NBER WORKING PAPER SERIES

RISK MANAGEMENT AND THE AUTONOMY OF LABOR

Andreas J. Dambaur
Keith Marzilli Ericson
Johannes G. Jaspersen
Sandra Zoller

Working Paper 30793
http://www.nber.org/papers/w30793

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2022

We acknowledge funding from the DFG (German Research Foundation) under grant number JA 2875/1-1. This RCT was registered in the American Economic Association Registry for randomized control trials under trial number AEARCTR-0008556. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Risk Management and the Autonomy of Labor
Andreas J. Dambaur, Keith Marzilli Ericson, Johannes G. Jaspersen, and Sandra Zoller
NBER Working Paper No. 30793
December 2022
JEL No. D90,J24,M54

ABSTRACT

Foresightful workers can take actions to reduce their exposure to risk in labor markets, but existing evidence on narrow bracketing suggests that individuals might not optimally integrate risk reduction decisions with subsequent labor decisions. In an online labor market, we vary the level of worker autonomy when transcribing a set of documents and measure the willingness to do preventative tasks to reduce the expected total task length. If workers integrate their risk management decisions with subsequent labor effort, the high-autonomy group should be more likely to engage in prevention, since they can adjust their work to accommodate the preventative effort. In contrast, if workers narrowly bracket the decision, it should not be affected by autonomy. We find that workers in the high autonomy condition are more likely to undertake prevention, consistent with an integrated view of risk management and later labor decisions. Our results have implications for designing effective incentive systems, particularly in times of the gig-economy and rising popularity of working from home.

Andreas J. Dambaur
Ludwig-Maximilians-Universität of Munich
School of Management
Andreas.Dambaur@lmu.de

Keith Marzilli Ericson
Boston University Questrom School
of Business
595 Commonwealth Avenue
Boston, MA 02215
and NBER
kericson@bu.edu

Johannes G. Jaspersen
Ludwig-Maximilians-Universität of Munich
Schackstr. 4
80539 Munich
Germany
jaspersen@bwl.lmu.de

Sandra Zoller
iptiQ
sandra_zoller@iptiq.com

A randomized controlled trials registry entry is available at
https://www.socialscienceregistry.org/trials/8556/

1 Introduction

Workers make many decisions to manage the risks they face, both big and small. An individual may avoid a career path with higher expected earnings if it also has a higher variance, and an individual may do extra work on a project upfront to reduce the chance that it needs to be revised. Autonomy at work has increased with the rise of the gig-economy and a shift toward working from home. Workers must decide how to allocate effort across tasks and how much total effort to put forth. Similarly, managers must decide how much autonomy to give to their workers. Understanding how worker decisions are made will inform our models and enable managers to design more effective incentive systems.

To optimally manage risk, an individual or firm must have an integrated view of the risks they face (Gamba and Triantis, 2014). When a worker with multiple projects has one that becomes more flexible or more risky, that should affect how they approach the other projects. However, individuals may narrowly bracket their decisions and not integrate different choices into one coherent strategy (Read et al., 1999; Rabin and Weizsäcker, 2009; Simonsohn and Gino, 2013). This may result from simply making decisions in isolation (Barberis et al., 2006) or from cognitive frictions that limit their ability to consider all relevant information (Lian, 2021). However, the evidence here is mixed, leaving open the question of whether and when workers approach projects in an integrated way.[1]

We study how individuals approach risk management at the level of a single project. We run a field experiment in an online labor market in which workers transcribe scanned documents. We examine workers' decisions to undertake additional preventative tasks (transcribe more documents) to avoid the risk that they must transcribe longer documents. An integrated view of risk management in this context requires individuals to consider how flexible they will be when completing the task, as that affects the cost of prevention. In our study, workers can skip a limited number of transcriptions. We randomly vary the flexibility that workers have: the high-autonomy group can flexibly allocate their total number of skipped transcriptions, while the low-autonomy group must allocate their skipped transcriptions in a prescribed manner. In our study, the high-autonomy group has a lower effective cost of engaging in prevention because

[1] We do observe individuals adjusting financial decisions in light of changes in their risk exposure. Gruber and Yelowitz (1999), for example, find that people save less once they are covered by health insurance and Gallagher et al. (2020) show that this effect depends on the more general financial situation of the household. Additionally, households utilize a variety of different channels after a wage shock to keep consumption levels relatively stable (Kochar, 1999; Heathcote et al., 2014; Blundell et al., 2016).

they can allocate their effort more flexibly and thus can more easily accommodate additional preventative effort. Thus, if workers integrate their choices in these two domains, the high-autonomy workers will be more likely to engage in risk-reducing prevention. Consistent with the predictions that workers integrate their risk management choice with their later effort decisions, we find that workers in the high-autonomy group are 9.7 percentage points more likely to engage in risk-reducing preventative effort. This constitutes a relative increase of about 23%, implying an economically meaningful effect.

Our experiment captures dynamics that are common beyond this particular decision. Preventative effort must come at the beginning. Thus, if low-autonomy workers need to meet stringent performance targets according to a fixed schedule, engaging in prevention can lead to particularly high effort costs in the beginning. Consider, for example, programmers coding an app. If they want to reduce the risk of additional work late in the project, they can spend time on planning the app's architecture such that the integration of new features is simple. However, if they have to deliver app features on a weekly basis or commit a certain number of code lines each day, any time spend on planning architecture leads to more stress at the beginning of the project. Thus, they might choose to forego these investments in reducing later risk (incurring what is termed "technical debt"). More flexible target schedules let them spread the cost of prevention over time, making it effectively cheaper.

This study helps us understand how risk management affects the provision of effort in labor markets. Workers here take into account that project flexibility impacts the effective price of risk management, and accordingly adjust their investment in prevention. This is not always the case in labor choices, where some evidence suggests individuals are myopic and bracket their labor choices narrowly (e.g., taxi drivers "taking one day at a time" in Camerer et al., 1997). Some studies have shown that people ignore at least part of their choice environment when making decisions such as stock market investments (Barberis and Huang, 2001; Barberis et al., 2006) or college admission decisions (Simonsohn and Gino, 2013). Our work is more consistent with laboratory evidence that individuals consider background factors such as risk and (to a certain extent) wealth when making decisions under risk (Beaud and Willinger, 2015; Andersen et al., 2018; Hinck et al., 2021) and thus validates the conjecture of Dohmen (2014) that narrow bracketing in labor choices needs to be researched further to obtain a comprehensive view of the phenomenon.

We also contribute to the discussion on the value of labor autonomy. There has been increased interest in the role of autonomy in labor decisions with the rise of the gig-economy in fields such as transportation, food deliveries, or professional services (Cook et al., 2021). Working from home also increases autonomy, and has received increased attention with the onset of the COVID pandemic and the resulting changes in work arrangements, many of which may be permanent (Dingel and Neiman, 2020; Barrero et al., 2021). Workers seem to value autonomy in its own right, irrespective of its instrumental value (Falk and Kosfeld, 2006; Bartling et al., 2014). Autonomy has instrumental value as well, with flexibility in the gig-economy leading to higher worker surplus (Chen et al., 2019). Similarly, working from home increases both worker productivity and job satisfaction (Bloom et al., 2015). A potential downside of autonomy is that workers might not exercise it optimally or start to shirk (e.g. Gjedrem and Rege, 2017; Hoffman et al., 2018; Kesavan and Kushwaha, 2020). In our experiment, however, workers are able to respond in a sophisticated manner to changes in the cost of reducing risk. Managers should account for this tradeoff when allocating autonomy to workers.

The paper proceeds as follows. We describe the experimental setting of our study in Section 2. The next section provides a theoretical framework and derives our hypothesis. Sections 4 and 5 describe the data and our results. The last section discusses our findings and concludes.

2 Experiment

We conducted a field experiment in an online labor market (Amazon's MTurk platform) to identify the causal effect of adding work autonomy on risk management choices, which allows us to make inferences about whether workers integrate their labor and risk management choices.

We posted a standard transcription task on MTurk in June 2022, advertising a fixed payment of $8 upon successful participation. To participate in our task (or HIT, as it is called on MTurk) workers had to first accept the HIT. They were then provided with a link to the decision interface, which was implemented via Qualtrics. Based on a power test, we recruited the pre-registered number of 500 individual workers.[2] 476 of the 500 workers successfully completed the HIT and

[2] We pre-registered our study in the AEA RCT Registry: `https://doi.org/10.1257/rct.8556`. Workers who had already participated in one of our test-runs were not eligible to participate. Furthermore, to qualify for participation, workers had to (i) be approved for at least 100 HITs, (ii) have an HIT approval rate for all requesters' HITs greater than 95%, (iii) and be located in the US. We used the procedure outlined in Kennedy et al. (2020) to detect VPS/VPN/Proxy usage through IPHub API. This way, we could verify that workers were indeed located in the US. If the verification failed, workers were excluded from participation. In addition, participation was only possible via a computer. If participants tried to participate via a mobile device, we

received payment.[3] We randomly assigned the workers to either the *low-autonomy* condition or the *high-autonomy* condition. In both groups, workers were tasked with transcribing scanned receipts. We chose this task because it is common on MTurk.

Each worker had to transcribe two *sets* of receipts. A single *receipt* consisted of one or more lines and had to be transcribed line by line. Receipts were in German language to make the task more challenging, but we avoided using any characters not available on an English language keyboard. An example of a receipt and the transcription screen is given in Figure 1. Each of the two receipt sets consisted of ten receipts which differed in the number of lines. With equal probability, workers either received two sets from a *collection* consisting of only short receipts or two sets from a collection with only long receipts. A set from the short collection had one receipt with only one line and then receipts increased in length in increments of two lines up to a maximum length of nineteen lines. Receipts in the long collection start at a length of eight lines and increase in increments of two up to a maximum length of twenty-six lines. A set of receipts is always drawn from a collection in such a way that it contains exactly one receipt of each length included in the respective collection. Thus, the average receipt length for sets from the short collection is always ten lines, while for sets from the long collection it is always seventeen lines. The receipts from one set were vertically arranged on one page. The order in which the receipts within one set were displayed was randomized, such that they were not ordered from short to long or from long to short. Such an ordering would have been too artificial and might have weakened the field experiment setting. As soon as workers indicated that they had finished working on the first set by clicking on "next", they could immediately start working on the second set.

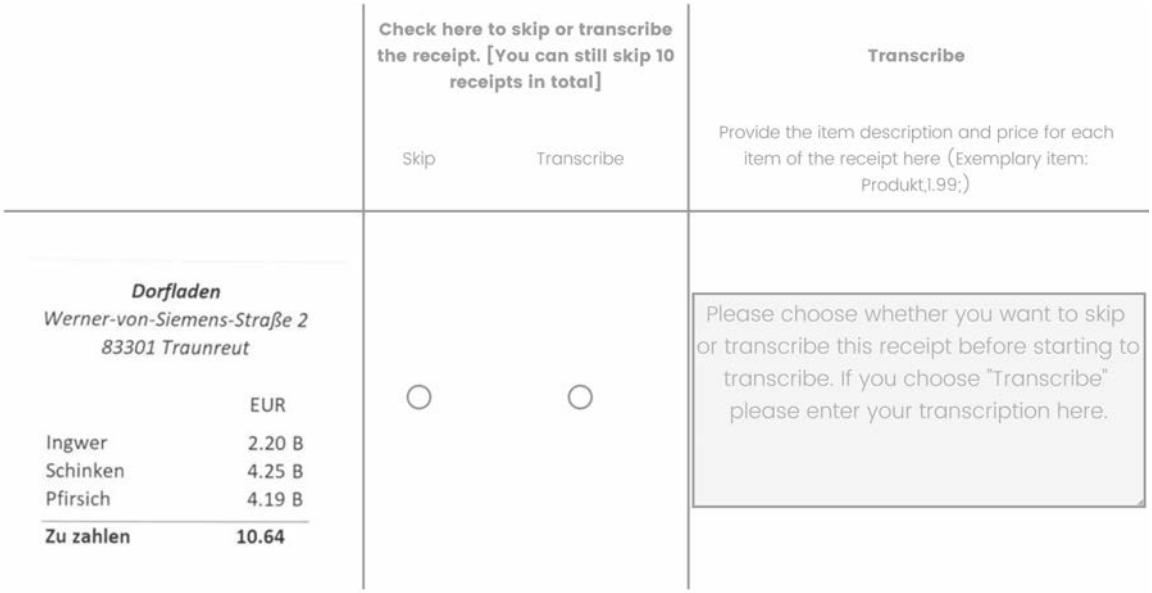
Workers were allowed to skip the transcription of ten receipts, but differed by treatment in how much discretion they had in how to allocate their effort. In the low-autonomy group, they could skip five receipts in each of the two sets. In the high-autonomy group, they could skip ten receipts combined, allocated however they wanted across both sets. The number of skippable

instructed them to use a computer or they would not be able to participate. Even though we set up the Qualtrics survey in a way such that starting the survey multiple times should be impossible, we identified 45 workers in our sample who nevertheless accessed it more than once. We repeat all analyses of the paper without these 45 workers. Results do not change in sign or significance. See Appendix A.3 for details.

