Asymmetric labor-supply responses to wage-rate changes: experimental evidence from an online labor market*

Philipp Doerrenberg Denvil Duncan Max Löffler

February 2019

Abstract

We test whether or not labor supply responds symmetrically to wage increases and decreases using a randomized field experiment with workers on Amazon's Mechanical Turk. The results show that wage increases have smaller effects on labor supply than wage decreases of equal magnitude, especially on the extensive margin where the elasticity for a wage decrease is twice that for a wage increase. This finding suggests that labor-supply responses to non-marginal wage changes are asymmetric. We discuss the potential mechanisms behind our results including standard models of labor supply, loss aversion and reciprocity.

JEL Classification: J20, J22, J31, D03
Keywords: Labor supply, labor supply elasticities, asymmetric effects of wages, field experiment, Amazon Mechanical Turk

*Doerrenberg: ZEW Mannheim, CESifo and IZA. Email: doerrenberg@zew.de. **Duncan**: Indiana University, Bloomington, ZEW Mannheim and IZA. Email: duncande@indiana.edu. **Löffler**: University of Cologne. Email: max.loeffler@wiso.uni-koeln.de. David Vjazner provided excellent research assistance. We are grateful to Ernst Fehr, Daniel Hamermesh, Johannes Hermle, Rafael Lalive, Harald Lang, Florian Morath, Andreas Peichl, Justin Ross, Jan Schmitz, Sebastian Siegloch, Christopher Zeppenfeld, and various seminar/conference participants for valuable comments and suggestions.

1 Introduction

Because the labor-supply literature typically focuses on *marginal* wage changes, a common prediction of theoretical models is that labor supply responds symmetrically to wage increases and decreases. In other words, wage increases and wage decreases of equal magnitude have the same effect (though with opposite signs) on labor supply decisions, implying that labor-supply elasticities with respect to wages do not depend on the sign of the wage variation. However, it is not difficult to show that *non*-marginal wage changes, which are the more relevant types of wage changes in the real world and therefore for empirical analysis, can lead to asymmetric responses.[1] Although this result has important implications for the empirical estimation of labor-supply responses, there is little empirical evidence regarding the symmetricity of the effect of wages on labor-supply.

This paper contributes to the literature by estimating the symmetricity of labor-supply responses to non-marginal wage changes. Our precise research question is: do wage increases and decreases of equal magnitude have symmetric effects on labor supply? Answering this research question requires a set-up that introduces (quasi-) randomly assigned wage increases and decreases at the same time for comparable individuals. Finding such types of experiments in 'natural' settings is difficult, if not impossible, and thus may partly explain the sparse literature on the symmetricity of labor supply responses to nominal wages.

We address these empirical challenges using a field experiment on labor supply where we randomly assign wage increases and decreases of equal magnitude to workers. Specifically, we set-up a real labor task and invite workers to work on this task in an actual online-labor-market, namely Amazon's Mechanical Turk (henceforth mTurk). The labor task is advertised on the mTurk website as any other labor task and workers receive wages that are comparable to other wages on mTurk. In addition, the labor task is designed to be perceived as realistic as possible; it requires workers to transcribe scanned German-language documents. Importantly, the workers in our experiment do not know that they are participating in an academic experiment. We announce a certain wage per transcribed picture in the advertisement of our task on mTurk and workers complete a batch of six transcriptions for the wage announced in the mTurk advertisement.[2] After transcribing the first batch of images, all workers are randomly assigned to one of three groups: 1) the wage remains constant (control group), 2) the wage increases by 20%, 3) the wage decreases by 20%. After the updated wages have been presented to workers, they

[1]We show below that several models predict asymmetric responses to non-marginal wage increases and decreases; in particular, a standard labor-supply and a model of loss aversion would predict asymmetry.

[2]As a result of this design feature, we induce an exogenously determined expectation regarding the per-unit wage throughout our labor task; workers expect the wage to stay constant at the wage which is advertised on the mTurk website and paid for the first six transcriptions. The field experiment therefore allows us to study the labor-supply responses to unanticipated wage changes.

can select to either stop working on our labor task or keep working as much as they wish. We identify the symmetricity of the labor-supply response by comparing labor-supply behavior between the three randomly assigned groups.

The results show that wage increases have a positive effect on labor supply while wage decreases reduce labor supply, providing clear support for a positive relationship between labor supply and wages. However, the labor-supply response to wage increases and decreases is asymmetric. This asymmetry is especially strong on the extensive margin, defined as the share of workers who quit our task conditional on seeing the treatment information. The estimated extensive-margin treatment effect for workers who experience a wage decrease is approximately twice that of workers who experienced a wage increase. Estimates of the intensive margin response are also suggestive of an asymmetric response where increases have smaller effects than decreases; differences in intensive-margin responses to wage increases and decreases are large, but imprecisely estimated. Our results further show that the wage changes did not have any effect on the quality of transcriptions, which is above 96 percent in all groups.

We discuss several mechanisms that help to rationalize our results regarding the asymmetry of labor-supply effects of wages. First, our results are consistent with standard labor-supply preferences. For example, there might be a positive number of workers in our task with a reservation wage that is between the initially announced wage and the wage in the decrease group; workers in this part of the distribution of reservation wages would quit the task once they learn about the wage decrease. Concurrently, the labor-supply curve might have a particular shape which induces asymmetric responses to *non*-marginal wage-rate changes. Second, the empirical findings are also consistent with a model of loss aversion where the reference wage is equal to the expected wage of $0.15 and the labor supply curve is kinked at this reference wage. In such a model, wage decreases are predicted to have a larger labor-supply effect than a wage increase of equal magnitude. Third, previous research by Kube et al. (2013) shows that asymmetric labor-supply responses can be explained by reciprocity (see below). We argue that this is an unlikely explanation for our results; since workers are paid per-unit wages, shirking (as a punishment for wage decreases) or supplying extra hard effort (as a reward to wage increases) is not possible in our context. We also rule out treatment induced skill-composition changes across the three groups due to skill-based exits as a possible explanation for our results. In particular, we do not observe that unskilled workers are more likely to exit in the wage-decrease group than unskilled workers in the other groups.

Our paper contributes to the literature on labor-supply effects of wage changes. Economists have explored the effect of wages on labor supply for several decades (see Keane, 2011, for a survey). Many of these studies use panel-data sets and exploit positive and negative variation in wages to estimate the wage elasticity of labor supply.[3] Because the elasticity estimated by these studies represents roughly an average of wage-increase-induced and wage-decrease-induced elasticities, our results suggest that existing estimates likely overestimate the effect of wage increases while underestimating the effect of wage decreases. Relatedly, our results further raise questions about the comparability of labor-supply elasticities across studies that differ in the sign of the wage changes used for identification. Our findings suggest that it cannot be concluded from the estimated elasticities that workers are more responsive in the one setting relative to another without knowing whether the sign of the wage changes is the same.[4]

Our finding that the largest asymmetry is along the extensive margin is especially important for understanding the labor-supply effects of wages since it is generally accepted that labor-supply elasticities are mainly determined by the extensive margin response (Blundell and MaCurdy, 1999; Meghir and Phillips, 2010; Bargain et al., 2014). Our results also highlight one possible reason for the downward rigidity in nominal wages (Kaur, 2017). Among the many explanations for this observed rigidity are the potentially detrimental effects on productivity and labor supply. We find large negative extensive margin responses, which suggest that nominal wage cuts could be damaging for firms. This is one potential reason for the reluctance of firms to reduce wages.

We further add to the experimental literature on the effect of wages on effort and labor supply. These studies provide credible randomized evidence in the absence of (discrete) work-time constraints, something which is difficult to obtain using observational data. Papers based on laboratory experiments provide robust evidence that labor effort and wages are characterized by a positive relationship (see the survey by Charness and Kuhn, 2011), which is consistent with our findings. However, laboratory experiments are subject to the usual concern that they cannot easily be generalized to real-world situations. Field experiments with higher external validity find mixed effects regarding the relationship between wages and effort. While some field experiments find a positive effect of wages on effort/labor supply (Fehr and Goette, 2007; DellaVigna and Pope, 2017), other studies find either no relationship (Hennig-Schmidt et al., 2010), short-run temporary effects which do not make a difference for final work outcomes (Gneezy and

[3]It is sometimes argued that nominal wage cuts are rare and therefore not relevant. While we acknowledge that nominal wage cuts occur less often than increases (see the literature on nominal wage rigidities, e.g., Kaur, 2017), it has been shown that wage cuts do happen; for example during recessions and bankruptcies, and for the self-employed and salary earners (Kahn, 1997). In addition, many studies on labor-supply elasticities use upward and downward variation in tax rates to instrument for wages (e.g., Eissa and Liebman, 1996; Rothstein, 2010). This generates downward variation in wages even in the absence of nominal wage cuts. Our study is also relevant for decreases in real wages, which occur more frequently than nominal wage cuts. Our results suggest that inflation-induced decreases of real wages have larger labor supply effects than previously thought.

[4]This is especially important for meta-analysis studies on labor supply (e.g., Evers et al., 2008). Our findings imply that in such meta-analyses one should carefully distinguish between labor supply estimates based on wage increases and those based on wage decreases.

List, 2006), or (positive) effects for only certain types of workers (Cohn et al., 2015). Our results add to the (ongoing) discussion on the wage-effort relation in field experiments, and provide evidence of a positive relationship between wages and labor effort in online labor markets.

To the best of our knowledge, no (lab or field) experimental study explores the potentially differential effects of wage increases and decreases on labor supply.[5] An exception is Kube et al. (2013), which is the study most closely related to ours. They conduct a field experiment with students working in a library for a given period of time (six hours). They generate an exogenous reference wage by announcing a projected hourly wage to all workers when the job is advertised. Immediately before the task starts, they announce a higher wage to workers in one treatment group and a lower wage to workers in another group. Workers in the control condition receive the initially announced wage. The study finds that the wage cut decreases work effort (i.e., output generated during the given period of time) whereas the wage increase does not have any effect relative to the control group. In line with our findings on transcription accuracy, their study also does not find any effects on quality of work.

