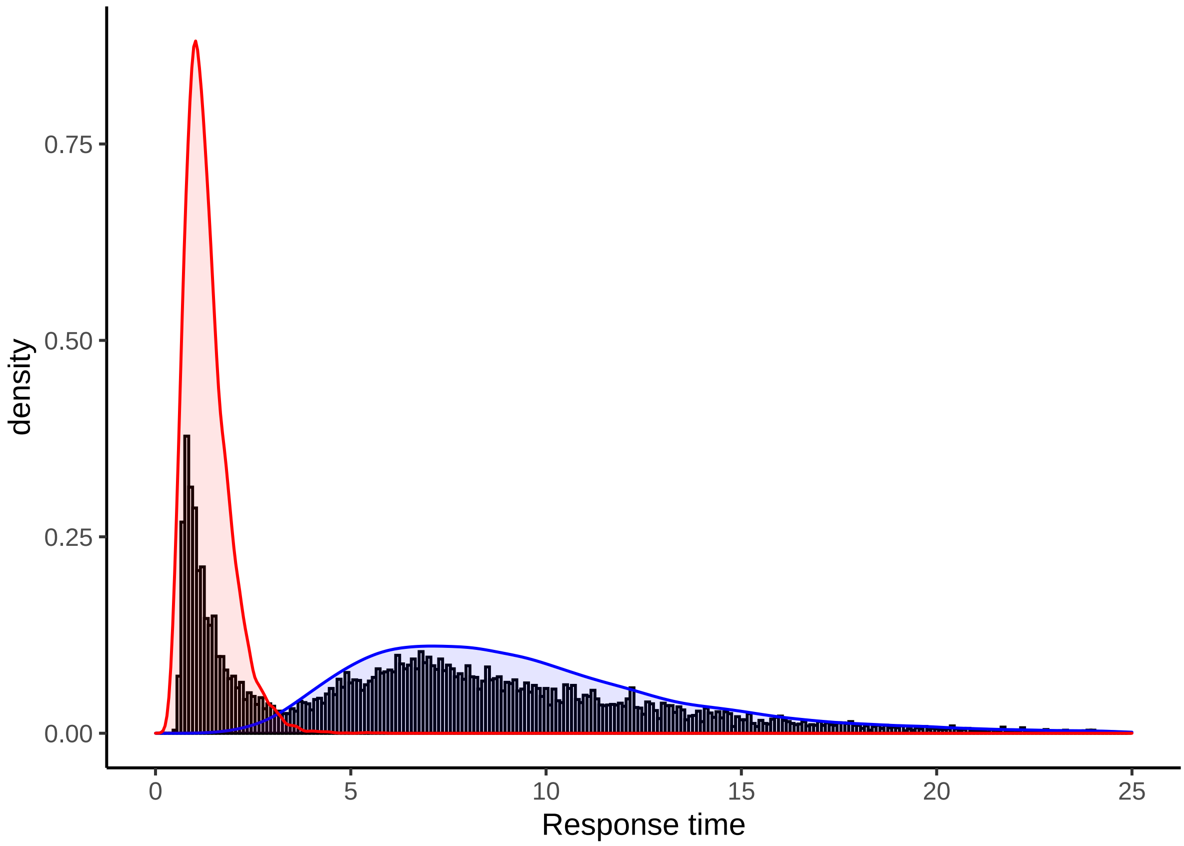


Figure 2: Empirical and modeled distributions of aolving time.

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.[10] 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.[11] 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