[3] The remaining 24 workers claimed to have completed the HIT, but the payment code entered on the MTurk platform was not generated through our Qualtrics experiment and there were no records of successfully completed tasks that could be matched to these workers. None of them complained about not being paid, which further corroborates that they did not complete the task.

receipts left was shown to the worker at every decision (see Figure 1). Our results show that workers sensibly skip the longest receipts in a set and choose to transcribe the shortest.[4] Thus, the structure of the sets implies that the total number of lines to be transcribed is a convex function of the number of transcribed receipts.

Figure 1: Sample transcription screen



Note: The figure displays an example for the task of transcribing a single receipt from a set of ten for a worker in the high-autonomy group who has chosen not to engage in preventative effort to guarantee a draw from the short collection. The receipt shown here has three lines and could only appear in the short collection. The number of receipts that can still be skipped was updated automatically. While workers in the high-autonomy group saw how many receipts they could still skip in total, workers in the low-autonomy group saw how many receipts they could still skip in the current set. Grey text in the transcription box would disappear as soon as the first character is entered in it. The text box was unlocked for input and its background color changed to white only if the worker previously checked "Transcribe".

The full task consisted of three stages: explanation, risk management, and transcription.[5] In the explanation stage, workers were told about the transcription task and had to complete one sample transcription. In this sample, they saw two receipts and were told to skip the longer of the two and transcribe the shorter one. This pre-test served to ensure that workers know how to correctly transcribe a receipt and how to indicate whether to skip or transcribe a receipt. It was also the only transcription in the entire task which was checked for accuracy. Workers were allowed to continue if they correctly chose to skip the longer receipt, chose to transcribe the shorter receipt, and their transcription provided for the shorter receipt was 100% accurate

[4] See Figure A1 in Appendix A.1 for a graphical analysis of the workers' efficiency at skipping receipts.

[5] See Appendix B for the full instructions provided to the workers.

ignoring capitalization. Workers had the chance to modify their choices and transcription until they fulfilled these requirements.

Transcriptions in the main part of the task were not checked for accuracy, because this would have been too artificial for a field experiment. However, as is the case in most transcription tasks on MTurk, we informed workers that they would only be paid the \$8 reimbursement if they successfully completed the task. For each receipt, we verified that the workers indicated whether they wanted to skip or transcribe the respective receipt once they submitted the set. We also checked how many skips workers had used. If they had used more skips than allowed, we informed them and asked them to transcribe the remaining demanded number of receipts. Otherwise, their task was considered incomplete and they were excluded from our experiment without receiving any payment.[6] This kept workers from shirking. For workers in the low-autonomy group, this check was performed after each of the two sets because these workers had a clearly defined number of allowed skips for each set. For workers in the high-autonomy group, this check was performed after set two because these workers only had a clearly defined number of totally allowed skips for both sets.

Next, in the risk management stage, workers were told how many receipts they would have to transcribe and saw a sample set of receipts for each collection as is shown in Figure 2. Workers could then choose to undertake prevention such that both sets of receipts came from the short collection. To do this, workers had to give up part of the option to skip receipts. π denotes how many of their ten skips workers had to give up. It was set randomly with $\pi \in \{2, 3, 4, 5\}$. Choosing prevention thus had the consequence that the workers had to transcribe π additional receipts. Workers in the low-autonomy group had to do this additional work in the first transcription set and thus the number of skips was reduced there. Workers in the high-autonomy group were able to choose how to split the additional receipts between set one and set two at their own discretion. This was the consequence of the high autonomy in their labor choices.

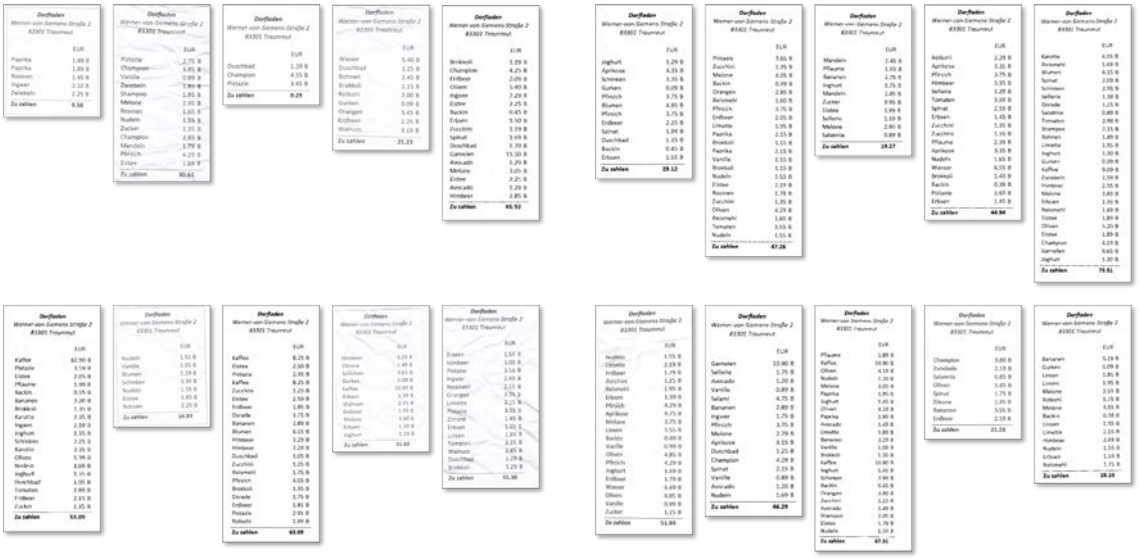
The third stage of the task was the actual transcription. It was set up as described above. After completing it, workers received a payment code which they could enter on the MTurk website to claim their reimbursement.

[6] Of the 24 subjects who did not complete the experiment, 11 could not generate a payment code because they skipped too many receipts and did not transcribe the remaining demanded number of receipts upon being prompted to do so.

Figure 2: Set examples

(a) Short collection

(b) Long collection



Note: The figure displays two example sets of receipts, one from a short and one from a long collection. Note that receipts in the short collection example did not start with a length of one, but rather started with a length of three lines. These examples were intended to illustrate the concept of a short vs. a long collection.

We framed this study as a field experiment and tried to uphold the integrity of this design choice as much as possible. Nevertheless, workers may still have inferred that they are not in a standard MTurk task because of the consent form, the sample transcription, and the risk management decision. However, our subject pool consists of real workers, working in their natural environment, on a task that is natural to them. Further, although we manipulated the composition of the receipt sets according to the procedure described above, workers were not aware of the specific structure of the receipt sets from the short vs. the long collection. Workers were only informed that a set of receipts consists of ten receipts of varying length and that both sets are drawn randomly from either a short collection containing (on average) shorter receipts or a long collection containing (on average) longer receipts. Therefore, the task retains its naturalness and is closer to a field experiment than an online laboratory experiment. Following Harrison and List's (2004) taxonomy, our experiment is not a "natural field experiment" but can be classified as a "framed field experiment".

3 Theory and hypothesis

We use a simple theoretical framework to analyze the effect greater autonomy should have on workers' preventative effort decisions in our experiment. Workers maximize expected utility over

the effort provided in transcribing the two receipt sets. We measure effort as the total number of lines which agents have to transcribe during the experiment. As described above, workers are not fully informed about the structure of the experiment. However, they do know they will transcribe two sets of ten receipts and can skip five receipts per set (low-autonomy group) or ten over both sets (high-autonomy group). They know that receipts are drawn from a long or a short collection and that receipts differ in length within a collection. This information can be summarized in a function $c(x,t)$ for each set, where x refers to the number of receipts to be transcribed, such that x equals 10 minus the number of skipped receipts in the set. $t \in \{s,l\}$ denotes whether the receipts are drawn from the short or the long collection. $c(x,t)$ then describes the total number of lines to be transcribed given x receipts from a collection of type t, satisfying $c_x(x,t) > 0 \; \forall \; t$ and $c(x,s) < c(x,l) \; \forall \; x$. Since workers can skip the longest receipts in a set, we can assume $c_{xx}(x,t) > 0 \; \forall \; t$. The structure of the experiment makes this function separable over the two sets. Utility is then formed over the total number of lines transcribed in both sets.[7] For ease of exposition, we assume $c(x,t)$ and the utility function to be twice continuously differentiable. Because transcribing requires effort, the utility function is decreasing, such that $U'(x) < 0$ for all values of $x \geq 0$.[8] Workers maximize their expected utility by choosing whether to undertake prevention and, if applicable, by deciding how many receipts to skip in each set.

When workers do not choose prevention, receipts are either short ($t = s$) or long ($t = l$). Each possibility has a probability of 50%. In the low-autonomy group, workers can skip five receipts per set and have no autonomy in their labor choices other than that they can choose to start at the shortest receipt. Workers in the low-autonomy group (superscript L) who do not undertake preventative effort (subscript R for risky) thus obtain expected utility

$$EU_R^L = \frac{1}{2}U(2c(5,s)) + \frac{1}{2}U(2c(5,l)). \tag{1}$$

Alternatively, these low-autonomy workers can undertake prevention (subscript P) and guarantee a draw from the short collection by transcribing π additional receipts in the first set, leading

[7] The form $U(c_1 + c_2)$ assumes that workers have no preferences regarding the set in which they transcribe the lines. Arguments about "time preferences" between the two sets are not applicable here, because after finishing the first set, the next one immediately begins. The number of transcribed receipts in the first set thus not only influences the effort in this set, but also how long the set lasts for the worker (or how soon the next set begins).

[8] Evidence for utility being concave as well as decreasing in experimental effort is, e.g., provided in Augenblick and Rabin (2019). We analyze whether this is true in Appendix A.5. Note, however, that such an assumption is not required for Proposition 1 to hold.

to expected utility

$$EU_P^L(\pi) = U(c(5+\pi, s) + c(5, s)). \tag{2}$$

In the high-autonomy group, denoted H, the workers can allocate their skipped receipts however they like between the two sets. Without prevention, those receipts are either from the short collection or from the long collection, but they are from the same collection in both sets. Thus, if high-autonomy workers (superscript H) choose the risky (subscript R) option, they have expected utility

$$EU_R^H = \frac{1}{2}U\left(\min_{x\geq 0}[c(x, s) + c(10-x, s)]\right) + \frac{1}{2}U\left(\min_{x\geq 0}[c(x, l) + c(10-x, l)]\right). \tag{3}$$

If instead they undertake prevention (subscript P), they need to transcribe $10+\pi$ receipts from the short collection across both sets, and their expected utility is

$$EU_P^H(\pi) = U\left(\min_{x\geq 0}[c(x, s) + c(10+\pi-x, s)]\right). \tag{4}$$

The prediction for differential behavior between the two treatments is given in the following proposition.[9]

Proposition 1. *When agents have the autonomy to move effort between sets and integrate their labor and risk management decisions, they are more likely to undertake prevention to draw from the short collection.*

With autonomy, workers are able to spread the additional receipts required to be transcribed across both sets, leading to less additional effort overall. This implies lower effective costs of prevention, which increases take-up. However, this mechanism requires that workers consider their later choices about how they will provide effort while making their decision about whether or not to undertake prevention.