While these results are broadly consistent with our findings, our paper differs from theirs in the design of the labor market institution, which has important implications for the interpretation and application of our findings. The institutions differ in that we pay workers for each transcribed picture instead of for a predetermined number of hours, and we allow workers to quit the labor task whenever the choose to do so. Furthermore, our analysis is based on a much larger sample of workers from a real-world labor market. Therefore, our design is representative of labor markets where workers receive piece-rate payment and have tremendous labor supply flexibility, whereas Kube et al. (2013) focus on labor markets where workers are required to work a predetermined number of hours for a fixed hourly wage rate. One advantage of our design is that workers are able to respond on two additional margins that are not included in Kube et al. (2013); the extensive margin and the intensive-*time* margin.[6] As a result, we are able to study asymmetric responses to wage changes on both margins. Additionally, because our workers receive a piece rate, subjects who reduce output earn a lower pay-off and have less scope to punish their employer through shirking. This implies that we do not study 'work morale' and it reduces the likelihood that our findings are driven by reciprocity as in Kube et al. (2013). Therefore, we are able to show that labor supply asymmetry exists even in the absence of a motive to reciprocate. The institutional frame-work of our study – large sample of

[5] Wage cuts are generally understudied in this literature; note that none of the experimental papers referenced above examines wage cuts. Also see footnote 3 regarding the prevalence of wage cuts.

[6] The participants in Kube et al. (2013) work for a pre-specified time period, thus precluding the possibility to study a time response. In our experiment, the participants can choose for how long they work, allowing us to study the intensive-time margin.

workers in their natural labor-market environment – also implies that our findings can be generalized to similarly situated labor markets; large crowd-sourcing labor markets characterized by low wage and high flexibility.[7]

To the extent that our findings can (partly) be explained by a model of loss aversion, our paper further makes a contribution to the behavioral-economics literature on loss aversion following Kahneman and Tversky (1979).[8] There is a large empirical literature showing that individuals indeed have preferences consistent with loss aversion and that individual expectations determine the reference point (e.g., Dunn, 1996; Post et al., 2008; Abeler et al., 2011; Card and Dahl, 2011; Marzilli Ericson and Fuster, 2011; Pope and Schweitzer, 2011), but there is scarce evidence regarding the role of loss aversion for labor-supply responses. We add to this literature in that we provide evidence that individuals may have preferences that are consistent with loss aversion in the context of labor supply. This finding is consistent with Ahrens et al. (2014) who derive the theoretical prediction that labor supply responds asymmetrically to wage rate changes in a framework with reference-dependent utility functions.[9]

The paper is organized as follows. Section 2 describes the real labor task and its implementation in Amazon's Mechanical Turk. Section 3 describes the data and our empirical approach and we present the results in Section 4. We discuss the potential economic mechanisms behind our findings, as well as their implications and generalizability, in Section 5. Section 6 concludes.

2 The Experiment

This section describes the field experiment used to estimate the impact of wage rate changes on labor supply. We begin by describing the labor task, the treatment design and

[7]The findings of two additional papers are relevant in the context of asymmetries in labor markets. Falk et al. (2006) find in a laboratory experiment that reservation wages respond asymmetrically to the introduction and removal of minimum wages. The results show that the introduction of a minimum wage has larger effects than the removal. Chemin and Kurmann (2014) study how reciprocal behavior of 12 fieldworkers responds to wage increases and decreases. Consistent with Kube et al. (2013), they find that wage increases had no effect while wage decreases had a negative effect on effort, and attribute this effect to reciprocity.

[8]This literature pursues the idea that individuals evaluate outcomes relative to reference points. These types of preferences are commonly termed reference-dependent preferences and have been formalized by Koszegi and Rabin (2006, 2007, 2009). Loss aversion describes the notion that individuals weight negative deviations (losses) from the reference point more than gains of equal magnitude.

[9]Our paper also relates to several studies showing that individual labor supply decisions are affected by target incomes. In a survey of the literature, Goette et al. (2004) show how empirical results on labor-supply behavior are consistent with reference-dependent preferences where workers provide high effort if they are below a target income, whereas they provide less effort if they have surpassed a target (also see Camerer et al., 1997, Crawford and Meng, 2011 and Fehr and Goette, 2007). Empirical evidence also suggests that loss aversion affects job searches (DellaVigna et al., 2017). While these studies demonstrate that workers have target incomes and loss-aversion preferences in the context of labor supply, they do not allow conclusions about the asymmetric effects of wages.

the implementation in Amazon's Mechanical Turk.

2.1 Design

Labor Task. We selected an online labor task that requires subjects to transcribe German text shown in a series of images. The German texts are taken from a recent publication, but each page of the document is ruffled so that the scanned versions appear much older than they really are. The advantage of changing the appearance of the images is that subjects are more likely to believe that the texts were scanned from old books for which a digital copy is not available. The task then, is to digitize these "old" German books.[10] Each image has approximately five lines and 43 words (344 characters). Figure 1 shows an example. Subjects are randomly shown one of 128 images at a time and are instructed to hit "save picture" when they are done transcribing the text in the image. A new image is shown after the subject hits "save picture".

Treatment Groups. We use a between-subjects design in order to identify the effect of wage changes on labor supply. Subjects are randomly assigned to one of three groups: one control group and two treatment groups. Subjects in all three groups work on the labor task described above and are paid a piece rate for each image that is transcribed. The piece rate[11] is set at \$0.15 for each of the first six transcribed images in all three groups. Subjects receive a notification thanking them for transcribing the images after the first six images have been transcribed. They are then told that they can transcribe additional images and that the piece rate for the additional images is either \$0.18, \$0.15 or \$0.12, for the wage-increase, control, and wage-decrease groups, respectively (see Figure 4 for an explanatory treatment notification). Notice that the wage rate remains fixed at \$0.15 for the control group, and that the wage rate change is the same for both treatment groups; in each case the rate changes by \$0.03 or 20%. We did not provide workers with a reason for their wage changes in order to keep a neutral framing (Kube et al., 2013). In addition, the reasons for the wage changes would have had to be different for wage increases and decreases, which would have complicated the comparability between the treatment groups.

Wage Expectations. The experiment is designed to establish an exogenous and salient expectation regarding the per-unit wage in the our mTurk task. Potential workers are told that the wage per transcribed picture is \$0.15 in the job announcement. Additionally,

[10] Horton et al. (2011) use a similar task and motivate it with the following advantages: transcribing text (i) is tedious, (ii) requires effort and attention and (iii) has a clearly defined quality measure.

[11] The piece rate is called *bonus* in the experiment. This is the usual wording if one is to implement per-piece payment within the same task in the mTurk labor market.

workers who start working on our task face the announced wage of \$0.15 for the first six transcribed pictures, after which the wage rate either increases or decreases. We argue that this design generates the expectation that the per-unit wage will remain constant at \$0.15 throughout the entire task. Our experimental design therefore allows us to study how unexpected wage increases and wage decreases affect labor supply. If we had initially told subjects that the wage would either increase or decrease, they could have adjusted their expectations and the labor supply response to varying wages would not have been comparable to real-situations where workers experience unanticipated wage changes. This design feature is also in accordance with Kube et al. (2013) who, following Bewley (2005), argue that deviations from an exogenous expectation capture the key aspects of wage changes (for example, disappointment and the break of trust relation in the case of wage cuts).

One potential drawback of our experimental design is that it may raise concerns of deception since the job description does not notify subjects of the possibility that the wage may increase or decrease after a certain number of transcribed pictures. This was a deliberate choice in an effort to establish a clear and salient wage expectation.[12] We avoid deception by including the following pieces of information in the treatment notification (see Figure 4). First, we thank the workers for completing the transcription task and remind them that, as promised in the introduction of the task, they will be paid \$0.15 for each of the pictures they transcribed so far. Next, we inform them that they have the option to transcribe additional images and that the piece rate for these additional transcriptions is different from that for the first batch of transcriptions. Finally, we make it clear that they can stop and exit the task at this point if they wish and instruct them on what to do next to ensure we are able to process their payment.[13] We argue that these design features make it clear to workers that they first transcribe pictures based on the piece rate announced in the introduction to the task, and that they can transcribe additional pictures at a new rate. The design of the task gives the impression to workers that the task consists of two parts and ensures that we did not deceive the workers regarding the wage in the second task.

[12]Informing subjects about the possibility of a wage change would have generated uncertainty about the eventual wage and the wage expectation would not have been as clear.

[13]The notification reads: "Thank you for transcribing these pictures. As written in the introduction, we will grant a bonus of \$0.15 for each of these pictures. There are additional pictures that you can transcribe. However, the bonus payment for each additional picture will be \$0.12/\$0.18 from now on. You will receive \$0.15 bonus for each of the six pictures you transcribed so far, though. If you want to stop and exit, just copy your Personal ID to the Amazon Turk Website and submit the HIT." Instead of the wage change, we include the following message for the control group: "There are additional pictures that you can transcribe. Just as before, the bonus for each additional picture will be \$0.15."

2.2 Implementation

Labor Market and Recruitment. The experiment is implemented in the field using workers on Amazon's Mechanical Turk. mTurk is an online labor market where job offers are posted and workers choose jobs for payment. It has numerous benefits for running experiments, including access to a large stable subject pool, diverse subject background, and low cost.[14] Furthermore, the behavior of online workers has been shown to be comparable to those of subjects in laboratory studies (Horton et al., 2011). Additionally, experimenter effects are avoided because subjects do not know that they participate in an experiment (Paolacci et al., 2010; Horton et al., 2011; Buhrmester et al., 2011; Mason and Suri, 2011). Importantly for us, we are able to identify the effect of wages changes in a naturally occurring labor market. In general, experiments on Amazon's Mechanical Turk therefore combine internal and external validity since it is a real labor market with actual workers where randomized trials can be conducted (Horton et al., 2011).[15]

Although we recruit workers through mTurk, they complete the labor task on an external website that we created for the purposes of the experiment. We first create a human intelligence task (HIT) that is advertised on mTurk. The HIT includes a description of the labor task and compensation. It also includes instructions for how to complete the task; see Figure 2. Particularly, subjects are told to accept the HIT and click on the weblink if they are interested in completing the task. Subjects who click on the link are taken to our external website where they are randomly assigned to one of three groups and shown the instructions in Figure 3. Subjects are instructed to click continue if they wish to work on the task, and those who do are shown images of scanned German text that they must transcribe for payment. Each page of our website shows the subjects their personal ID, number of pictures transcribed so far, and the current piece rate. We implement treatment after six images have been transcribed and limit the total number of images that each subject can transcribe to 50. However, subjects are not aware of either of these limits until they reach them. In other words, subjects do not know that the HIT has six images, that they will have the opportunity to continue working after the first six images, that the piece rate might be different if they continue working, or that they can only transcribe up to 50 images if they chose to continue working. Subjects in wage-decrease group who complete six transcriptions are shown the treatment information illustrated in Figure 4. A similar text is shown to subjects in the wage-increase group and the control group; the only difference is the piece rate for the additional images.