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

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

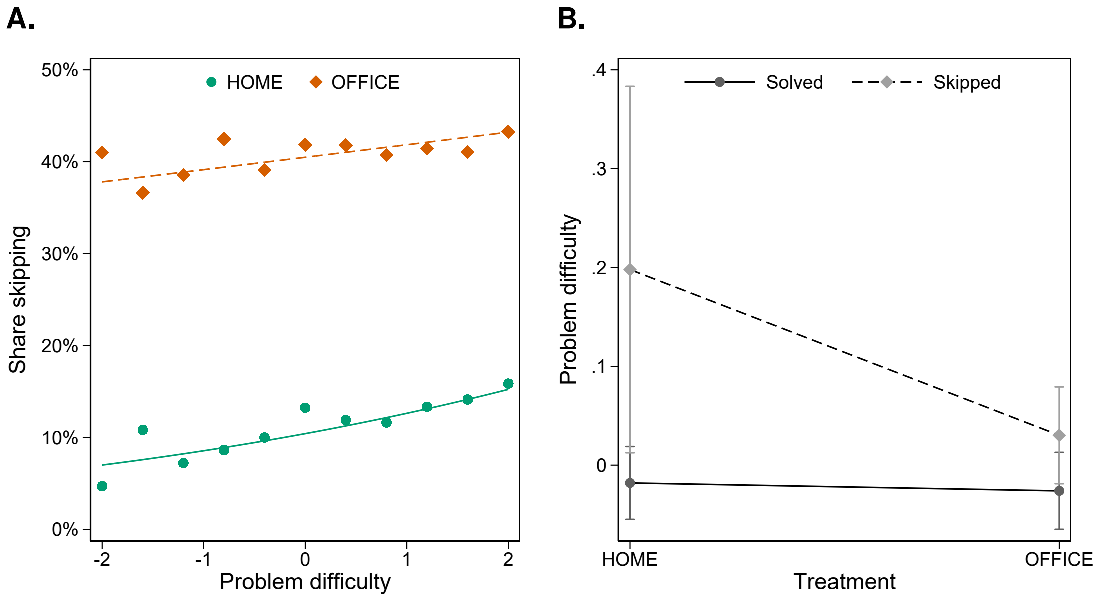


Figure 3: Problem difficulty and skipping.

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.[12] 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 Appendix B.

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 3 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. Within treatments, more difficult problems are more likely to be

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

Table 3: 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. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

skipped in the 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 3 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 3. Support for Hypotheses 4 and 5 is strongest when workers transition from OFFICE to HOME. The following result summarizes the strategy analysis.

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 analyzes 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 counterintuitive 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 "nondiagnostic" to the task at hand (Nisbett, Zukier, and Lemley, 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 some of the explanation for recent research pointing at the benefits of working from home (Angelici and Profeta, 2020; Bloom et al., 2015; Choudhury, Foroughi, and Larson, 2021).

Finally, this study focuses on the effect of the signal structore 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.

7. References

Angelici, Marta and Paola Profeta (2020). Smart-working: work flexibility without constraints. *CESifo Working Paper* 8165.

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. *FRB of Dallas Working Paper* 2017.

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.

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 (2020). Collaborating during coronavirus: the impact of covid-19 on the nature of work. *NBER working paper* (27612).

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.

Engel, Russell (2011). Why work when you can shirk?: worker productivity in an experimental setting. *Journal of Applied Business and Economics* 11(2).

Fehr, Ernst, Alexander Klein, and Klaus M. Schmidt (2001). Fairness, incentives and contractual incompleteness. *University of Zurich IEW working paper* 72.

— (2007). Fairness and contract design. *Econometrica* 75(1), pp. 121–154.

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).

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

Gibbs, Michael, Friederike Mengel, and Christoph Siemroth (2022). Work from home & productivity: evidence from personnel & analytics data on it professionals. *Journal of Political Economy: Microeconomics*. Forthcoming.

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.

Hoskin, Keith (1996). The 'awful idea of accountability': inscribing people into the measurement of objects. In: *Accountability: power, ethos and the technologies of managing*. Ed. by R. Munor and J. Mouritsen. London: International Thomson Business Press, pp. 265–282.

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

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.

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.

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.

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.

Appendix A. 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 A1 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.