---

[9] *Proof.* We denote the maximum willingness to transcribe additional receipts for prevention in treatments L and H as $\bar{\pi}^L$ and $\bar{\pi}^H$, respectively. They are implicitly defined by $EU_P^L(\bar{\pi}^L) = EU_R^L$ and $EU_P^H(\bar{\pi}^H) = EU_R^H$. Given that $c_{xx} > 0$, we can see that for any values y and t, it follows that $\min_{x\geq 0}[c(x,t) + c(y-x,t)] = 2c(y/2, t)$. From this, we can see that $EU_R^H = EU_R^L$ and thus $EU_P^H(\bar{\pi}^H) = EU_P^L(\bar{\pi}^L)$. From $U(c(5+\bar{\pi}^L, s) + c(5, s)) = U\left(\min_{x\geq 0}[c(x, s) + c(10+\bar{\pi}^H - x, s)]\right)$, we know $1/2c(5+\bar{\pi}^L, s) + 1/2c(5, s) = c(5+\bar{\pi}^H/2, s)$. Because $c_{xx} > 0$, it follows that $1/2c(5+\bar{\pi}^L, s) + 1/2c(5, s) > c(5+\bar{\pi}^L/2, s)$ which implies $c(5+\bar{\pi}^L/2, s) < c(5+\bar{\pi}^H/2, s)$ and thus $\bar{\pi}^L < \bar{\pi}^H$. For any value of π, workers in group H are thus *ceteris paribus* more likely to choose prevention than workers in group L. □

4 Data and results

Table 1 summarizes the descriptive statistics of our sample of 476 MTurk workers, and shows that the characteristics of our workers are balanced across treatment. With a total sample average of 3.45, the randomly assigned number of additional receipts required for prevention (that is, effectively the price) does not differ statistically significantly between the two groups. Workers in the low-autonomy group are somewhat more efficient in correctly completing the pre-test (6.04 seconds/character) than workers in the high-autonomy group (7.70 seconds/character), but the difference is only marginally significant and does not extend to the main task (see Table A2 in Appendix A.2). An average of 98% for quality of transcription in both groups indicates that workers in both the low-autonomy and high-autonomy group provide high-quality transcriptions and do not shirk.

Our key result is that take-up of prevention is more likely in the high-autonomy group, consistent with the predictions of Proposition 1 for individuals making integrated decisions. 42% of the 241 workers in the low-autonomy group undertake prevention. Of the 235 workers in the high-autonomy group, 52% choose this option. The difference is statistically significant according to a two-sided t-test ($p = 0.029$) and a two-sided Fisher's Exact test ($p = 0.035$).

Table 1: Summary statistics

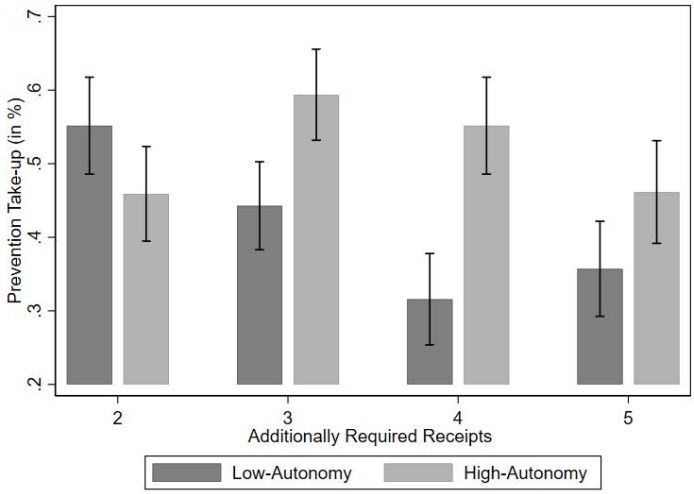
	Low-Autonomy	High-Autonomy	p-value Δ	Total Sample
Prevention Take-up	0.42 (0.49)	0.52 (0.50)	0.03**	0.47 (0.50)
Add. Receipts Required	3.46 (1.10)	3.43 (1.10)	0.76	3.45 (1.10)
Efficiency Pre-Test	6.04 (8.97)	7.70 (12.54)	0.10*	6.86 (10.90)
Quality of Transcription	0.98 (0.07)	0.98 (0.07)	0.80	0.98 (0.07)
N. Observations	241	235		476

Notes: This table reports means and standard deviations (in parentheses) in the treatments low-autonomy and high-autonomy and in the total sample. The third column presents p-values for two-sided t-tests of differences in the means. "Prevention Take-up" is a dummy variable that equals one if the worker undertakes prevention and zero else. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention. "Efficiency Pre-Test" is the time (in seconds) the worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

To illustrate the main effect graphically, Figure 3 plots the mean take-up rate separately for the high-autonomy and low-autonomy groups for each number of required receipts. The take-up rate is higher in the high-autonomy group for most, but not all of the randomly assigned prices

(i.e., required number of additional receipts). The small non-monotonicities observed in Figure 3 are likely the result of randomly chance (we are engaging in multiple comparisons here).

Figure 3: Take-up of prevention by additionally required receipts and experimental group.



Notes: This figure displays the average shares of take-up of prevention in percent for each possible number of additionally required receipts, by treatment. Error bars indicate standard errors of the mean.

To further quantify the impact of autonomy on take-up of prevention at the individual worker level, we estimate linear probability models of the form:

$$\text{Take-up}_i = \beta_0 + \beta_1 \text{High-Autonomy}_i + \beta_2 \pi_i + \delta X_i + \varepsilon_i. \tag{5}$$

Here, *Take-up*$_i$ is a dummy variable that equals one if worker i undertakes prevention and zero else. *High-Autonomy*$_i$ is an indicator variable for being randomly assigned to the high-autonomy group. π_i is the randomly determined number of additionally required receipts from worker i if they choose prevention. To improve statistical power, we include X_i, a vector of individual control variables. ε_i constitutes the error term.

The results of the estimation with heteroscedasticity robust standard errors are reported in Table 2. The estimation in Column 1 includes only the treatment assignment and the additionally required receipts as explanatory variables. In Column 2, all control variables are added. Consistent with Proposition 1, positive coefficient estimates that are statistically significant at the 5%-level in both specifications indicate that workers randomly assigned to the high-autonomy group are more likely to take-up prevention than workers randomly assigned to the low-autonomy group. The coefficient estimate in Column 2 (1) indicates that the difference in take-up is 9.7 (9.9) percentage points. Given a baseline take-up rate of 41.9%, this constitutes

Table 2: Multivariate analysis of prevention take-up

Dependent Variable	(1) Prevention Take-up	(2) Prevention Take-up
High-Autonomy	0.099**	0.097**
	(0.046)	(0.046)
Add. Receipts Required	-0.037*	-0.033
	(0.021)	(0.021)
Efficiency Pre-Test		0.003
		(0.002)
Quality of Transcription		0.750***
		(0.237)
N. Observations	476	476
Daytime FE	No	Yes
R^2	0.017	0.049

Notes: Estimations are OLS. Heteroscedasticity robust standard errors according to procedure hc-3 are in parentheses. The dependent variable is "Prevention Take-up", a dummy variable that equals one if the worker undertakes prevention and zero else. "High-Autonomy" is a dummy indicating that the worker was randomly assigned to the high-autonomy group. All control variables are defined as in Table 1. Daytime fixed effects are categorized in 6 hour increments. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

a relative increase of 23.2% (23.6%) Thus, the causal effect of workers' autonomy in setting their work schedule on their take-up decision is economically meaningful.

Consistent with the results of Figure 3, the number of additionally required receipts seems to have a negative impact on take-up. However, this effect is only significant at the 10%-level in Column 1 and turns insignificant when adding all controls (Column 2), consistent with the small non-monotonicities seen in Figure 3. Furthermore, the coefficient estimate on quality of transcription in Column 2 suggests that workers who provide more accurate transcriptions are also more likely to undertake prevention. One potential explanation is that providing transcriptions with higher accuracy is more costly and therefore, the risk of having to transcribe receipts from the long collection is more pronounced for accurate workers. This, in turn, increases the attractiveness of a guaranteed draw from the short collection. Of the remaining control variables, neither the time of day workers start working on the task nor the efficiency with which they complete the pre-test has a statistically significant impact on their decision to take-up prevention.[10]

[10] As outlined in the pre-registration, we repeat the estimations of Table 2 using Probit instead of OLS. We conduct this robustness check, because the dependent variable is binary and linear probability models like OLS could, in theory, distort the results. Table A1 in Appendix A reports the corresponding estimates. Results regarding our hypothesis remain unchanged in sign and significance.

As pre-registered, we also consider the influence of autonomy on how long workers take for their task and on how efficient they are while doing it. As can be seen in Appendix A.4, being assigned to the high-autonomy group has no effect on either outcome. This is an indication that the treatment had no unintended consequences which could have affected the prevention decision.

5 Integration of prevention and effort decisions

Our experiment's main outcome is the the take-up of prevention. However, using some additional assumptions, we can use this outcome to calculate the implied average maximum acceptable number of additional receipts for undertaking prevention in each experimental group, which we denote $\bar{\pi}^G$ for experimental group G. We assume a linear demand curve going through two points: first, $(0, 1)$, implying that at zero required receipts, all agents would undertake prevention; second, the point $(E[\pi^G], E[\text{Take-up}^G])$ which is the average take-up in group G at the average price π faced by that group. Taking the integral under the demand curve renders an average $\bar{\pi}^G$ of $\frac{E[\pi^G]}{2(1-E[\text{Take-up}^G])}$. Using these assumptions renders $\bar{\pi}^L$ of 2.98 receipts and $\bar{\pi}^H$ of 3.57 receipts.[11]

The result in Proposition 1 depends on the extent to which the agents integrate their labor choices with their risk management decisions. We now provide a back-of-the-envelope calculation about the degree to which this integration takes place. For this, we need to assume that the workers correctly anticipated the starting length of receipts in both collections and that they anticipated that the length of a receipt increases by two lines with each additional receipt. Using this information, we can infer a parametric form for the function $c(x, t)$. Assuming that workers skip the longest receipts, the function has the form $c(x, t) = x^2 + \mathbf{1}_{t=l}7x$, where $\mathbf{1}_{t=l}$ is an indicator function for the long collection.[12]

To calculate the degree of integration, we assume that agents in the low-autonomy group act according to our theoretical model and use their maximum acceptable number of additional

[11] A different calibration approach would be to calculate a ratio of the coefficients of being assigned to a group and the coefficient of the additionally required effort in a regression with take-up as the dependent variable. However, we follow our pre-registration plan. Also, the non-monotonicities documented in Figure 3 combined with the implied extrapolation of the regression results make such an approach unfeasible here.

[12] It is simple to verify that this form has the properties $c(0, t) = 0, c(x, t) > 0, c_x > 0$, and $c_{xx} > 0 \ \forall \ t$ and $c(x, l) > c(x, s) \ \forall \ x > 0$.

receipts as reference to predict that value of $\bar{\pi}$ which fully integrating agents should show in the high-autonomy group. We denote this predicted value as $\hat{\pi}^H$. From the proof of Proposition 1, we can infer that $c(a+\bar{\pi}^L, s) + c(a, s) = 2c(a+\hat{\pi}^H/2, s)$ needs to hold. Applying the functional form of c, this is equivalent to $(a+\bar{\pi}^L)^2 + a^2 = 2(a+\hat{\pi}^H/2)^2$. Full integration of labor and risk management choices would thus imply that $\hat{\pi}^H = \sqrt{4a^2 + 4\bar{\pi}^L a + 2(\bar{\pi}^L)^2} - 2a = 3.32$. This is lower than the observed value of 3.57, implying that the workers react more to work autonomy than full integration of choices would predict. However, this observed overshooting could be due to the participant's remaining uncertainty about the specific parameters resulting from our field experiment design. We thus conclude that workers fully integrate their prevention and subsequent labor choices in our experiment.