Transcribing text from an image can be a tedious task. However, given that the text

[14] According to Amazon, there are over 500,000 workers from 190 countries in the mTurk labor market: `https://requester.mturk.com/tour`.

[15] Kuziemko et al. (2015) and DellaVigna and Pope (2017) are recent examples of economics papers using Amazon's Mechanical Turk.

in the images is short, the task could be perceived as mostly costless for German speakers. In order to reduce this possibility and ensure that the labor costs are non-zero, we restrict the subject pool to workers with a US IP address. The idea here is that the labor cost of transcribing German text is much higher for non-Germans than for Germans. Of course, our restriction does not preclude the possibility that German speakers participated in the task. However, any Germans who participated in our experiment are randomly distributed across our treatments and therefore have no effect on our outcome of interest.

The experiment is programmed on mTurk to expire after 750 workers accept the HIT or 10 days have passed, which ever comes first. Our initial run of the experiment, which started on June 15, 2015, expired after 10 days with only 418 workers. Therefore, we initiated a second run on July 20, 2015, and this run expired after hitting the 750 worker threshold six days later. In total, 1,168 workers participated in the two runs. Note that the HIT is designed such that workers cannot work on the task more than once. We also excluded workers who participated in the first run from participating in the second run. Moreover, it is highly unlikely that individuals have multiple worker accounts to avoid these constraints: First, when registering for mTurk, Amazon requires workers to confirm in the Participation Agreement that they "may not use multiple Amazon Accounts to register with Mechanical Turk". Second, the Participation Agreement further requires workers to provide "true and accurate" information on a worker's name, email address, phone number and physical address.[16] Third, workers are required to provide a tax identification number (Social Security Number or Individual Tax Identification Number) after their mTurk lifetime earnings have exceeded a set threshold. Workers who fail to provide this number are not allowed to accept additional HITs on mTurk.

Payment. The experiment ends for each subject when she decides to stop or when she transcribes 50 pictures, whichever comes first. In either case, each subject is instructed to copy her personal ID number, which is shown in the top right corner of each page, and paste it in the entry box on the mTurk website. This process is necessary for us to match subjects to their mTurk worker ID and thus process their payments. Subjects receive a participation reward of \$0.10, which is paid as long as a subject accepts the HIT and completes at least one transcription. Additionally, subjects are paid a piece rate of \$0.15 for each of the first six transcribed pictures, and depending on treatment group, \$0.12, \$0.15 or \$0.18 for each transcribed image above the first six transcriptions. Given the payment restrictions imposed by the mTurk platform, we frame the piece rate as a bonus in all communications to the subjects. For example, subjects in the control group are told they will be paid \$0.10 for participating in our HIT and a bonus of \$0.15 for each transcribed picture.

[16] The Participation Agreement is online at: `https://www.mturk.com/mturk/conditionsofuse`.

We chose this payment structure based on a small test of the real effort task that we implemented with English-speaking students in a class at a major public university in the US before we started the field experiment.[17] This test revealed that it takes about 4 minutes on average to transcribe foreign-language text paragraphs that have the same size as the paragraphs in our experiment. This suggests that approximately 15 pics can be transcribed per hour, resulting in an hourly wage of about \$2.35 ($= 0.1 + 6 \times 0.15 + 9 \times 0.15$) in the control group, \$2.62 ($= 0.1 + 6 \times 0.15 + 9 \times 0.18$) in the increase group, and \$2.08 ($= 0.1 + 6 \times 0.15 + 9 \times 0.12$) in the decrease group. In light of a median reservation wage of between \$1.12 and \$1.38 per hour for mTurkers, according to Horton and Chilton (2010) and Horton et al. (2011), this payment structure seemed adequate from an ex-ante perspective. From an ex-post perspective (see results), it turns out that the average time needed per picture in our small test was an appropriate, if not even too conservative, predictor of the transcription speed in our actual experiment. We observe that participants in our sample (across all groups) needed about 3.7 minutes per picture (i.e., about 16.22 pictures per hour), which results in hourly wages of \$2.53, \$2.84, \$2.23 in the three groups, respectively. That is, both the per-hour wages that we expected before we ran the experiment and the per-hour wages that we observe for the workers in our experimental sample are considerably higher than the hourly median reservation wage reported in Horton and Chilton (2010) and Horton et al. (2011).

3 Data and empirical approach

This section describes our outcome variables, details on the sample, and the empirical strategy used to identify the symmetry of wage effects.

3.1 Outcome Variables

We construct several outcome variables that measure different aspects of labor supply in order to identify the effect of wage changes on labor supply. These include the quit rate, number of transcribed pictures, time spent transcribing, transcription rate, and accuracy. Each of these variables is described in greater detail below.

Transcriptions and Hours. Because workers are paid for each transcribed image, we expect that they will respond to the wage changes by changing the number of images they transcribe. Therefore, one variable of interest is the total number of transcribed images per worker. We further explore the total time spent working on the task and the time per transcribed text (transcription rate). Because we do not have an exact measure of

[17] This test did not include any wage variations. The sole purpose was to test the functionality of the website and to infer the average time it takes to transcribe one of the images.

the time workers actually spent working on a picture, we proxy the transcription rate by counting the time between the submission of two transcriptions. We acknowledge that this likely overstates the transcription time for any given image. However, the difference in transcription rate between groups should still be instructive of the impact of wage changes.

Extensive Margin. Recall that workers are notified of treatment after transcribing six images. The notification makes it clear that the worker has completed the HIT, but that there are additional (optional) images to transcribe. Workers are also informed that they can quit the task at this point or continue transcribing the additional images at the newly announced wage rate. Given these features of the treatment notification, we interpret the decision to stop working at this point as an extensive margin decision. Therefore, one of our key outcome variables is the share of workers who quit the task immediately after receiving the notification. Because the treatment notification has a modest nudge to quit, we expect that the share of quitters will be reasonably high in the control group despite the fact that the wage remains constant. The important question for us is: does the wage increase/decrease have any effect beyond this modest nudge.

An important feature of online-labor markets such as mTurk is that they facilitate almost instantaneous switching of labor tasks. In other words, a worker can quit one job this second and start a new job the next second. This is not unlike what one would observe in traditional labor markets where a worker secures a new job before quitting her existing job. Unfortunately, we do not observe what subjects do when they quit our task. Therefore, the extensive margin response in our study simply means that the worker quits our task. We cannot say whether or not they quit working online or switch to a more profitable task.

Accuracy. Recall that the transcriptions are based on text for which we have the original digital copy. This makes it possible for us to measure accuracy by comparing the transcribed text for each worker to the actual text.

3.2 Sample

Our HIT was accepted by 1,168 mTurk workers. We restrict the sample to those workers who completed at least one picture, and therefore received the participation fee; this leaves us with 1,158 workers. We observe in the data that a few workers worked on the task for an unreasonable number of time, e.g., several days. To avoid this source of noise, we drop the top 0.05% of workers in the distribution of minutes worked; these are six workers who worked for more than 385 minutes on the task. Table 1 presents summary statistics for our sample of workers ($N = 1,152$) with regard to our main variables: number of transcribed pictures, accuracy of transcription, and total time worked. We observe that, on average, workers transcribed 12.8 pictures[18] over an average time span of 39.79 minutes. The transcription quality was very high with an average accuracy of 96.97%. This is reassuring as it suggests that workers take the task seriously and provided high-quality transcriptions. Note that we intended to avoid giving the impression that subjects are participating in an experiment, and therefore did not survey any demographic characteristics.

Because the treatment variation in wages only appears after the first batch of six transcriptions, only a share of the total 1,152 participants are exposed to the treatment condition. Table 2 shows that 62.5% (720) of the 1,152 workers completed at least six pictures and therefore saw the treatment notification. This share ranges from 59% in the wage-increase group to 65% in the wage-decrease group. The number of observations in each treatment group is summarized in Table 2. In total, we have 248, 215, and 257 workers who saw the treatment notification in the control, increase and decrease groups, respectively. Because workers did not know they were in an experiment or that the wage rate would change, self-selection into the treatments was impossible. We therefore argue that the groups are balanced with respect to the characteristics that predict the probability of quitting before seeing the treatment, and thus we mostly restrict the empirical analysis that follows to the sample of 720 participants who saw the treatment (see section 4.1 for data-based evidence that there is no difference between groups before treatment notification).

A common feature of mTurk is that workers discuss HITs on forums. This can raise issues for experimenters as those workers who have completed the experiment will unknowingly share the details of treatments with other workers who have yet to complete the experiment. We followed the forums on mTurk in order to determine if our HIT was being discussed and discovered that our HIT did in fact show up on one of the forums.[19] The first mention of our HIT occurred on July 24 during the second run of the experiment. We noticed the mention on the 26^{th} when the HIT had already expired. The discussion on the forum was favorable towards our HIT, but workers discussed the fact that the wage rate changed as well as the magnitude of the changes. They also discussed potential reasons for rate changes, and mostly speculated that the wage variation must be due the quality of work. Nobody speculated that this task is an experiment; people therefore still did not know they were part of an experiment.

The forum post led to a significant spike in acceptance of our HIT; approximately 58% of the workers accepted the HIT after the forum discussion began. Because some of these subjects knew of a potential wage variation before accepting the HIT, self-selection

[18]Figure 18 in the Appendix provides the distribution of completed pictures for all workers in the sample.

[19]See `https://www.reddit.com/r/HITsWorthTurkingFor/comments/3eg39l/us_transcribe_texts_from_an_image_payment_bonus/`.

might be a problem. For example, it is possible that only workers who are willing to work for our lowest wage rate accepted our HIT. If this is the only source of selection, then our analysis produces a lower bound estimate in both groups. A more troubling source of selection is a case where workers sign up with the hope of receiving a wage increase. These subjects would effectively have the expectation that the wage will be \$0.18, and would be more likely to quit the task if assigned to the wage decrease group. This source of selection would lead to a downward bias in the wage-decrease group and upward bias in the wage-increase group. Because of this potential problem, we present estimates with and without the post-forum sample. There is no evidence that the forum had an effect on the results (see Appendix B).