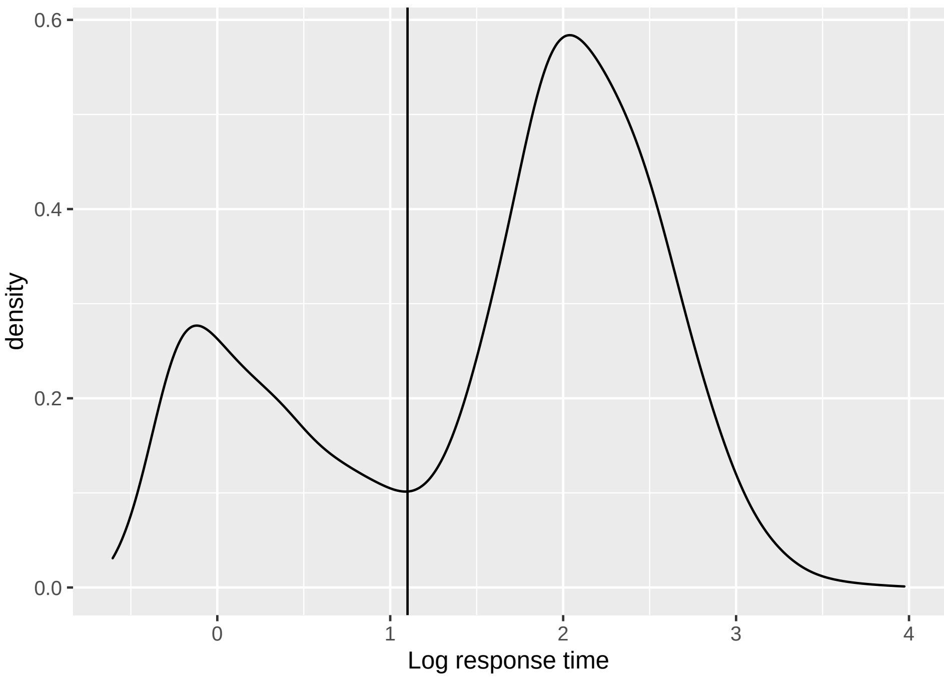
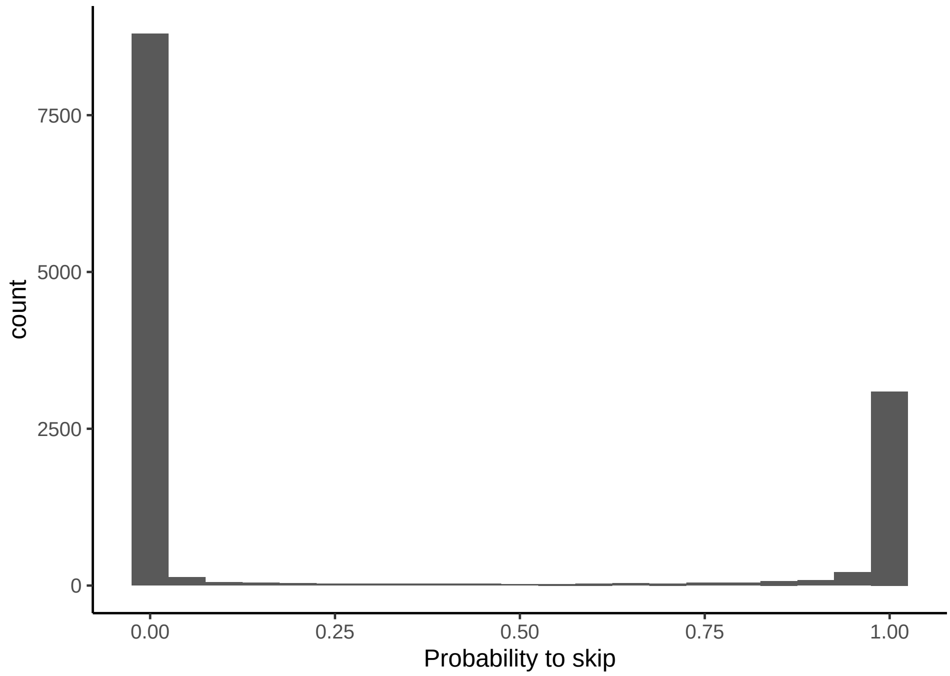


Figure A1

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 A2 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 A2

Appendix B. 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.

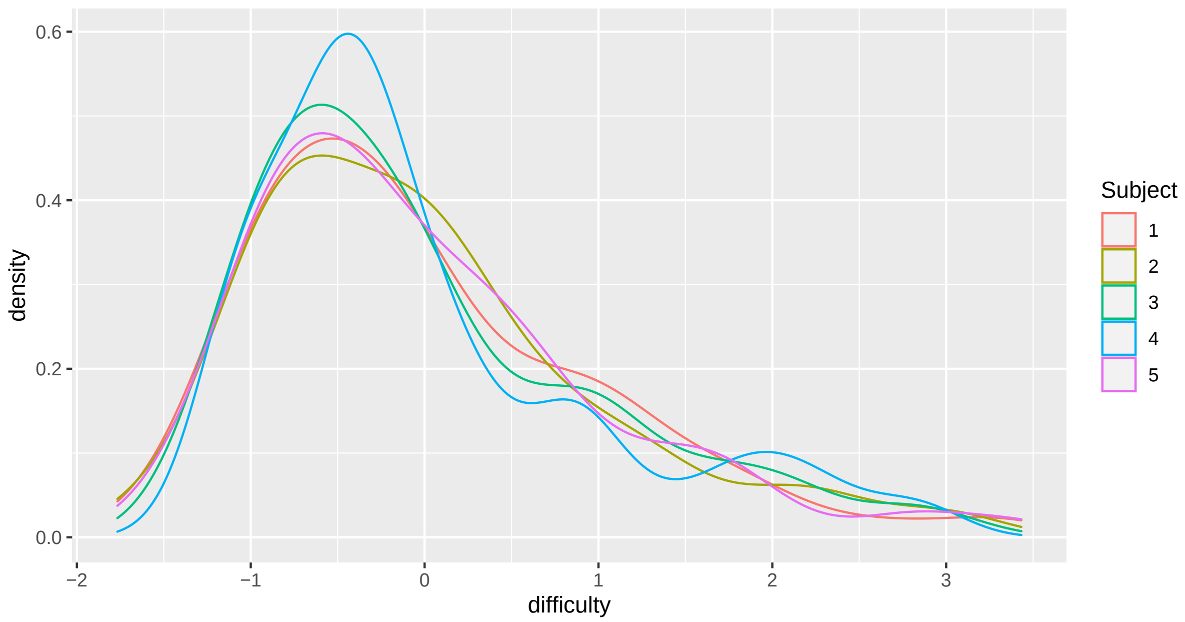


Figure B3: Distribution of problem difficulty in the training set.

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 B3.

As predictors, we used indicators for the features listed in Table B1. 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 B1: 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.

Appendix C. 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 mathematic 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 mathematic 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.