The maximum acceptable number of additional receipts in the low-autonomy group can further be used to show that the workers have concave utility over the number of transcribed lines, consistent with notion of a convex cost of effort function. The necessary assumptions and the calibration are given in Appendix A.5.

6 Conclusion

We show that gig-economy workers in an online labor market were able to engage in effective risk management by putting forth preventative effort and appropriately responding to the relative rewards and benefits of doing so. Despite previous research suggesting that individuals narrowly bracket monetary lottery choices, our results imply that they are able to broadly frame naturalistic tasks that require effort for uncertain reward.

Our study focused on the issue of broad versus narrow framing in labor choices. More complex and longer-term projects may be more difficult to manage, with additional issues involved in managing these tasks, such as prioritization and time tradeoffs. In such situations, preventative effort may be lower than optimal if the benefits are further delayed from the costs, affecting the design of optimal incentive schemes (Ericson and Laibson, 2019; Cohen et al., 2020). Running experiments in labor markets allows the researcher to isolate specific theoretical effects (see, e.g., Cooper et al., 2021; Huffman and Bognanno, 2018). Thus, we intentionally removed time tradeoffs, to focus on a single mechanism: narrow framing.

The result that flexibility in labor decisions can increase preventative activities is important for managers who need to set the autonomy of their workforce. The inherent tradeoff that on the one hand workers value autonomy and increase their productivity (Bartling et al., 2014; Bloom

et al., 2015), but on the other hand might not exercise their discretion optimally (Gjedrem and Rege, 2017; Hoffman et al., 2018), should be amended by this potential upside of flexible work arrangements. Managers should thus take into account how employees will engage in preventative effort when choosing the level of the autonomy for their workforce.

References

Andersen, S., Cox, J. C., Harrison, G. W., Lau, M. I., Rutström, E. E., and Sadiraj, V. (2018). Asset integration and attitudes toward risk: Theory and evidence. *Review of Economics and Statistics*, 100(5):816–830.

Augenblick, N. and Rabin, M. (2019). An experiment on time preference and misprediction in unpleasant tasks. *Review of Economic Studies*, 86(3):941–975.

Barberis, N. and Huang, M. (2001). Mental accounting, loss aversion, and individual stock returns. *The Journal of Finance*, 56(4):1247–1292.

Barberis, N., Huang, M., and Thaler, R. H. (2006). Individual preferences, monetary gambles, and stock market participation: A case for narrow framing. *American Economic Review*, 96(4):1069–1090.

Barrero, J. M., Bloom, N., and Davis, S. J. (2021). Why working from home will stick. Working paper, National Bureau of Economic Research.

Bartling, B., Fehr, E., and Herz, H. (2014). The intrinsic value of decision rights. *Econometrica*, 82(6):2005–2039.

Beaud, M. and Willinger, M. (2015). Are people risk vulnerable? *Management Science*, 61(3):624–636.

Bloom, N., Liang, J., Roberts, J., and Ying, Z. J. (2015). Does working from home work? Evidence from a Chinese experiment. *The Quarterly Journal of Economics*, 130(1):165–218.

Blundell, R., Pistaferri, L., and Saporta-Eksten, I. (2016). Consumption inequality and family labor supply. *American Economic Review*, 106(2):387–435.

Camerer, C., Babcock, L., Loewenstein, G., and Thaler, R. (1997). Labor supply of New York City cabdrivers: One day at a time. *The Quarterly Journal of Economics*, 112(2):407–441.

Chen, M. K., Rossi, P. E., Chevalier, J. A., and Oehlsen, E. (2019). The value of flexible work: Evidence from Uber drivers. *Journal of Political Economy*, 127(6):2735–2794.

Cohen, J., Ericson, K. M., Laibson, D., and White, J. M. (2020). Measuring time preferences. *Journal of Economic Literature*, 58(2):299–347.

Cook, C., Diamond, R., Hall, J. V., List, J. A., and Oyer, P. (2021). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. *The Review of Economic Studies*, 88(5):2210–2238.

Cooper, D. J., Saral, K., and Villeval, M. C. (2021). Why join a team? *Management Science*, 67(11):6980–6997.

Dingel, J. I. and Neiman, B. (2020). How many jobs can be done at home? *Journal of Public Economics*, 189:104235.

Dohmen, T. (2014). Behavioral labor economics: Advances and future directions. *Labour Economics*, 30:71–85.

Ericson, K. M. and Laibson, D. (2019). Intertemporal choice. In *Handbook of Behavioral Economics: Applications and Foundations 1*, volume 2, pages 1–67. Elsevier.

Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.

Gallagher, E. A., Gopalan, R., Grinstein-Weiss, M., and Sabat, J. (2020). Medicaid and household savings behavior: New evidence from tax refunds. *Journal of Financial Economics*, 136(2):523–546.

Gamba, A. and Triantis, A. J. (2014). Corporate risk management: Integrating liquidity, hedging, and operating policies. *Management Science*, 60(1):246–264.

Gjedrem, W. G. and Rege, M. (2017). The effect of less autonomy on performance in retail: Evidence from a quasi-natural field experiment. *Journal of Economic Behavior & Organization*, 136:76–90.

Gruber, J. and Yelowitz, A. (1999). Public health insurance and private savings. *Journal of Political Economy*, 107(6):1249–1274.

Harrison, G. W. and List, J. A. (2004). Field experiments. *Journal of Economic Literature*, 42(4):1009–1055.

Heathcote, J., Storesletten, K., and Violante, G. L. (2014). Consumption and labor supply with partial insurance: An analytical framework. *American Economic Review*, 104(7):2075–2126.

Hinck, S., Peter, R., and Steinorth, P. (2021). How does multiplicative background risk affect risk taking? Theoretical predictions and experimental evidence. Working paper, University of Hamburg.

Hoffman, M., Kahn, L. B., and Li, D. (2018). Discretion in hiring. *The Quarterly Journal of Economics*, 133(2):765–800.

Huffman, D. and Bognanno, M. (2018). High-powered performance pay and crowding out of nonmonetary motives. *Management Science*, 64(10):4669–4680.

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., and Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods*, 8(4):614–629.

Kesavan, S. and Kushwaha, T. (2020). Field experiment on the profit implications of merchants' discretionary power to override data-driven decision-making tools. *Management Science*, 66(11):5182–5190.

Kochar, A. (1999). Smoothing consumption by smoothing income: Hours-of-work responses to idiosyncratic agricultural shocks in rural India. *Review of Economics and Statistics*, 81(1):50–61.

Lian, C. (2021). A theory of narrow thinking. *The Review of Economic Studies*, 88(5):2344–2374.

Rabin, M. and Weizsäcker, G. (2009). Narrow bracketing and dominated choices. *American Economic Review*, 99(4):1508–43.

Read, D., Loewenstein, G., Rabin, M., Keren, G., and Laibson, D. (1999). Choice bracketing. In *Elicitation of Preferences*, pages 171–202. Springer.

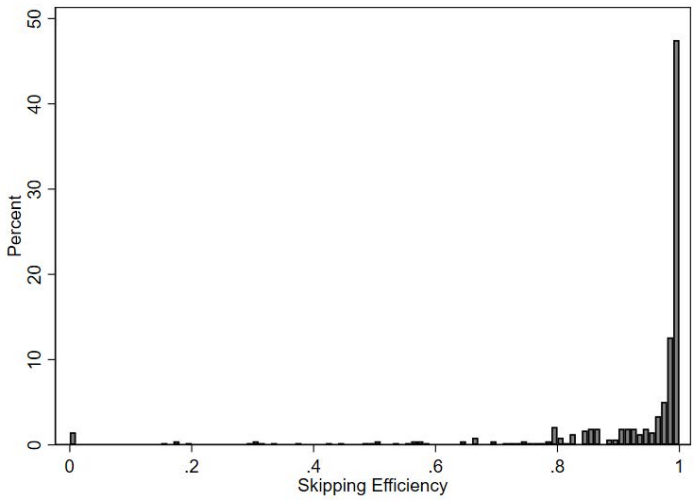
Simonsohn, U. and Gino, F. (2013). Daily horizons: Evidence of narrow bracketing in judgment from 10 years of MBA admissions interviews. *Psychological Science*, 24(2):219–224.

A Supporting Analyses

A.1 Additional analyses on the full sample

As referenced in Footnote 4, Figure A1 shows the efficiency of workers when making skipping decisions. Efficiency is calculated as the number of lines skipped divided by the maximum number of lines that could have been skipped. An efficiency of 1 thus implies the best possible efficiency. As can be seen in the graph, slightly under 50% of our sample have perfect efficiency when skipping receipts, that is, they always skip the longest receipts. In addition, most of the sample has an efficiency above 80%. This makes our assumption of a convex cost function c innocuous.

Figure A1: Worker efficiency at skipping receipts



Notes: Histogram of the distribution of skipping efficiency in the full sample. Efficiency is calculated as the number of lines skipped divided by the maximum number of lines that could have been skipped.

As referenced in Section 4, we also analyse our main multivariate analyses with a Probit estimator, to acknowledge the fact that the prevention take-up is a binary dependent variable. Table A1 shows the result of this estimation. The table shows that using a Probit estimator does not change the results in sign and significance.

Table A1: Probit analysis of prevention take-up

Dependent Variable	(1) Prevention Take-up	(2) Prevention Take-up
High-Autonomy	0.251**	0.241**
	(0.116)	(0.118)
Add. Receipts Required	-0.093*	-0.079
	(0.053)	(0.055)
Efficiency Pre-Test		0.009
		(0.006)
Quality of Transcription		6.136**
		(2.435)
N. Observations	476	476
Daytime FE	No	Yes
Pseudo-R^2	0.012	0.046

Notes: Estimations are Probit. Heteroscedasticity robust standard errors are in parentheses. The dependent variable is "Prevention Take-up", a dummy variable that equals one if the worker undertakes prevention and zero else. "High-Autonomy" is a dummy indicating that the worker was randomly assigned to the high-autonomy group. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention. "Efficiency Pre-Test" is the time (in seconds) a worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. Daytime fixed effects are categorized in 6 hour increments. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

A.2 Extended summary statistics

Table A2: Extended summary statistics

	Low-Autonomy	High-Autonomy	p-value Δ	Overall
Prevention Take-up				
Mean (SD)	0.42 (0.49)	0.52 (0.50)	0.03**	0.47 (0.50)
Median [Min, Max]	0.00 [0.00, 1.00]	1.00 [0.00, 1.00]		0.00 [0.00, 1.00]
Add. Receipts Required				
Mean (SD)	3.46 (1.10)	3.43 (1.10)	0.76	3.45 (1.10)
Median [Min, Max]	3.00 [2.00, 5.00]	3.00 [2.00, 5.00]		3.00 [2.00, 5.00]
Duration (in mins)				
Mean (SD)	49.63 (40.50)	51.61 (43.82)	0.61	50.60 (42.14)
Median [Min, Max]	36.77 [4.67, 255.45]	37.38 [10.42, 262.23]		37.10 [4.67, 262.23]
Efficiency Pre-Test				
Mean (SD)	6.04 (8.97)	7.70 (12.54)	0.10*	6.86 (10.90)
Median [Min, Max]	3.14 [0.23, 70.94]	3.12 [0.19, 71.37]		3.14 [0.19, 71.37]
Efficiency Set 1				
Mean (SD)	1.52 (1.08)	1.72 (1.77)	0.14	1.62 (1.46)
Median [Min, Max]	1.22 [0.33, 8.52]	1.31 [0.37, 15.85]		1.26 [0.33, 15.85]
Efficiency Set 2				
Mean (SD)	1.53 (2.37)	1.38 (0.96)	0.37	1.46 (1.82)
Median [Min, Max]	1.21 [0.39, 35.76]	1.23 [0.35, 10.69]		1.21 [0.35, 35.76]
Quality of Transcription				
Mean (SD)	0.98 (0.07)	0.98 (0.07)	0.80	0.98 (0.07)
Median [Min, Max]	0.99 [0.14, 1.00]	0.99 [0.24, 1.00]		0.99 [0.14, 1.00]
Daytime				
Mean (SD)	1.53 (0.60)	1.48 (0.61)	0.36	1.51 (0.60)
Median [Min, Max]	2.00 [0.00, 3.00]	1.00 [0.00, 3.00]		1.00 [0.00, 3.00]
N. Observations	241	235		476