3.3 Empirical Strategy

Random assignment to treatment groups ensures that our empirical approach is straight forward. The empirical analysis proceeds as follows. First, for each experimental group, we plot the share of subjects in each 'period',[20] relative to the total number of subjects who initially started the working task. We use all available subjects (i.e., not only those who saw the treatment notification) for this exercise. This descriptive analysis sheds light on differential drop out rates across the experimental groups before and after treatment.

Second, for each outcome variable, we compare the means of the respective outcome across treatment groups and use non-parametric Wilcoxon rank-sum tests for differences in distributions between the groups (Wilcoxon, 1945; Mann and Whitney, 1947). In addition, we run simple OLS regressions of the outcome variables on the treatment dummies.

$$Y_i = \alpha + \beta_{Increase}\mathbb{1}(i \in Increase) + \beta_{Decrease}\mathbb{1}(i \in Decrease) + \epsilon_i \tag{1}$$

where Y_i is an outcome of interest for subject i, e.g., the number of transcribed pictures. Indicator functions $\mathbb{1}(i \in Increase)$ and $\mathbb{1}(i \in Decrease)$ evaluate to one if worker i is part of the increase group or the decrease group, respectively, and zero otherwise. α is a constant, ϵ_i denotes the unexplained error term. These empirical analyses allow us to identify the effect of wage increases and decreases on our outcome variables. These parametric and non-parametric analyses are restricted to the subjects who saw the treatment notification (this is sufficient because there is no selection prior to the treatment notification – as discussed before and shown below in section 4.1).

To test for symmetry of these responses, we use the $\widehat{\beta}$ coefficients of the above OLS regressions for each outcome and t-tests to test the null that the sum of the estimated coefficients for the wage-increase group and wage-decrease group (both relative to the

[20]We use the term 'period' to indicate the number of the picture which is to be transcribed. For example, the treatment notification occurred after 6 periods; i.e., after subjects transcribed 6 pictures.

control group) is zero:

$$H_0 : \beta_{Increase} = -\beta_{Decrease}. \tag{2}$$

We then use the estimated regression-based treatment effects to calculate implied elasticities separately for each treatment group. Using the control group as a counterfactual, we derive the elasticity of an outcome variable Y with respect to wages for each treatment group g (either wage increase or decrease) as follows:

$$\epsilon_g = \frac{(Y_g - Y_c)/Y_c}{(w_g - w_c)/w_c} \qquad\qquad \widehat{\epsilon}_g = \frac{\widehat{\beta}_g/\widehat{\alpha}}{(w_g - 0.15)/0.15}, \tag{3}$$

where subscript c indicates the control group, w is the wage per transcribed picture, $(w_g - w_c)$ is the change in wages in group g (either $+3$ or -3), and $(Y_g - Y_c)$ is the difference between the outcome variable in group g and the control group. Specifically, $(Y_g - Y_c)$ is the difference in means between the relevant treated group and the control group or, equivalently, the regression coefficient β_g of the respective treatment dummy.

Finally, to shed more light on the dynamics of the effects, we regress the probability of working in a given period on a full interaction of an increase-group dummy and period dummies as well as a full interaction of a decrease-group dummy and period dummies (standard errors clustered on individual level). Formally, this regression reads:

$$D_{i,t} = \sum_{t=1}^{50} \gamma^t_{Increase} \mathbb{1}(i \in Increase) + \sum_{t=1}^{50} \gamma^t_{Decrease} \mathbb{1}(i \in Decrease) + \nu_t + \epsilon_{i,t} \tag{4}$$

where $D_{i,t}$ equals one if worker i transcribed picture t and zero otherwise. We normalize coefficients of the last pre-treatment period – $\widehat{\gamma}^6_g$ – for both the increase and the decrease group to zero. The coefficients from this regression, which we present in an event-study type Figure, provide insight about the differential dynamics of our treatment effects between the wage-increase and wage-decrease groups (always relative to the control group). Interested in the asymmetry of treatment effects between groups, we then use these regression results to sum up the estimated coefficient for the wage-increase group $\widehat{\gamma}^t_{Increase}$ and the corresponding estimate for the wage-decrease group $\widehat{\gamma}^t_{Decrease}$ for each period t. These results are also shown in an event-study type Figure and provide insights about the dynamics of asymmetric effects.

4 Results

In this section we present the empirical results of our experiment. We start by analyzing workers' drop-out rates by looking at the full sample with all subjects to check that there are no pre-treatment trends in outcomes in Section 4.1. This descriptive exercise also allows to check for first indications of asymmetric responses across the three groups. In a second step, we study the treatment effects at the extensive margin restricting our sample to workers who transcribed at least six pictures and saw the treatment (Section 4.2). Third, we analyze intensive margin responses regarding the time spent working and the transcription rate in Section 4.3, and on the number of transcribed pictures and the quality of transcriptions (see Section 4.4).

4.1 Descriptive Evidence

We begin the analysis by calculating the drop-out rate in each 'period' by control and treatment groups. Using all workers who completed at least one picture, Figure 5 shows that around 7-8 percent of those subjects who started working in our experiment quit the task in each pre-treatment period. Importantly, Figure 5 shows that pre-treatment drop-out rates are equal across groups, suggesting that there is no differential selection across groups *before* treatment. After seeing the treatment, the exit-rate in the increase group decreases to three percent, while the exit rate for the decrease group rises to 21 percent. These differential exit-rates are a first indicator of asymmetric effects between wage increases and decreases.

4.2 Extensive Margin

We now turn to the statistical analysis of our results. Because we find no evidence of self selection into treatment – as illustrated by non-differential pre-treatment trends – all following analyses are based solely on the sample of workers who saw the treatment notification.[21]

Figure 6 displays the treatment effects on the extensive margin, i.e., the share of workers who quit immediately after having seen the treatment. We observe that 14.1% of the workers in the control group quit the labor task after receiving the treatment notification. Relative to the control group, the share of quitters is 8.5 percentage points lower in the wage-increase group and 17.8 percentage points higher in the wage-decrease group. These group-wise differences between means are all statistically significant at the 1% level according to non-parametric ranksum tests, and suggest that wage increases induce workers to keep working while wage decreases increase the likelihood of quitting. These results are also demonstrated in OLS regressions (based on equation 1) of the extensive-margin indicator variable on the treatment dummies; see Model I of Table 3.

An important observation is that the extensive margin response is asymmetric; the treatment effect for the wage-increase group is economically and statistically different

[21] Note that this implies that all subjects in this sample have transcribed at least six pictures and we are left with around 60% ($N = 720$) of the original sample at this point.

from that for the wage-decrease group (p-value: 0.094, calculated based on equation 2). The asymmetry is also evident in the implied elasticities (as calculated by equation 3), which is 3.0 in the increase group and 6.3 in the decrease group.

We further investigate the dynamics of the treatment effects in Figure 7, which graphically presents the coefficients from regression equation 4. The left panel shows that our treatment affects subjects throughout the entire experiment until the maximum number of transcribed pictures is reached. In each period after the treatment notification, individuals in the increase group are more likely to still participate in our experiment relative to the control group. Similarly, workers in the decrease group are less likely to continue working compared to the control group. Both effects are significantly different from zero for each period until the end of the experiment in $t = 50$. As expected, treatment effects at the extensive margin are especially strong immediately after the treatment and become less important the longer subjects continue working. The right panel of Figure 7 investigates the asymmetry of this effect, i.e., the sum of the estimated coefficients $\widehat{\gamma}^t_{Increase}$ and $\widehat{\gamma}^t_{Decrease}$ from equation (4). We find that the asymmetry in the extensive margin response mainly occurs immediately after the treatment notification and we cannot reject symmetry five or more periods after treatment notification.

4.3 Time responses

This section describes results for time related outcome variables, again using the subsample of workers who saw the treatment notification. We show means along with medians to account for potential outliers in the time participants spent working.

Time Spent Working. Figure 8 shows that, on average, subjects in the control group spent about 61 minutes working on the labor task with a median of 44 minutes. Relative to the control group, workers who experienced a wage increase worked on the task for 6 additional minutes on average (11 minutes difference at the median) while those who experienced a wage decrease spent on average 11 fewer minutes (16 minutes at the median) working on the task. A non-parametric test shows that the treatment effect is statistically different from zero for the wage-decrease group, but not for the wage increase group. The non-parametric results are also reflected by the regressions; see Model III of Table 3. Again, these results indicate that labor supply and wages are positively related.

The differences indicate an asymmetric effect; the treatment effect is larger in the wage-decrease group than in the wage-increase group. This is also evident by the implied elasticities, which are 0.50 in the increase group and 0.87 in the decrease group. However, we cannot reject the null that the difference between the treatment effects is zero. In other words, though the relative magnitude of the treatment effects is indicative of an asymmetric response, we cannot rule out symmetry in a statistical sense.

The effect on the total time spent working described above can be decomposed into two parts; the first due to the extensive margin response and the second due to the intensive margin response. We identify the contribution of the intensive margin response in Figure 9, which plots the mean and median of the total number of minutes worked conditional on *not* quitting right away after the treatment. The Figure shows that, conditional on transcribing at least one picture after the treatment notification, workers in the control group spent an average of 68 minutes on the task (with a median of 50). Relative to the control group, workers in the wage-increase group worked for one additional minute on average (7 additional minutes at the median) while workers in the wage-decrease group spent on average 4 fewer minutes on the task (median 8). Subtracting these average intensive-time-margin treatment effects from the total treatment effects implies that the extensive margin response explains the overwhelming majority of the effect on time spent working on the task. In fact, the extensive margin response explains 83% (= (6 − 1)/6) and 64% (= (11 − 4)/11) of the time margin response in the wage increase and decrease groups, respectively.

Transcription Rate. The results for the transcription rate are shown in Figure 10. Workers on average spent 3.8, 3.4 and 3.9 minutes for one picture in the control, increase and decrease groups, respectively. The differences between groups are not statistically significant (also see Model IV in Table 3). We further separate this total effect into its intensive and extensive margin components and find that there is no statistically significant effect on either margin (see Figure 11 which reports the transcription rate conditional on completing at least one transcription after the treatment notification).