Notes: This table reports means, standard deviations (SD), medians, minimums (Min) and maximums (Max) in the low-autonomy group, the high-autonomy group and the total sample. The third column presents p-values for two-sided t-tests of differences in the means. "Prevention Take-up" is a dummy variable that equals one if the worker undertakes prevention and zero else. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention.' "Duration (in mins)" is the time (in minutes) the worker needed to complete all parts of our experiment. "Efficiency Pre-Test" is the time (in seconds) the worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Efficiency Set 1" is the time (in seconds) the worker needed for the transcription of set 1 divided by the total number of characters contained in the receipts that the worker chose to transcribe in set 1. "Efficiency Set 2" is the time (in seconds) the worker needed for the transcription of set 2 divided by the total number of characters contained in the receipts that the worker chose to transcribe in set 2. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. "Daytime" is a categorical variable that equals zero if the worker started working on the task between 0 and 5.59 a.m., equals one if the worker started between 6 and 11.59 a.m., equals two if the worker started between noon and 5.59 p.m., and equals three if the worker started between 6 and 11.59 p.m.. All times refer to the worker's local time. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

A.3 Excluding multi-access workers

Our MTurk request was for 500 completed HITs of the transcription task. No MTurk worker was allowed to complete the task twice. We also set up the Qualtrics survey such that starting the survey multiple times should be impossible. Nevertheless, for 45 of the 476 workers in our final sample, we registered at least one additional access to the survey. Such situations could, for example, appear if the worker first accessed the survey on a mobile device, immediately closed the window without pressing the closing button on the notification screen, and then accessed the link again with a desktop computer. Even though the process would be lengthy and not worth the effort, it was possible that some workers tried to "game" our survey through multiple accesses. To make sure that such behavior, however unrealistic, did not drive our results, we show the multivariate analysis of the main paper and the Probit analysis without the 45 identified workers in this appendix. As one can see from Table A3 and Table A4, results are virtually unchanged by this sample restriction.

Table A3: Multivariate analysis of prevention take-up without multi-access workers

	(1)	(2)
Dependent Variable	Prevention Take-up	Prevention Take-up
High-Autonomy	0.121**	0.116**
	(0.048)	(0.048)
Add. Receipts Required	-0.042*	-0.037
	(0.022)	(0.022)
Efficiency Pre-Test		0.002
		(0.002)
Quality of Transcription		0.916**
		(0.365)
N. Observations	431	431
Daytime FE	No	Yes
R^2	0.023	0.055

Notes: Results are corresponding to those reported in Table 2, but estimations exclude the 45 subjects who accessed the survey more than once. Heteroscedasticity robust standard errors according to procedure hc-3 are in parentheses. The dependent variable is "Prevention Take-up", a dummy variable that equals one if the worker undertakes prevention and zero else. "High-Autonomy" is a dummy indicating that the worker was randomly assigned to the high-autonomy group. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention. "Efficiency Pre-Test" is the time (in seconds) a worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. Daytime fixed effects are categorized in 6 hour increments. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

Table A4: Probit analysis of prevention take-up without multi-access workers

Dependent Variable	(1) Prevention Take-up	(2) Prevention Take-up
High-Autonomy	0.307**	0.293**
	(0.122)	(0.124)
Add. Receipts Required	-0.107*	-0.092
	(0.056)	(0.057)
Efficiency Pre-Test		0.008
		(0.007)
Quality of Transcription		7.421**
		(3.027)
N. Observations	431	431
Daytime FE	No	Yes
Pseudo-R^2	0.017	0.052

Notes: Results are corresponding to those reported in Table A1, but estimations exclude the 45 subjects who accessed the survey more than once. Heteroscedasticity robust standard errors are in parentheses. The dependent variable is "Prevention Take-up", a dummy variable that equals one if the worker undertakes prevention and zero else. "High-Autonomy" is a dummy indicating that the worker was randomly assigned to the high-autonomy group. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention. "Efficiency Pre-Test" is the time (in seconds) a worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. Daytime fixed effects are categorized in 6 hour increments. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

A.4 Analysis of HIT duration and efficiency

Below are the results of the effect of being in the high-autonomy group on our secondary outcome variables duration and efficiency. Duration is the time in minutes the worker needed to complete all parts of our experiment.

Table A5: Results for duration

	(1)	(2)
Dependent Variable	Duration	Duration
High-Autonomy	1.706	-1.839
	(3.873)	(3.037)
Prevention Take-up	3.025	-0.433
	(3.930)	(3.187)
Add. Receipts Required	0.914	0.910
	(1.916)	(1.492)
Efficiency Pre-Test		2.411***
		(0.218)
Quality of Transcription		4.484
		(20.556)
N. Observations	476	476
Daytime FE	No	Yes
R^2	0.002	0.391

Notes: Estimations are OLS. Heteroscedasticity robust standard errors according to procedure hc-3 are in parentheses. The dependent variable is "Duration (in mins)", the time (in minutes) the worker needed to complete all parts of our experiment. "High-Autonomy" is a dummy indicating that the worker was randomly assigned to the high-autonomy group. "Prevention Take-up" is a dummy variable that equals one if the worker undertakes prevention and zero else. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention. "Efficiency Pre-Test" is the time (in seconds) a worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. Daytime fixed effects are categorized in 6 hour increments. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

Efficiency for one set of transcriptions is the time (in seconds) the worker needed for the transcription of the set divided by the total number of characters contained in the receipts that the worker chose to transcribe in the set.

Table A6: Results for efficiency

Dependent Variable	(1) Efficiency Set 1	(2) Efficiency Set 1	(3) Efficiency Set 2	(4) Efficiency Set 2
High-Autonomy	0.206	0.193	-0.109	-0.128
	(0.133)	(0.132)	(0.155)	(0.173)
Prevention Take-up	-0.077	-0.074	-0.411***	-0.307***
	(0.138)	(0.141)	(0.143)	(0.089)
Add. Receipts Required	-0.060	-0.052	0.005	-0.012
	(0.047)	(0.047)	(0.052)	(0.045)
Efficiency Pre-Test		0.018**		0.013**
		(0.008)		(0.006)
Quality of Transcription		-1.823		-8.523
		(1.354)		(8.023)
N. Observations	475	475	476	476
Daytime FE	No	Yes	No	Yes
R^2	0.007	0.046	0.014	0.150

Notes: Estimations are OLS. Heteroscedasticity robust standard errors according to procedure hc-3 are in parentheses. In columns one and two, the dependent variable is "Efficiency Set 1", the time (in seconds) the worker needed for the transcription of set 1 divided by the total number of characters contained in the receipts that the worker chose to transcribe in set 1. In columns three and four, the dependent variable is "Efficiency Set 2", the time (in seconds) the worker needed for the transcription of set 2 divided by the total number of characters contained in the receipts that the worker chose to transcribe in set 2. "High-Autonomy" is a dummy indicating that the worker was randomly assigned to the high-autonomy group. "Prevention Take-up" is a dummy variable that equals one if the worker undertakes prevention and zero else. "Add. Receipts Required" is the randomly determined number of additionally required receipts if the worker choose prevention. "Efficiency Pre-Test" is the time (in seconds) a worker needed for the pre-test divided by the total number of characters to be transcribed in the pre-test. "Quality of Transcription" is the percentage of similarity of the transcription provided by the worker and the solution measured by Levenshtein distance and averaged across all the receipts that the worker chose to transcribe. Daytime fixed effects are categorized in 6 hour increments. Note that one worker in the high-autonomy group chose to skip all receipts in set 1 and transcribe all ten receipts in set 2. Therefore, only 475 observations are used in columns one and two. *, ** and *** denote statistical significance at the 10%, 5% and 1% level.

A.5 Calibrating utility curvature from prevention choices

The maximum acceptable number of additional receipts allows us to answer the question, whether or not subjects in our experiment have non-linear utility over the number of transcribed lines. For this, we use the same parametric form for the function c as in the integration calibration of Section 5.

In the notation of Section 3, the maximum acceptable number of additional receipts for prevention from the short collection in the low-autonomy group, denoted $\bar{\pi}^L$, solves the equality $EU_P^L = EU_R^L$. We denote the number of receipts to be transcribed per set as a and then solve

$$U\big(c(a+\bar{\pi}^L, s) + c(a,s)\big) = \frac{1}{2}U\big(2c(a,s)\big) + \frac{1}{2}U\big(2c(a,l)\big). \tag{6}$$

We define Γ as the amount of additional lines to be transcribed in the long collection $\Gamma := 2\big(c(a,l) - c(a,s)\big)$. This and the functional form of the effort function lets us use a second-order Taylor approximation around $2c(a,s)$ to obtain

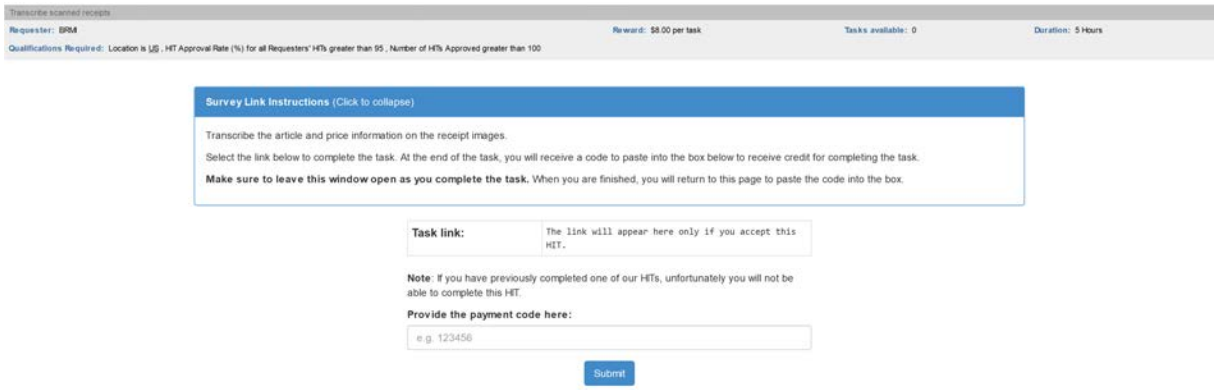
$$\frac{U''\big(2c(a,s)\big)}{U'\big(2c(a,s)\big)} = \frac{8a\bar{\pi}^L + 4(\bar{\pi}^L)^2 - 2\Gamma}{\Gamma^2}. \tag{7}$$

From this equation, we can see that $\frac{U''\left(2c(a,s)\right)}{U'\left(2c(a,s)\right)} = \frac{8*5*2.98+4*2.98^2-2*70}{70^2} = \frac{14{,}72}{4900} = 0.003$. Because of $U' < 0$ a positive value of $\frac{U''\left(2c(a,s)\right)}{U'\left(2c(a,s)\right)}$ implies concave utility over the number of lines to be transcribed.

B Experimental Instructions

B.1 Amazon MTurk

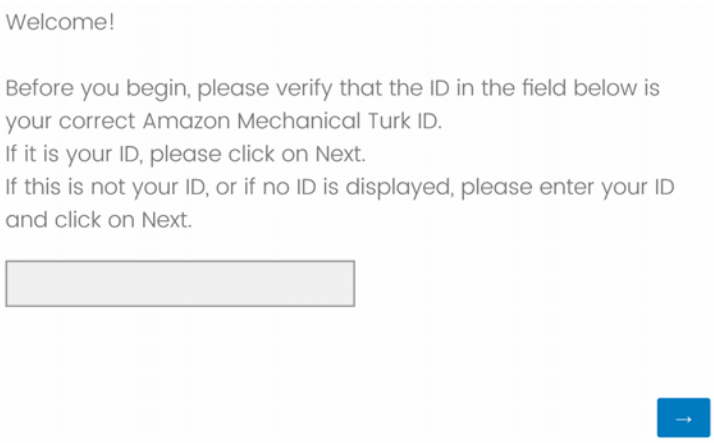
Figure B1: MTurk HIT appearance



Notes: The Figure shows the appearance of the HIT on Amazon MTurk.

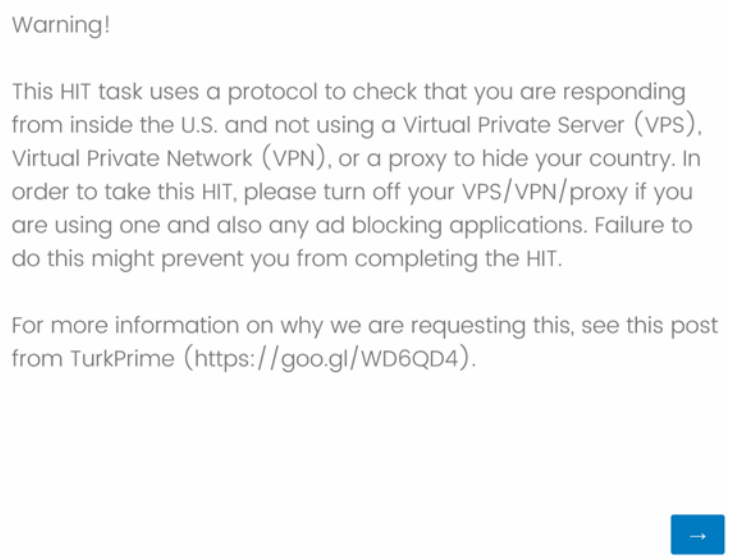
B.2 Qualtrics Survey

Figure B2: MTurk ID verification



Notes: The text field was prefilled with the ID retrieved from the URL that the worker used to access the Qualtrics survey. If the entered ID was in the list of workers who had already participated in one of our test-runs the worker was excluded from participation (see Figure B12). When the worker clicked the next-button it was also automatically checked what type of device the worker was using. If a mobile device was detected the worker was prevented from further participation and was asked to use a computer instead (see Figure B13).

Figure B3: VPN detection warning



Notes: Once the worker clicked next on this screen the IPHub API was used to detect VPN/VPS/Proxy usage. If such usage was detected the worker was excluded from the experiment and the corresponding screen was displayed (see Figure B14). Additionally, the API was used to verify that the worker's IP address was located in the United States. If this verification failed the worker was excluded from the experiment and the corresponding screen was displayed (see Figure B15).

Figure B4: Welcome

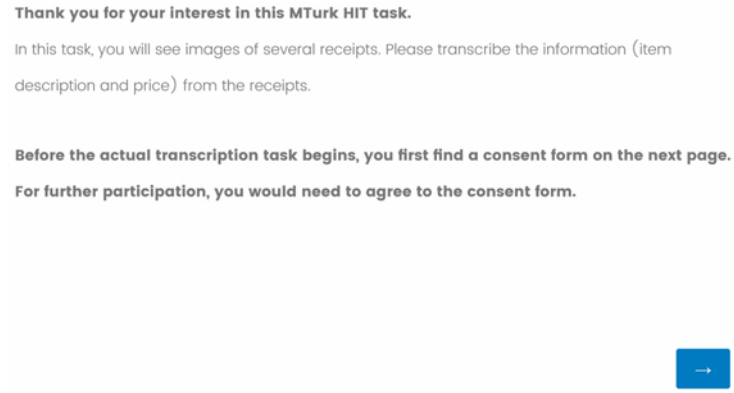
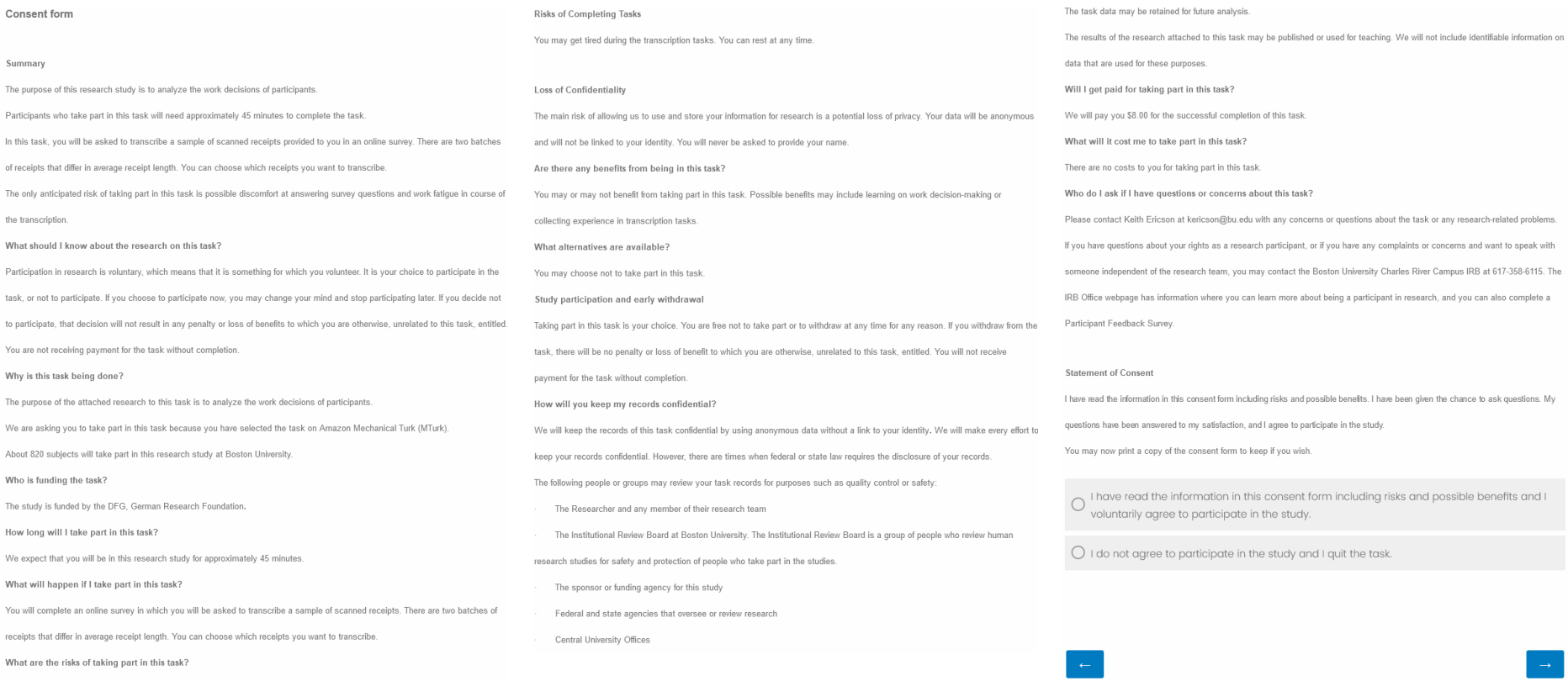


Figure B5: Consent form



Consent form

Summary

The purpose of this research study is to analyze the work decisions of participants.

Participants who take part in this task will need approximately 45 minutes to complete the task.

In this task, you will be asked to transcribe a sample of scanned receipts provided to you in an online survey. There are two batches of receipts that differ in average receipt length. You can choose which receipts you want to transcribe.

The only anticipated risk of taking part in this task is possible discomfort at answering survey questions and work fatigue in course of the transcription.

What should I know about the research on this task?

Participation in research is voluntary, which means that it is something for which you volunteer. It is your choice to participate in the task, or not to participate. If you choose to participate now, you may change your mind and stop participating later. If you decide not to participate, that decision will not result in any penalty or loss of benefits to which you are otherwise, unrelated to this task, entitled. You are not receiving payment for the task without completion.

Why is this task being done?

The purpose of the attached research to this task is to analyze the work decisions of participants.

We are asking you to take part in this task because you have selected the task on Amazon Mechanical Turk (MTurk).

About 820 subjects will take part in this research study at Boston University.

Who is funding the task?

The study is funded by the DFG, German Research Foundation.

How long will I take part in this task?

We expect that you will be in this research study for approximately 45 minutes.

What will happen if I take part in this task?

You will complete an online survey in which you will be asked to transcribe a sample of scanned receipts. There are two batches of receipts that differ in average receipt length. You can choose which receipts you want to transcribe.

What are the risks of taking part in this task?

Risks of Completing Tasks

You may get tired during the transcription tasks. You can rest at any time.

Loss of Confidentiality

The main risk of allowing us to use and store your information for research is a potential loss of privacy. Your data will be anonymous and will not be linked to your identity. You will never be asked to provide your name.

Are there any benefits from being in this task?

You may or may not benefit from taking part in this task. Possible benefits may include learning on work decision-making or collecting experience in transcription tasks.

What alternatives are available?

You may choose not to take part in this task.

Study participation and early withdrawal

Taking part in this task is your choice. You are free not to take part or to withdraw at any time for any reason. If you withdraw from the task, there will be no penalty or loss of benefit to which you are otherwise, unrelated to this task, entitled. You will not receive payment for the task without completion.

How will you keep my records confidential?

We will keep the records of this task confidential by using anonymous data without a link to your identity. We will make every effort to keep your records confidential. However, there are times when federal or state law requires the disclosure of your records.

The following people or groups may review your task records for purposes such as quality control or safety:

- The Researcher and any member of their research team
- The Institutional Review Board at Boston University. The Institutional Review Board is a group of people who review human research studies for safety and protection of people who take part in the studies.
- The sponsor or funding agency for this study
- Federal and state agencies that oversee or review research
- Central University Offices

The task data may be retained for future analysis.

The results of the research attached to this task may be published or used for teaching. We will not include identifiable information on data that are used for these purposes.

Will I get paid for taking part in this task?

We will pay you $8.00 for the successful completion of this task.

What will it cost me to take part in this task?

There are no costs to you for taking part in this task.

Who do I ask if I have questions or concerns about this task?

Please contact Keith Ericson at kericson@bu.edu with any concerns or questions about the task or any research-related problems.

If you have questions about your rights as a research participant, or if you have any complaints or concerns and want to speak with someone independent of the research team, you may contact the Boston University Charles River Campus IRB at 617-358-6115. The IRB Office webpage has information where you can learn more about being a participant in research, and you can also complete a Participant Feedback Survey.

Statement of Consent

I have read the information in this consent form including risks and possible benefits. I have been given the chance to ask questions. My questions have been answered to my satisfaction, and I agree to participate in the study.

You may now print a copy of the consent form to keep if you wish.

○ I have read the information in this consent form including risks and possible benefits and I voluntarily agree to participate in the study.

○ I do not agree to participate in the study and I quit the task.

←

→

Notes: The consent form was presented on one page with text flowing vertically. It is presented here in three horizontally arranged fragments only for readability.

Figure B6: Pre-test (1 of 2)

Pre-Test

Below you see two receipts as a pre-test. Transcribe the item description and price of each purchased article on the shorter scanned receipt. You don't need to transcribe any other information (such as "Zu zahlen xx.xx") from the receipt.

Separate item description and price with a comma (",").

Begin a new line per article and type a semicolon (";") at the end of each line.

In the following table, you see one exemplary receipt and the corresponding required transcription of it.