4.4 Number and quality of transcriptions

This section describes the treatment effects on the number of transcribed pictures and the quality of transcription. All analyses are again based on the subsample of workers who saw the treatment notification.

Number of Transcribed Pictures. Figure 12 shows that the treatment variation clearly affected the number of transcribed pictures per worker. While the average worker transcribed 19.04 images in the control group, the average worker completed 22.35 and 15.25 pictures in the wage-increase and wage-decrease groups, respectively. All group-wise differences between groups are distinguishable from zero at the 1%-level according to non-parametric rank-sum tests. The relationship between labor supply and wages is therefore again positive. These results are confirmed in Model II of Regression Table 3, which also shows that we cannot reject the null that the wage effect on total output is symmetric.

As in section 4.3, we decompose the total effect on number of transcribed pictures into its intensive and extensive margin components. We begin with the contribution of the intensive margin response by calculating the per-worker number of transcriptions for each group conditional on completing at least one picture after seeing the treatment information. These results, which are presented in Figure 13, show that output is higher when wages rise and lower when wages fall. While the non-parametric tests reveal that the difference between the control and increase group is statistically significant, the difference between control and decrease is not significant (p-value: 0.15). More importantly, the magnitude of these intensive-time-margin effects is not asymmetric in a statistical sense.

We next identify the contribution of the extensive margin response by subtracting the intensive-time-margin effect from the total effect. For example, the total treatment effect for the wage-increase group is 3.3 transcriptions. From Figure 13, we know that 2.14 of this effect is due to the intensive-time-margin response. Therefore, the balance of 1.17 ($= 3.31 - 2.14$) is due to the extensive margin response. A similar calculation for the wage-decrease group reveals that the contribution of the extensive margin is 2.19.

Quality of transcriptions. Figure 14 depicts that the wage-rate changes did not have any effects on the quality of transcription. The differences are tiny and indistinguishable from zero, which confirms that workers in all groups paid careful attention to the task. This result is in line with the field experiment of Kube et al. (2013) who do not find any effects of wages on work quality either.

4.5 Robustness

Because the workers discussed our task on the mTurk forum, it is possible that our findings are driven by selection into the HIT. We explore this by performing the analyses separately on the sample of workers who worked on our task before it was discussed online and the sample of workers who worked on it afterwards. These results, which are presented in Appendix B, show no evidence that our results are driven by selection among workers who participated in the post-forum period. In addition, we regress each outcome variable on a dummy variable indicating whether the subject worked on the task before or after the forum post; we do not find any significant effects of working on the task after the forum post (results not reported).

5 Discussion of results

In this section, we first discuss the potential economic explanations behind our results, and then describe the policy implications and generalizability of our findings.

5.1 Mechanisms

Our results show that the extensive margin response to wage changes is strongly asymmetric. We also find evidence of an asymmetric intensive-time-margin response, but this effect is not statistically distinguishable from zero. Similarly, the wage-induced effect on number of transcribed images is marginally asymmetric, though not statistically significant. How can these results be rationalized? In the following, we discuss several channels that help to understand the economic mechanisms behind our results.

Standard Labor Supply. One possible explanation of our *extensive-margin* results is that they are driven by a rational response to the difference between reservation wages and the newly announced wage: A worker's decision to work or not is determined by the wage rate relative to the worker's reservation wage, and the worker chooses to work if the wage rate is greater than her reservation wage. Since participation in our experiment is voluntary, it is reasonable to assume that the reservation wage for our workers has a distribution that is bounded between \$0 and \$0.15. This raises the possibility that some workers have reservation wage between \$0.12 and \$0.15. If true, this would make the observed responses consistent with a rational calculus of reservation wages. In particular, we would expect all rational workers with reservation wage between \$0.12 and \$0.15 to quit the labor task when the wage rate decreases to \$0.12. In contrast, workers whose wages stay constant or increase keep working because the new wages are at least as large as their reservation wage. As a result, the response to the wage decrease is larger than to the wage increase.

There are two potential reasons why it is not immediately clear that our results can be explained by this story of reservation wages. First, previous studies by Horton and Chilton (2010) and Horton et al. (2011) find that mTurkers have a median reservation wage between \$1.12 and \$1.38 per hour,[22] which is substantially lower than the implied hourly wage of \$2.10 in our wage decrease group.[23] However, although the median hourly reservation wage is considerably lower than the hourly wage in our decrease group, it is likely that the number of workers with reservation wage between \$0.12 and \$0.15 is *not* zero, suggesting that we should see a response to the wage decrease at the extensive margin even in light of a very low average reservation wage.

[22]Horton et al. (2011) estimate a median reservation wage of \$1.12 using data generated from an mTurk task that is similar to our task. This task required mTurk workers to transcribe paragraph-sized chunks of text that are written in Tagalog, a language of the Philippines. That is, as in our task, subjects are required to transcribe foreign language (workers in their task were not from the Philippines) and are paid per transcribed text.

[23]The implied hourly wage in the decrease group is calculated based on the observation that workers in the decrease group transcribe about 15.18 pics per hour: $2.10 = 0.1 + 6 \times 0.15 + 9.18 \times 0.12$. Note that this hourly wage is a lower bound because our measure of the time it takes to transcribe one picture overstates the actual time per picture; see Section 3.1.

Second, we observe a statistically significant extensive margin response in the wage-increase group, which, at first glance, appears inconsistent with the reservation-wage argument since every worker in this group would have been paid above her reservation wage from the beginning of the experiment. However, the reservation-wage story might explain this result if workers have imperfect information about the disutility of the labor task before they start working on our task. Because workers make their decision to start working on our task based solely on our description on the mTurk website, they are only able to form an expectation regarding the disutility of the labor task. Once workers transcribe the first batch of six pictures, they are able to update their estimate of the costs of working and the decision to continue working after this first batch of transcriptions is based on this updated estimate. Some subjects may have underestimated the disutility of working on the task and will quit the task after the first pictures even in the absence of any wage changes. This is consistent with our observation of positive quit-rates in the control group.[24] This mechanism additionally suggests that the share of workers who quit after a wage decrease is larger than in the control group without wage changes. It also suggests that the share of workers who quit in the wage-increase group is smaller than in the control group, but potentially still positive. As a result, this argument of updated beliefs about the disutility of the labor task, along with a non-zero number of workers whose reservation is between \$0.12 and \$0.15, provides a rationale for an asymmetric labor supply response to wage increases and decreases, which is consistent with a standard labor-supply model with reservation wages.

So what about the *intensive-margin* results? Could these results be due to the standard model? Notice that the intensive margin response is based on the difference between the marginal disutility of transcription and the wage rate. Assuming the disutility of transcription is increasing in the number of transcriptions, we would expect workers in the wage-increase group to work longer and faster, relative to the control group. On the other hand, because the wage-decrease group faces a lower wage than the control group, we would expect workers in this group to spend less time working and to do so at a slower rate. This is exactly what we find, implying that we can confirm a positive relation between labor supply and wages. We further find indications that the economic magnitude of these responses is asymmetric; e.g., the intensive-margin treatment effect for the time spent working in the wage-decrease group is four times that in the wage-increase group. The standard neoclassical labor supply model may yield such asymmetric labor supply responses if the labor supply function has a particular shape where an increase triggers a smaller response than a decrease.[25] Even then, asymmetry would only arise

[24]Note that another reason for quits in the control group is that we intentionally nudge workers to quit after six pictures; see the notification that we display to workers in all groups after the first six transcribed pictures.

[25]For example, a CRRA utility function $\frac{1}{1-\gamma}(wL)^{1-\gamma}$ with linear costs aL yields asymmetric labor

for non-marginal wage changes. Although our absolute change in the wage rate of \$0.03 is small, the relative change is 20% and therefore unlikely to be perceived as marginal. This implies that our findings regarding the intensive margin are also consistent with a standard model of labor supply.

Loss Aversion. Our findings can also be rationalized by a model of loss aversion where the reference wage is equal to the expected wage of \$0.15 and the labor supply curve is kinked at this reference wage (a model of this type has for example been put forward by Ahrens et al., 2014; see Section C in the Appendix).[26] In such a model, workers are loss averse with reference-dependent utility functions; workers' labor-supply functions are kinked at at a reference wage and have steeper slopes in the gain domain than in the loss domain. As a result, workers are less willing to supply an additional unit of labor when the wage is above the reference wage than when it is below, and a wage decrease is predicted to have a larger labor-supply effect than a wage increase of equal magnitude.

The main insight from this potential mechanism is sketched in Figure 17, which relates leisure and wages. A worker who is located at the reference wage, denoted R, will respond differently to wage increases and decreases of equal magnitude. In particular, a worker at the reference point weights wage decreases more heavily than wage increases. As a result, she will respond more strongly to a wage decrease (by working less) than an equally sized wage increase (to which she will respond through more labor supply). This result implies that labor supply elasticities identified from wage increases are predicted to be smaller than labor supply elasticities identified from wage decreases. Our findings are consistent with this prediction.

Reciprocity. Another potential explanation of our findings is reciprocity; workers interpret the wage changes as punishment or reward, and respond accordingly. Workers who receive a wage decrease feel punished and thus lower their labor supply in an effort to punish the employer, while workers who receive a wage increase feel rewarded for their effort and thus work harder to return the favor to their employer. To the extent that the degree of induced reciprocity is asymmetric, this explanation is potentially consistent with our findings. Although we have no way of ruling out this motivation behind our results, we argue that this is an unlikely explanation based on our experimental design. Recall that subjects are paid for each completed transcription and not per unit of time. This implies that workers in our experiment have little scope for punishing the employer

supply responses for non-marginal wage changes.

[26] It is reasonable to assume that workers in our field experiment expect the per-unit wage to remain constant at \$0.15. The literature typically finds that reference points depend on expectations (e.g., Koszegi and Rabin, 2006; Abeler et al., 2011; Ahrens et al., 2014), suggesting that \$0.15 would constitute the reference wage if a model of loss aversion was applied to our context.

through shirking. Reducing the number of transcription implies that a worker punishes herself in the form of lower pay-off, and potentially lower performance rating, which affects her prospects of being allowed to work on other mTurk tasks.[27] One strategy to punish the employer without incurring a cost is to continue to work hard, but submit transcriptions that are of low enough quality to be mostly useless to the employer, but high enough quality to avoid a negative performance review. Because we have the transcriptions and the actual texts, we can check to see if workers used this strategy; there is no evidence that they did (see Section 4.4).