Receipt	Transcription
Dorfladen *Werner-von-Siemens-Straße 2* *83301 Traunreut* EUR Pistazie 3.05 B Melone 3.80 B Bohnen 1.95 B **Zu zahlen 8.80**	Pistazie,3.05; Melone,3.80; Bohnen,1.95;

In the main transcription task, you will have the option to skip the transcription of some receipts. If you decide to skip the transcription of a receipt, please indicate your choice by checking "Skip". If you decide to transcribe a receipt, please indicate your choice by checking "Transcribe" and enter your transcription in the corresponding text field. To familiarize you with this, please skip the longer of the two receipts below and provide the transcription for the shorter one.

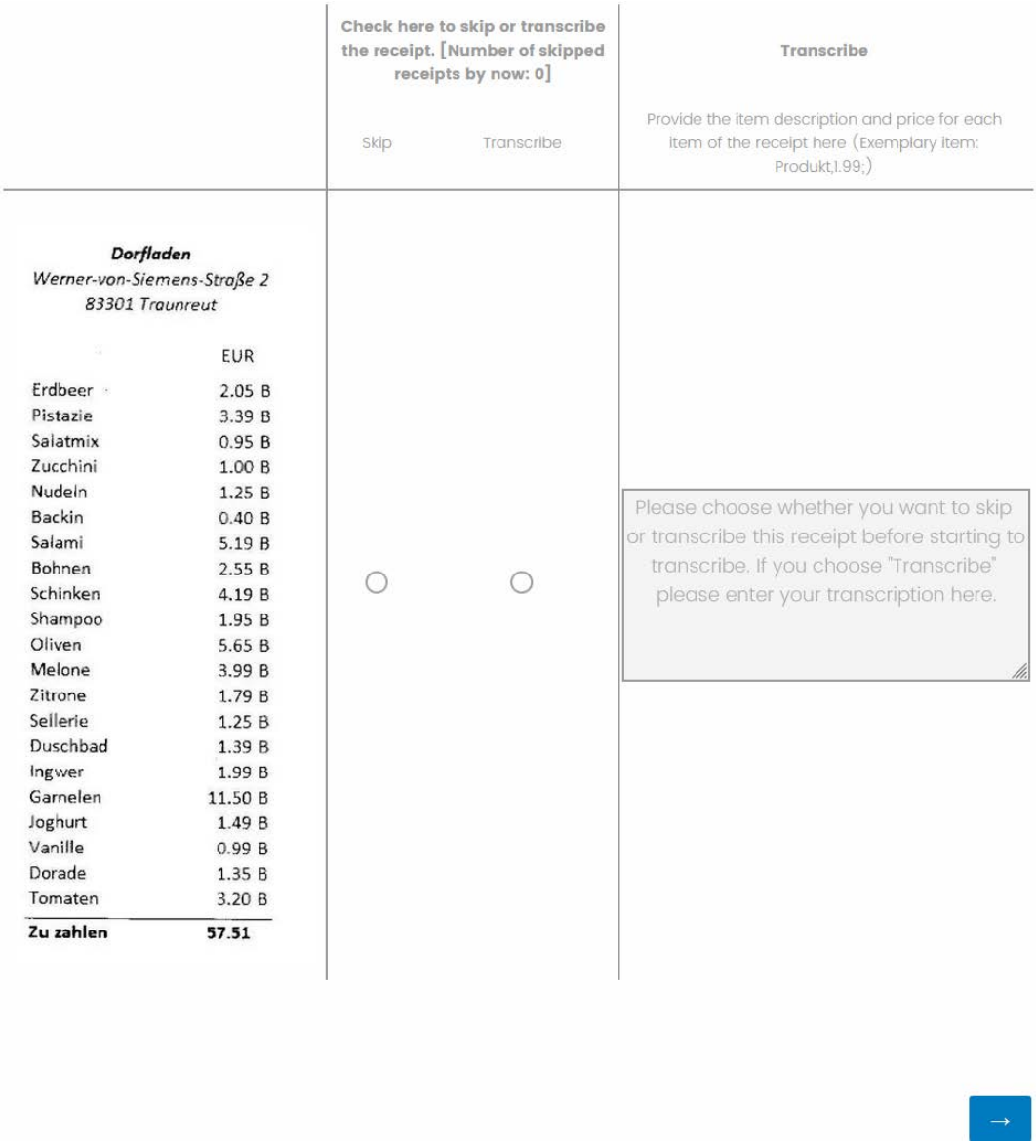
After the completion of the pre-test transcription, you are allowed for the main transcription task.

Your entries will be validated automatically and you are immediately admitted.

	Check here to skip or transcribe the receipt. [Number of skipped receipts by now: 0]		**Transcribe**
	Skip	Transcribe	Provide the item description and price for each item of the receipt here (Exemplary item: Produkt,1.99;)
Dorfladen *Werner-von-Siemens-Straße 2* *83301 Traunreut* EUR Blumen 4.85 B Joghurt 1.29 B Melone 3.69 B Erdbeer 2.35 B Backin 0.39 B **Zu zahlen 12.57**	○	○	Please choose whether you want to skip or transcribe this receipt before starting to transcribe. If you choose "Transcribe" please enter your transcription here.

Notes: The pre-test was presented on one page. For better readability, it is divided into two parts here. The second part follows on the next page.

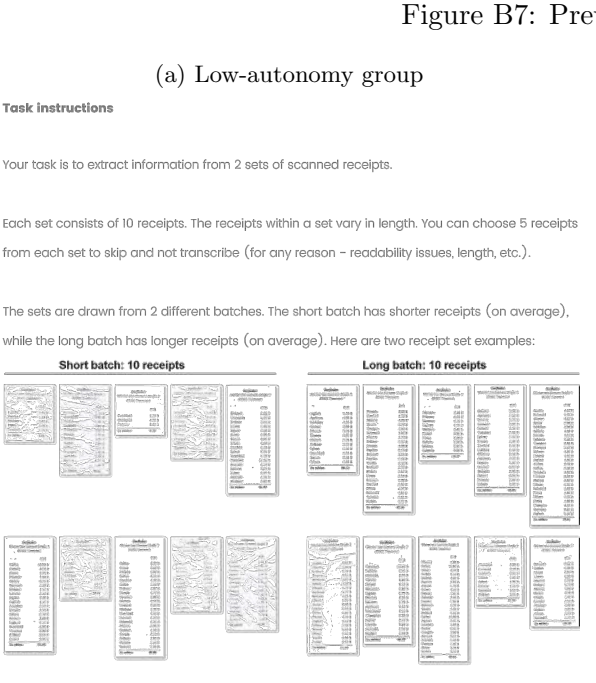
Figure B6: Pre-test (2 of 2)



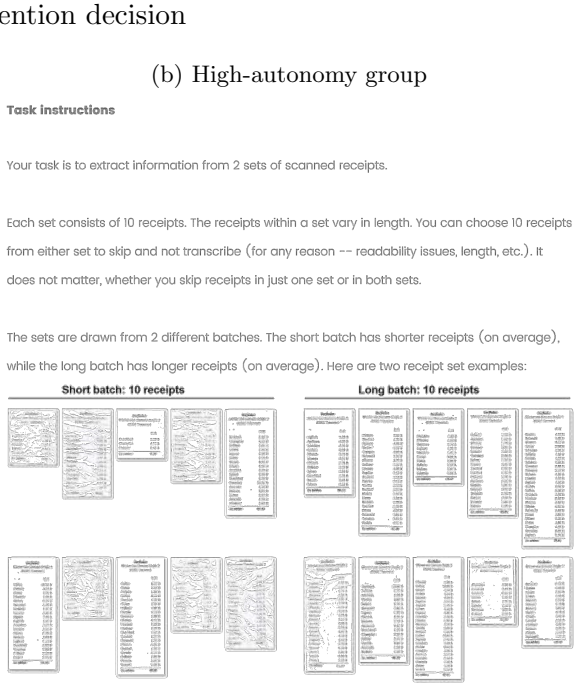
Notes: The pre-test was presented on one page. For better readability, it is divided into two parts here. The first part is on the previous page.

Figure B7: Prevention decision

(a) Low-autonomy group



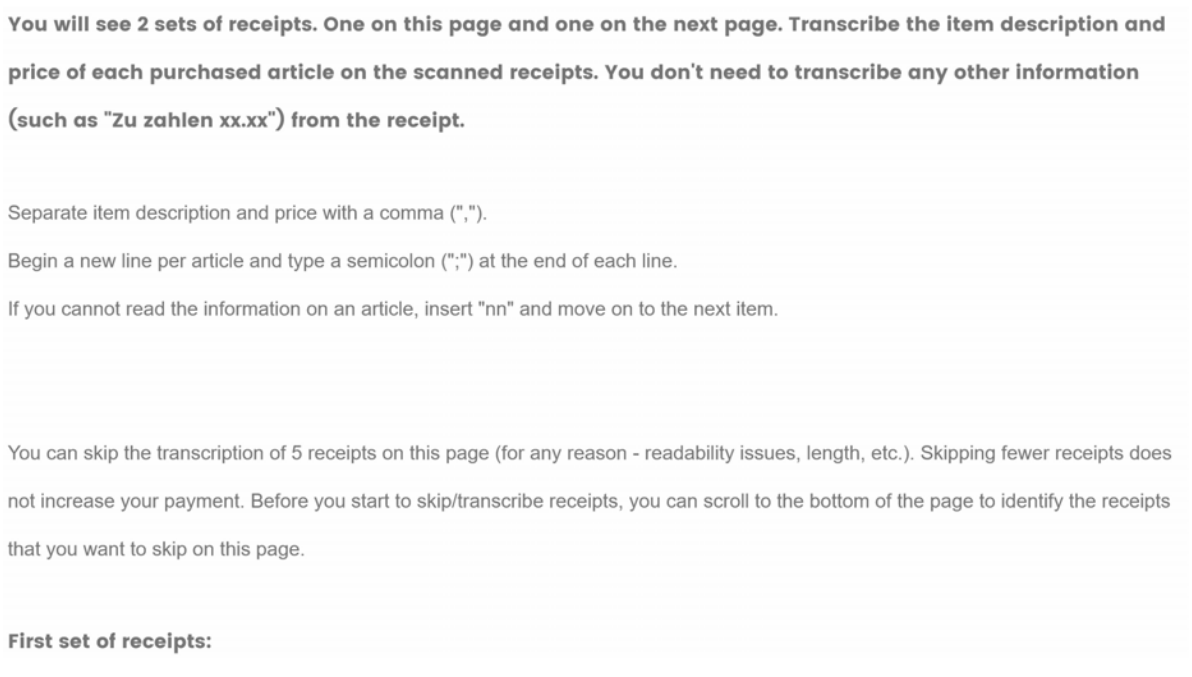
(b) High-autonomy group



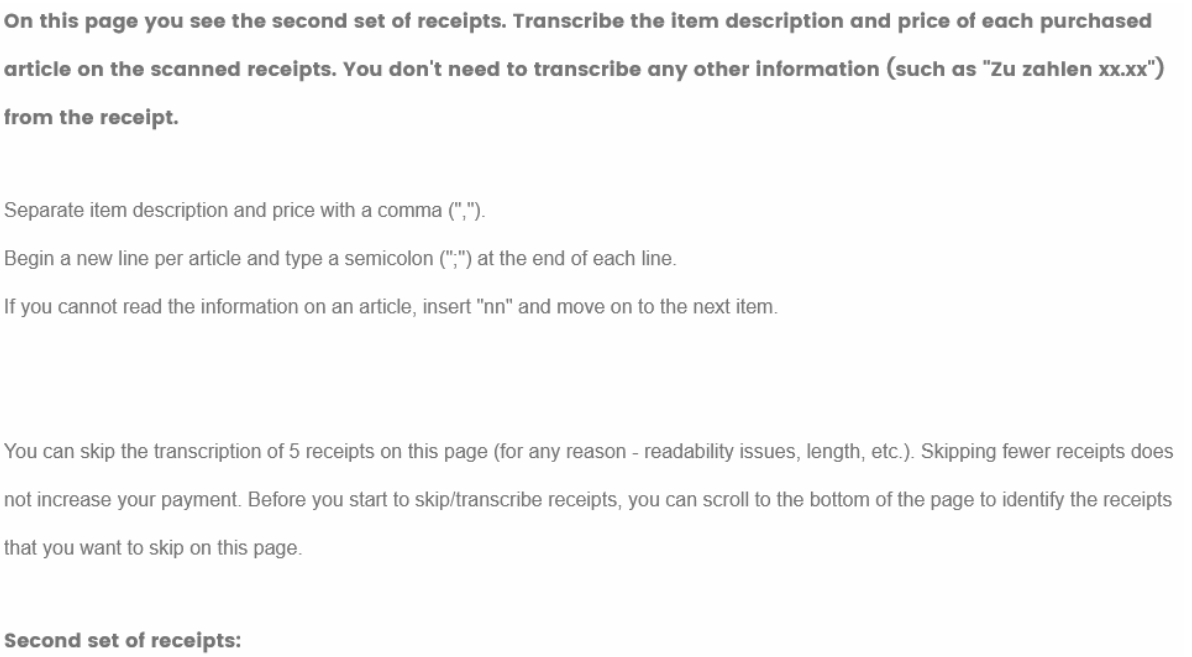
Notes: Panel (a) contains the instructions for the prevention decision shown to workers in the low-autonomy group. In this example, they would have to transcribe two additional receipts for prevention. Panel (b) contains the instructions for the prevention decision shown to workers in the high-autonomy group. In this example, they would have to transcribe three additional receipts for prevention.