Similarly, as opposed to settings where workers are paid per hour, transcribing more pictures is not a reward for the employer in our experiment since this increases the costs to the employer. Workers are also likely to know that employers can easily recruit other workers to transcribe pictures and that employers therefore do not face the risk that pictures remain untranscribed.

Treatment-induced heterogeneous skill composition. One could think that the skill composition (i.e., the ability to transcribe pictures) of workers in the three experimental groups is affected by the treatment variations. For example, it might be plausible that the least productive workers drop out of the labor task once they are faced with a wage decrease, whereas similar unskilled workers stay after a wage increase. This would then imply that the share of unskilled workers in the decrease group would be smaller after treatment notification than in the increase group. As a result, differences across groups might simply be a result of heterogeneous attrition due to the treatments.

Using all individuals in our work task (thus not only those who saw the treatment), Figures 15 and 16 show that transcription skills do not affect drop-out rates. Figure 15 shows the median time per transcribed picture over the course of all periods.[28] We use median time per transcribed picture as a measure of transcription skills. The Figure shows that average transcription skills in the three experimental groups are not affected by the treatment notification; the lines evolve similarly over the periods in all three groups. Figure 16 shows the estimates of a regression of a 'Last-picture-transcribed' dummy (i.e., a dummy indicating if the respective period is the last period of the respective individual) on a full interaction of period dummies and the transcription time in the respective period. The Figure shows that transcription skills (as measured by transcription time) do not determine exit rates. In other words, being talented or not in transcribing pictures does not predict if a participant drops out of the labor task.

[27]Workers on mTurk receive performance rating for each task they complete. Employers often use workers' performance rating to screen out low performers from their tasks. Therefore, a worker who decides to punish us because their wage has been reduced, runs the risk of limiting the number of tasks she will qualify to work on in the future.

[28]For example, in period 10 the Figure shows the median time that it took workers in the experimental groups to transcribe the 10th picture.

5.2 Implications

The existing labor-supply literature often identifies labor supply elasticities by exploiting panel data comprised of both wage increases and decreases. This approach makes sense in the context of *marginal* wage changes where the elasticity is shown to be symmetric. However, this approach becomes problematic when on considers that wage changes are generally non-marginal. The reason is that both the standard and behaviorally-inspired models can be used to show that labor supply responses to non-marginal wage changes need not be symmetric. Consistent with the theoretical finding of Ahrens et al. (2014) and the empirical results of Kube et al. (2013), we show that labor supply responds asymmetrically to non-marginal wage changes.

Our findings suggest that ignoring the direction of wage changes when estimating labor supply elasticities leads to biased own-wage labor-supply elasticities; elasticities are overestimated when wages rise and underestimated when wages fall. Importantly, and as an addition to the previous literature, we find that the asymmetry of labor supply w.r.t. wages is more pronounced on the extensive margin relative to the intensive margin. This refinement of the asymmetry is especially important since it is generally accepted that labor supply elasticities are mainly determined by the extensive margin response (Blundell and MaCurdy, 1999; Meghir and Phillips, 2010; Bargain et al., 2014). Our findings are also practically useful, since labor supply elasticities play an important role in quantifying the economic impacts of policy changes that affect wages; e.g., minimum wage polices.

Additionally, our results highlight one potential reason for the downward rigidity in wages. In particular, it is widely observed that nominal wages tend to move in one direction only. Prominent explanations for downward nominal rigidities include institutions such as minimum wages and collective bargaining. Recent evidence by Kaur (2017), however, shows that wages are downward rigid even in the absence of such institutions. Among the many potential explanations for this observed rigidity are the potentially significant effects on productivity and labor supply. However, the scarcity of nominal wage cuts makes it challenging to determine if the labor supply and production impacts of nominal wage cuts are indeed large and negative. Our study adds to the small literature that have explored this question. As mentioned before, we find large negative extensive margin responses, which suggest that nominal wage cuts could be potentially damaging for firms. This is one reason behind the reluctance of firms to reduce wages.

A policy area where our findings are likely to be particularly useful is taxation. First, tax reforms generally result in either tax increases or tax decreases, which translate into changing after-tax wages. In fact, upward and downward changes are more prominent for tax rates than for wages. We know that the tax elasticity of labor supply is generally larger than the wage elasticity; e.g., due to tax aversion (Kessler and Norton, 2016). This suggests that the labor-supply asymmetry with respect to tax-rate changes is likely to be more pronounced than what our findings for labor supply responses to wage changes suggest. This makes the distinction between rate increases and decreases particularly important in the context of tax rate. Second, our findings also raise questions about the elasticity of taxable income which plays a crucial role in our understanding of the efficiency costs of taxation and which is often estimated based on panel-data exploiting multiple variations in tax rates (e.g., Saez et al., 2012). In particular, our results suggest that failure to distinguish between ETI estimated with tax rate increases and ETI estimated with tax rate decreases is likely to lead to an underestimation of the efficiency cost of tax rate increases. This problem is likely to be even more important than with wage changes since tax rates generally move freely in both directions. Of course, the labor supply response to wage changes is not necessarily identical to the response to tax rate changes. Therefore, we are cautious in generalizing our results to the case of tax rate changes.[29]

5.3 Generalizability

The results described above are obtained using an experimental design in a large real-world labor market. Importantly, workers did not know they participated in an experiment and thus behaved as they would in their natural occurring environment. At least two aspects of our empirical results suggest that workers took the task seriously and supplied labor in a "plausible" way. First, the overall quality is very high with an accuracy rate of more than 97% in all three groups. Second, we find evidence for an upward sloping labor supply curve – that is, workers work more as wages increase and less when wages go down –, which is consistent with the empirical labor supply literature (e.g., Keane, 2011; Bargain et al., 2014). These two points also indicate that labor-supply behavior in our online task has some implications for labor-supply behavior in other, more traditional labor markets.

Due to randomization, our experimental design also guarantees internal validity. Though our findings are based on an actual real-world labor market, we are careful not to generalize our results to all types of labor markets. Nonetheless, we argue that the findings are applicable to labor markets with piece rate, flexibility and multiple outside options. One example of such labor markets is on-line crowd-sourcing labor markets, which are becoming increasingly common in the current technological age.[30] A common feature of these labor markets is that workers tend to work for relatively low wages and have extremely high levels of labor supply flexibility. Because the labor supply effects are predominately on the extensive margin, we argue that the results are also likely to be

[29] We are not aware of any evidence on the asymmetric effects of tax increases and decreases on labor supply or taxable income. Benzarti et al. (2017) show that increases in Value Added Taxes (VAT) have a larger effect on prices than VAT reductions.

[30] See `https://sites.google.com/site/johnjosephhorton/miscellany/online-labor-markets` for a list of crowd-sourcing online labor markets.

equally applicable to traditional labor markets where workers face greater restrictions on labor hours.

6 Conclusion

We estimate the effect of (non-marginal) wage changes on labor supply using data generated in a field experiment on Amazon's Mechanical Turk. Our findings show that the labor-supply curve for workers on mTurk is upward sloping; the relationship between wages and labor supply is positive both for the case of wage increases and wage decreases. We further find strong evidence of an asymmetric response on the extensive margin; the extensive-margin response for wage decreases is twice as large as for equally sized increases. The magnitude of the intensive-time margin responses is also indicative of an asymmetric response, but we cannot rule out symmetry in a statistical sense. These results are consistent with a standard labor-supply model as well as reference-dependent preferences. Importantly, they are not driven by treatment-induced heterogeneous skill compositions.

References

Abeler, J., Falk, A., Goette, L. and Huffman, D. (2011). Reference Points and Effort Provision, *American Economic Review* **101**(2): 470–92.

Ahrens, S., Pirschel, I. and Snower, D. J. (2014). A Theory of Wage Adjustment under Loss Aversion, *IZA Discussion Papers No 8699*.

Bargain, O., Orsini, K. and Peichl, A. (2014). Comparing Labor Supply Elasticities in Europe and the United States – New Results, *The Journal of Human Resources* **49**(3): 723–838.

Benzarti, Y., Carlon, D., Harju, J. and Kosonen, T. (2017). What Goes Up May Not Come Down: Asymmetric Incidence of Value Added Taxes, *mimeo, http://media.wix.com/ugd/d10d21_7e02d776a50445c7b80ffcdb03ec51f7.pdf*.

Bewley, T. F. (2005). Fairness, Reciprocity, and Wage Rigidity, *in* H. Gintis, S.Bowles, R. T. Boyd and E. Fehr (eds), *Moral Sentiments and Material Interests: The Foundations of Cooperation in Economic Life*, MIT Press.

Blundell, R. and MaCurdy, T. E. (1999). Labor Supply: A Review of Alternative Approaches, *in* O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics*, North-Holland, Amsterdam.

Buhrmester, M., Kwang, T. and Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?, *Perspectives on Psychological Science* **6**(1): 3–5.

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.

Card, D. and Dahl, G. B. (2011). Family Violence and Football: The Effect of Unexpected Emotional Cues on Violent Behavior, *The Quarterly Journal of Economics* **126**(1): 103–143.

Charness, G. and Kuhn, P. (2011). Lab labor: What can labor Economists learn from the lab?, *in* O. Ashenfelter and D. Card (eds), *Handbook of Labor Economics Vol 4A*, North Holland, Amsterdam, pp. 229–330.

Chemin, M. and Kurmann, A. (2014). Do Workers Feel Entitled to High Wages? Evidence from a Field Experiment, *mimeo, available on author website: http://www.andrekurmann.com/test*.

Cohn, A., Fehr, E. and Goette, L. (2015). Fair Wages and Effort Provision: Combining Evidence from a Choice Experiment and a Field Experiment, *Management Science* **61**(8): 1777–1794.

Crawford, V. P. and Meng, J. (2011). New York City Cab Drivers' Labor Supply Revisited: Reference-Dependent Preferences with Rational-Expectations Targets for Hours and Income, *American Economic Review* **101**(5): 1912–32.

DellaVigna, S., Lindner, A., Reizer, B. and Schmieder, J. F. (2017). Reference-Dependent Job Search: Evidence from Hungary, *Quarterly Journal of Economics* **132**(4): 1969–2018.