Figure B8: Providing transcriptions

(a) Set 1



(b) Set 2



Notes: Panels (a) and (b) show the instructions for transcribing set 1 and set 2, respectively, which were shown to a worker in the low-autonomy group who did not take up prevention. On the same page, after these instructions, the ten receipts of the corresponding set were arranged vertically in random order. Figure 1 shows an example of how a single receipt was presented. The instructions for the high-autonomy group were identical, except that workers there were told how many receipts they could skip in total, rather than on the respective page.

Figure B9: Payment code and debriefing

Thank you for your work!

Here is your payment code. Please enter this number on MTurk to receive your payment: 6522430

When you have copied the payment code, please **click on the next button to submit your survey.**

Debriefing statement

On behalf of our academic partners, thank you for participating in the research study.

Study title and investigators

Title: Risk management and effort fungibility

Investigators:

- Andreas Dambaur (Ludwig-Maximilians-Universität München (LMU))
- Keith Ericson (Boston University)
- Johannes Jaspersen (LMU)
- Sandra Zoller (LMU)

What you should know about this study

The purpose of the attached research to this task is to analyze the work decisions of participants. In this study, participants were first asked to make a risk management choice and then work on transcription tasks. There have been two groups with different flexibility in choosing transcription tasks. The study will analyze the work decisions of the different groups.

Right to withdraw data

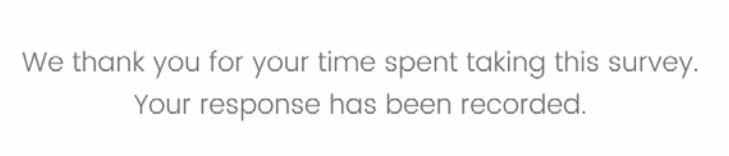
Taking part in this task is your choice. You are free not to take part or to withdraw at any time for any reason. If you withdraw from the task, there will be no penalty or loss of benefit to which you are otherwise, unrelated to this task, entitled. You will not receive payment for the task without completion.

If you have questions

Please contact Keith Ericson at kericson@bu.edu with any concerns or questions about the task or any research-related problems. If you have questions about your rights as a research participant, or if you have any complaints or concerns and want to speak with someone independent of the research team, you may contact the Boston University Charles River Campus IRB at 617-358-6115. The IRB Office webpage has information where you can learn more about being a participant in research, and you can also complete a Participant Feedback Survey.

→

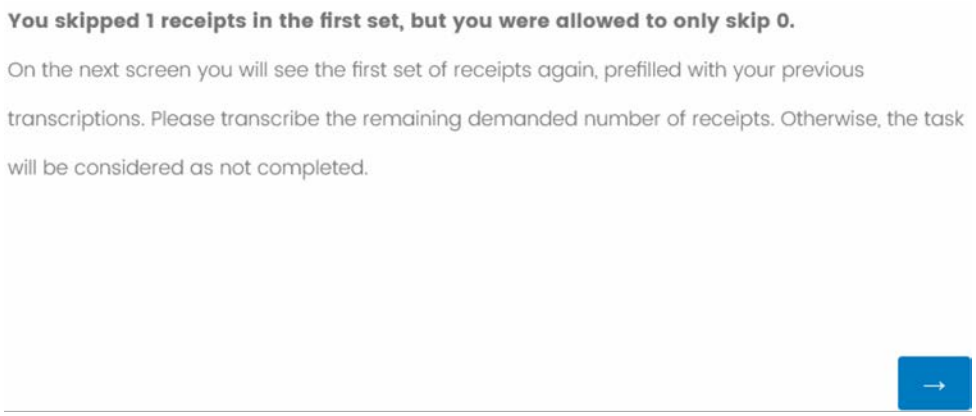
Figure B10: End of survey

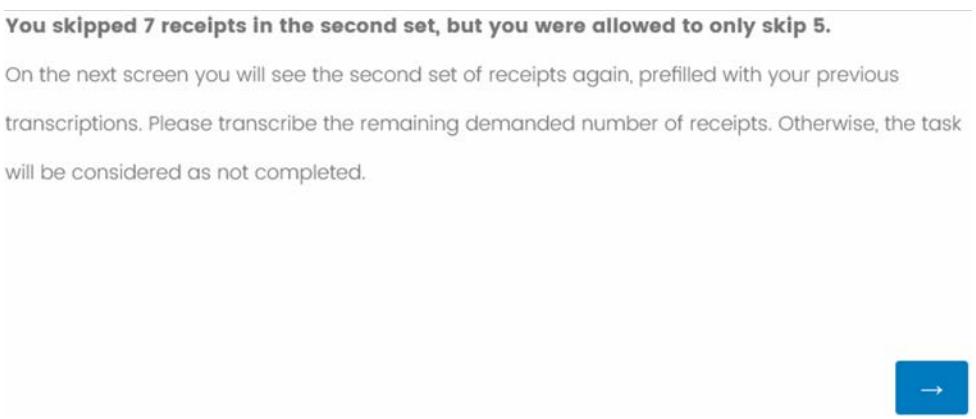


B.3 Qualtrics: Optional Screens

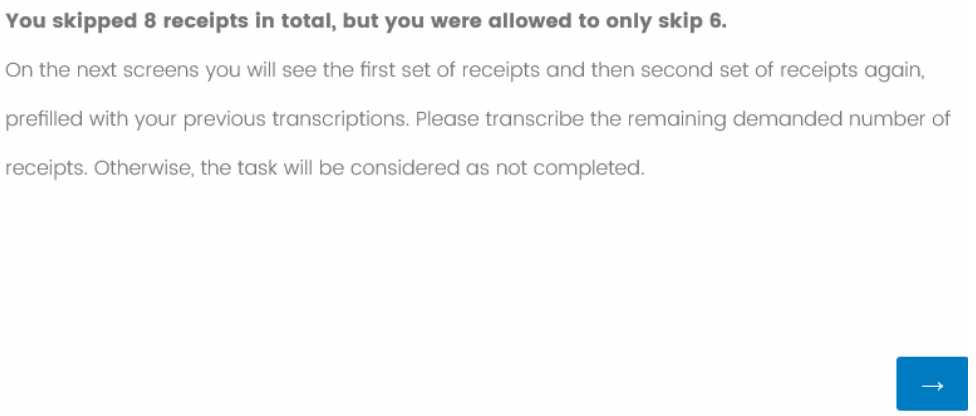
Figure B11: Too many skips

(a) Low-autonomy group





(b) High-autonomy group



Notes: Panel (a) shows the screen that was displayed to workers in the low-autonomy group if the check of the number of skipped receipts in the first/second set, as described in Section 2, revealed that they had skipped too many receipts in the respective set. The worker in this example chose prevention and therefore had to transcribe five additional receipts in the first set, resulting in zero allowed skips. Note that workers in the low-autonomy group were always allowed to skip five receipts in the second set, regardless of their prevention decision. Panel (b) shows the screen that was displayed to workers in the high-autonomy group if the check of the number of skipped receipts in total, as described in Section 2, revealed that they had skipped too many receipts in total. In this example, the worker took up prevention and therefore had to transcribe four additional receipts in total, resulting in six allowed skips.

Figure B12: Participation in test-run detected

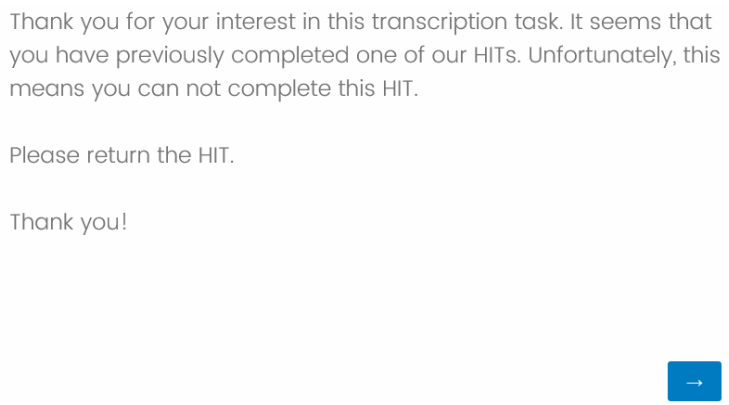


Figure B13: Mobile device detected

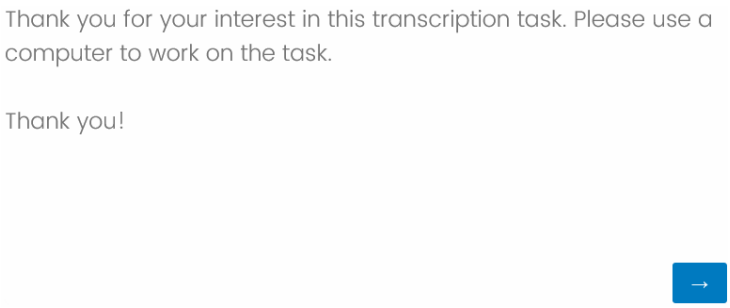


Figure B14: VPS/VPN/Proxy detected

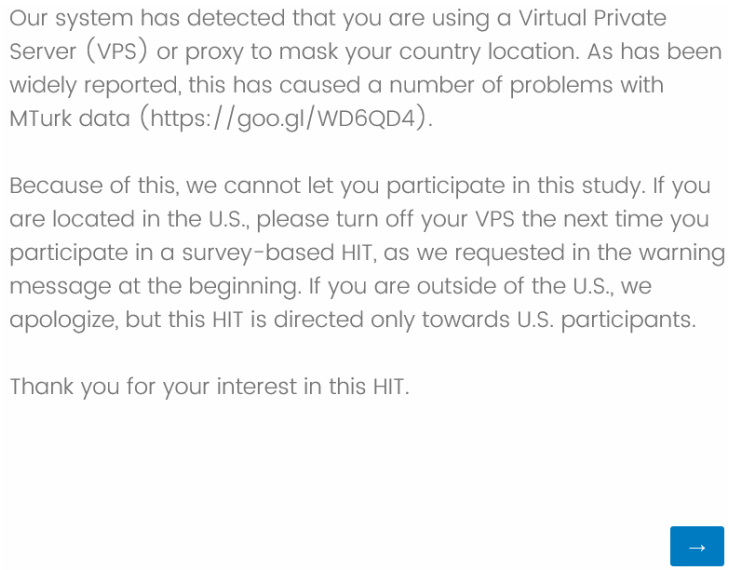


Figure B15: Outside US detected