DellaVigna, S. and Pope, D. (2017). What Motivates Effort? Evidence and Expert Forecasts, *Review of Economic Studies* . Forthcoming.

Dunn, L. F. (1996). Loss Aversion and Adaptation in the Labor Market: Empirical Indifference Functions and Labor Supply, *The Review of Economics and Statistics* **78**(3): 441–450.

Eissa, N. and Liebman, J. B. (1996). Labor Supply Response to the Earned Income Tax Credit, *The Quarterly Journal of Economics* **111**(2): 605–637.

Evers, M., De Mooij, R. and Van Vuuren, D. (2008). The Wage Elasticity of Labour Supply: A Synthesis of Empirical Estimates, *De Economist* **156**(1): 25–43.

Falk, A., Fehr, E. and Zehnder, C. (2006). Fairness Perceptions and Reservation Wages – the Behavioral Effects of Minimum Wage Laws, *The Quarterly Journal of Economics* **121**(4): 1347–1381.

Fehr, E. and Goette, L. (2007). Do Workers Work More if Wages Are High? Evidence from a Randomized Field Experiment, *American Economic Review* **97**(1): 298–317.

Gneezy, U. and List, J. A. (2006). Putting Behavioral Economics to Work: Testing for Gift Exchange in Labor Markets Using Field Experiments, *Econometrica* **74**(5): 1365–1384.

Goette, L., Huffman, D. and Fehr, E. (2004). Loss Aversion and Labor Supply, *Journal of the European Economic Association* **2**(2-3): 216–228.

Hennig-Schmidt, H., Sadrieh, A. and Rockenbach, B. (2010). In Search of Workers' Real Effort Reciprocity - a Field and a Laboratory Experiment, *Journal of the European Economic Association* **8**(4): 817–837.

Horton, J. J. and Chilton, L. B. (2010). The Labor Economics of Paid Crowdsourcing, *Proceedings of the 11th ACM Conference on Electronic Commerce* .

Horton, J. J., Rand, D. G. and Zeckhauser, R. J. (2011). The online laboratory: conducting experiments in a real labor market, *Experimental Economics* **14**: 399–425.

Kahn, S. (1997). Evidence of Nominal Wage Stickiness from Microdata, *The American Economic Review* **87**(5): 993–1008.

Kahneman, D. and Tversky, A. (1979). Prospect Theory: An Analysis of Decision under Risk, *Econometrica* **47**(2): 263–91.

Kaur, S. (2017). Nominal Wage Rigidity in Village Labor Markets, *American Economic Review* . Forthcoming.

Keane, M. P. (2011). Labor Supply and Taxes: A Survey, *Journal of Economic Literature* **49**(4): 961–1075.

Kessler, J. B. and Norton, M. I. (2016). Tax Aversion in Labor Supply, *Journal of Economic Behavior and Organization* **124**: 15–28.

Koszegi, B. and Rabin, M. (2006). A Model of Reference-Dependent Preferences, *The Quarterly Journal of Economics* **121**(4): 1133–1165.

Koszegi, B. and Rabin, M. (2007). Reference-Dependent Risk Attitudes, *American Economic Review* **97**(4): 1047–1073.

Koszegi, B. and Rabin, M. (2009). Reference-Dependent Consumption Plans, *American Economic Review* **99**(3): 909–36.

Kube, S., Marechal, M. A. and Puppe, C. (2013). Do Wage Cuts Damage Work Morale? Evidence From a Natural Field Experiment, *Journal of the European Economic Association* **11**(4): 853–870.

Kuziemko, I., Norton, M. I., Saez, E. and Stantcheva, S. (2015). How Elastic Are Preferences for Redistribution? Evidence from Randomized Survey Experiments, *American Economic Review* **105**(4): 1478–1508.

Mann, H. B. and Whitney, D. R. (1947). On a test whether one of two random variables is stochastically larger than the other, *Annals of Mathematical Statistics* **18**: 50–60.

Marzilli Ericson, K. M. and Fuster, A. (2011). Expectations as Endowments: Evidence on Reference-Dependent Preferences from Exchange and Valuation Experiments, *The Quarterly Journal of Economics* **126**(4): 1879–1907.

Mason, W. and Suri, S. (2011). Conducting behavioral research on Amazon's Mechanical Turk, *Behavioural Research* **44**: 1–23.

Meghir, C. and Phillips, D. (2010). Labour Supply and Taxes, *in* J. Mirrless, S. Adam, T. Besley, R. Blundell, S. Bond, R. Chote, M. Gammie, P. Johnson, G. Myles and J. Poterba (eds), *Dimensions of Tax Design: The Mirrlees Review*, Oxford University Press, pp. 202–274.

Paolacci, G., Chandler, J. and Ipeirotis, P. G. (2010). Running experiments on Amazon Mechanical Turk, *Judgment and Decision Making* **5**(5).

Pope, D. G. and Schweitzer, M. E. (2011). Is Tiger Woods Loss Averse? Persistent Bias in the Face of Experience, Competition, and High Stakes, *American Economic Review* **101**(1): 129–57.

Post, T., van den Assem, M. J., Baltussen, G. and Thaler, R. H. (2008). Deal or No Deal? Decision Making under Risk in a Large-Payoff Game Show, *American Economic Review* **98**(1): 38–71.

Rothstein, J. (2010). Is the EITC as Good as an NIT? Conditional Cash Transfers and Tax Incidence, *American Economic Journal: Economic Policy* **2**(1): 177–208.

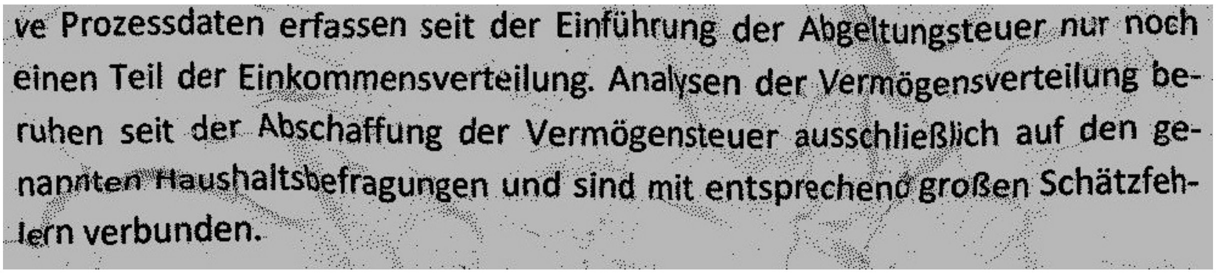
Saez, E., Slemrod, J. and Giertz, S. H. (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review, *Journal of Economic Literature* **50**(1): 3–50.

Wilcoxon, F. (1945). Individual comparisons by ranking methods, *Biometrics* **1**: 80–83.

Tables and Figures

Figures

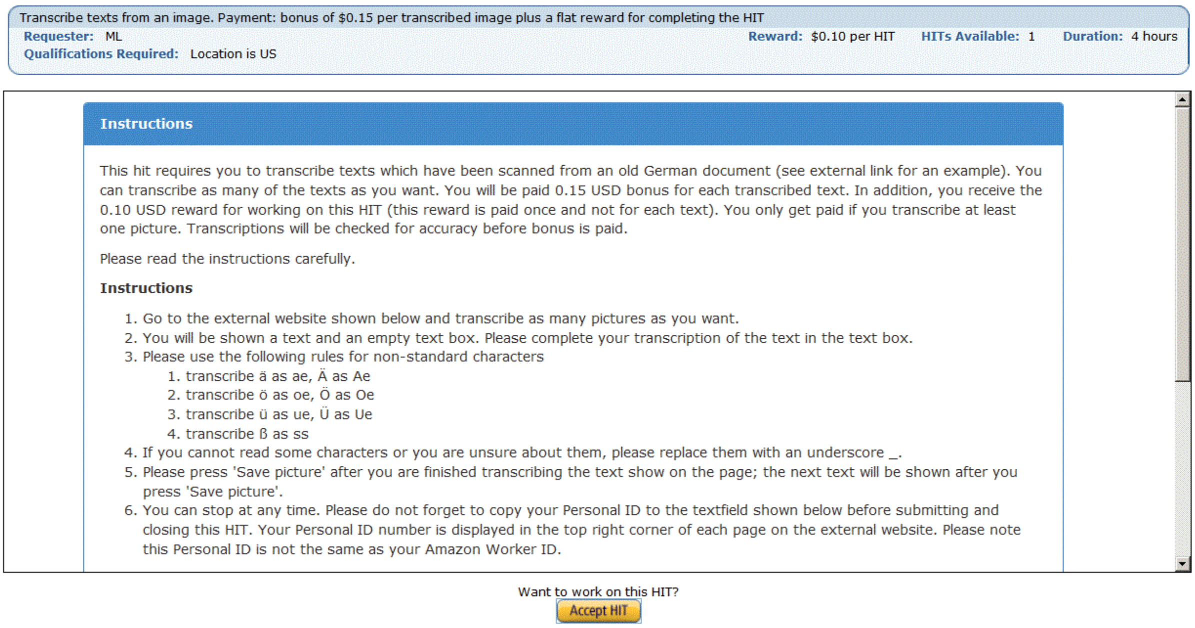
Figure 1: Image of Text to be Transcribed



ve Prozessdaten erfassen seit der Einführung der Abgeltungsteuer nur noch
einen Teil der Einkommensverteilung. Analysen der Vermögensverteilung be-
ruhen seit der Abschaffung der Vermögensteuer ausschließlich auf den ge-
nannten Haushaltsbefragungen und sind mit entsprechend großen Schätzfeh-
lern verbunden.

Notes: The Figure depicts a screenshot of an image of text that was to be transcribed by the subjects. Subjects were randomly shown one of 128 images. All images are comparable to the image depicted in the Figure. All images are in German and taken from a recent policy-report publication.

Figure 2: Human Intelligence Task Shown on mTurk



Notes: The Figure depicts a screenshot from Amazon's Mechanical Turk website. It shows how the labor task used for the field experiment was advertised on mTurk. Subjects are taken to our external website once they click the "Accept Hit" button.

Figure 3: Instructions Shown on Our Website

Transcribe pictures

Personal ID: 789db7d48af873208f7f253a6cd5a24c
Transcribed pictures: 0
Current bonus per picture: 0.15 USD

Welcome.

Thank you for working on this hit. This hit requires you to transcribe texts which have been scanned from an old German document (see below for an example). You can transcribe as many of the texts as you want; a new text will be presented when you hit the 'Save picture' button. You will be paid 0.15 USD bonus for each transcribed text. In addition, you receive the 0.10 USD reward as shown on the Amazon Mechanical Turk page for working on this HIT (this reward is paid once and not for each text). You only get paid if you transcribe at least one picture. Transcriptions will be checked for accuracy before bonus is paid.

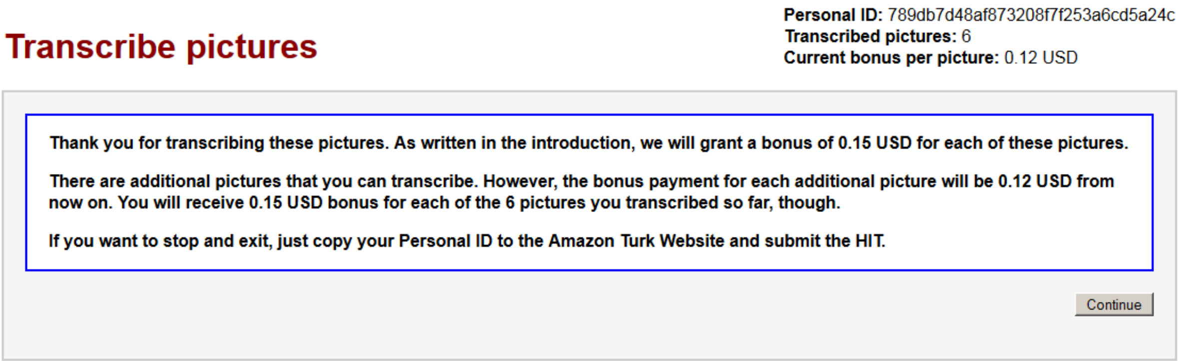
To the top right of this web page you see your personal ID. Please submit this personal identifier to Amazon Mechanical Turk in order to complete this assignment.

Instructions

1. Your Personal ID number is shown in the top right corner of each page. Please submit this personal identifier to Amazon Mechanical Turk in order to complete this assignment.
2. You will be shown a text and an empty text box. Please complete your transcription of the text in the text box.
3. Please use the following rules for non-standard characters
    1. transcribe ä as ae, Ä as Ae
    2. transcribe ö as oe, Ö as Oe
    3. transcribe ü as ue, Ü as Ue
    4. transcribe ß as ss
4. If you cannot read some characters or you are unsure about them, please replace them with an underscore _.
5. Please press 'Save picture' after you are finished transcribing the text show on the page; the next text will be shown after you press 'Save picture'.
6. You can stop at any time. Please do not forget to copy your Personal ID to the Amazon Turk Website before submitting and closing this HIT

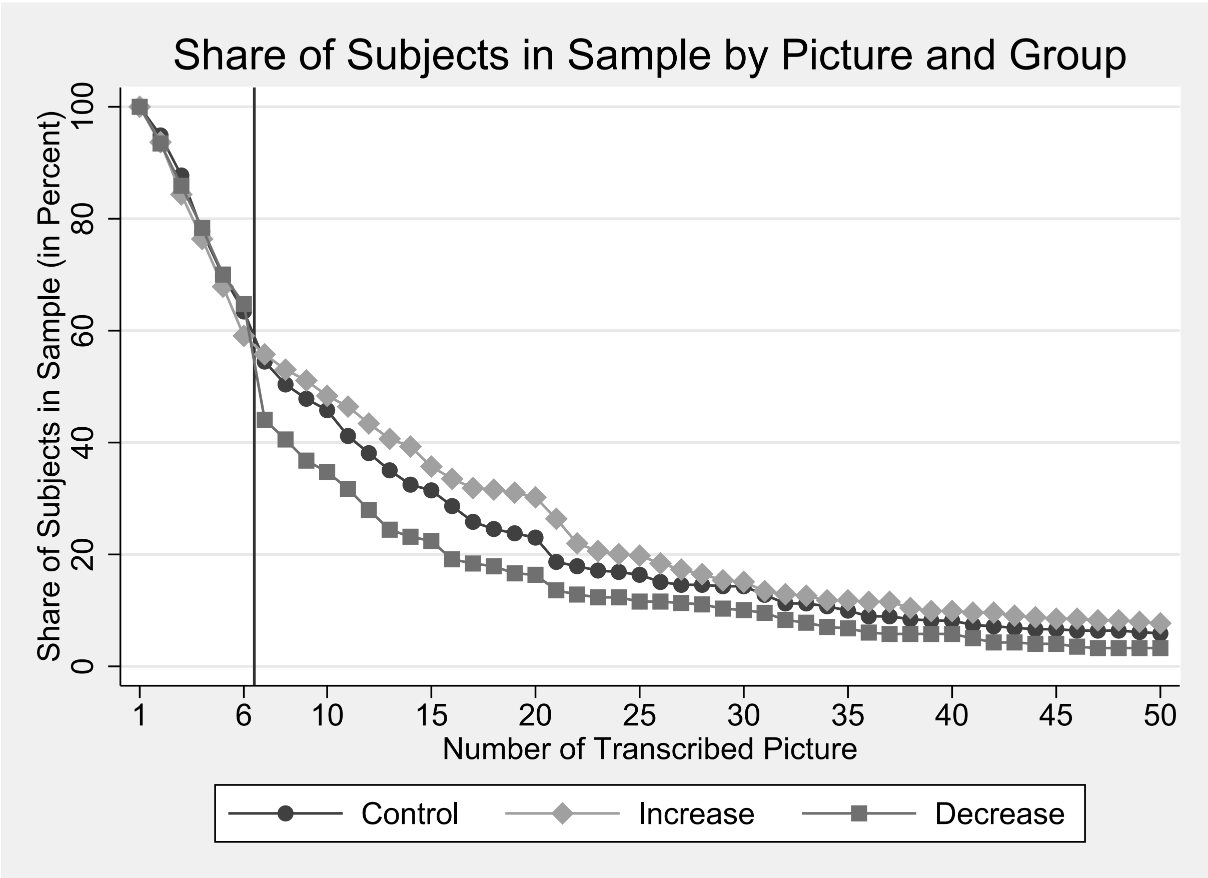
Notes: The Figure depicts a screenshot of the external website that we set up for the purpose of the field experiment. Subjects were taken to this website once they decided on Amazon's Mechanical Turk website that they would like to work on the task. The depicted screenshots provides subjects all information relevant for the task.

Figure 4: Treatment Variation



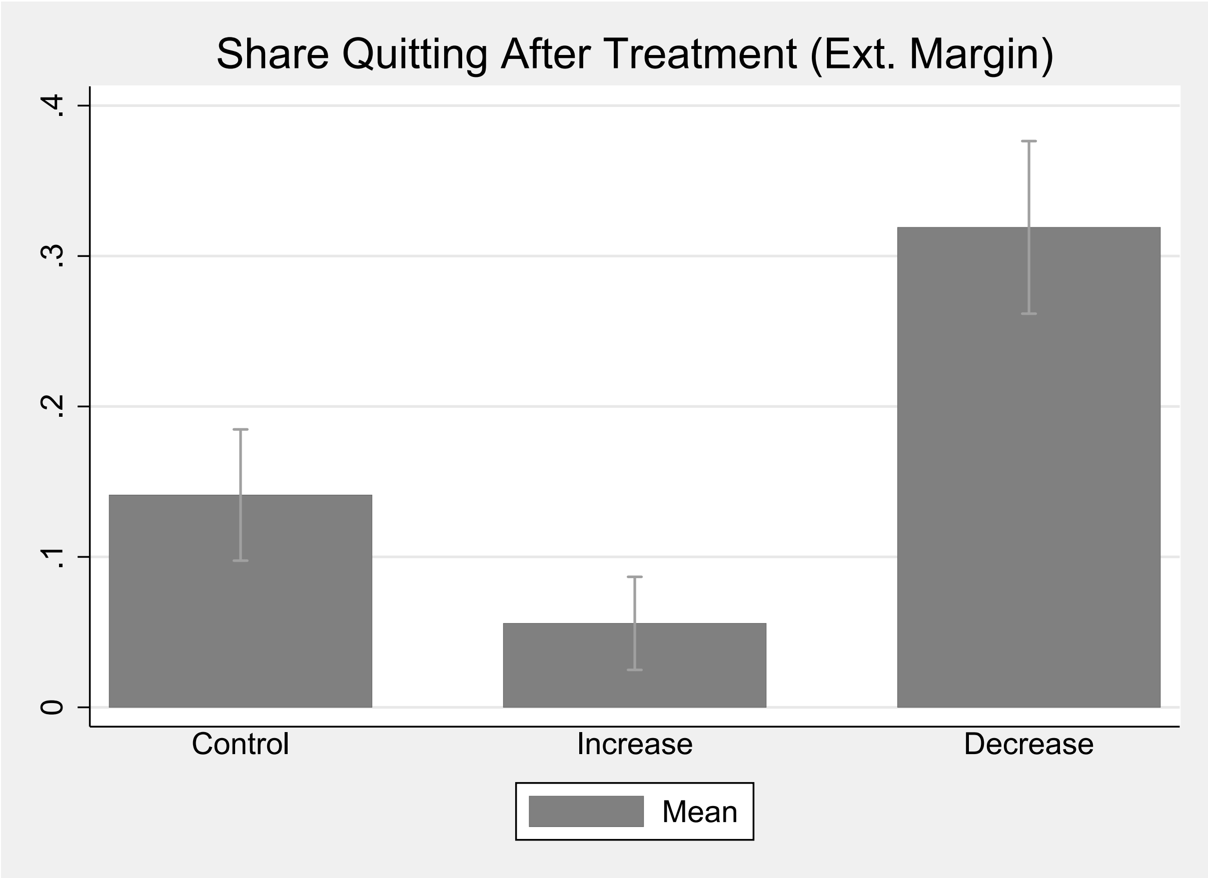
Notes: The Figure depicts a screenshot of the treatment notification in the "wage decrease" group. The treatment notifications for the "control" and "wage increase" groups were identical except for the information regarding the piece-rate wage for the subsequent images. The treatment notification popped up after a subject transcribed six images.

Figure 5: Share of Workers over Periods by Treatment Group



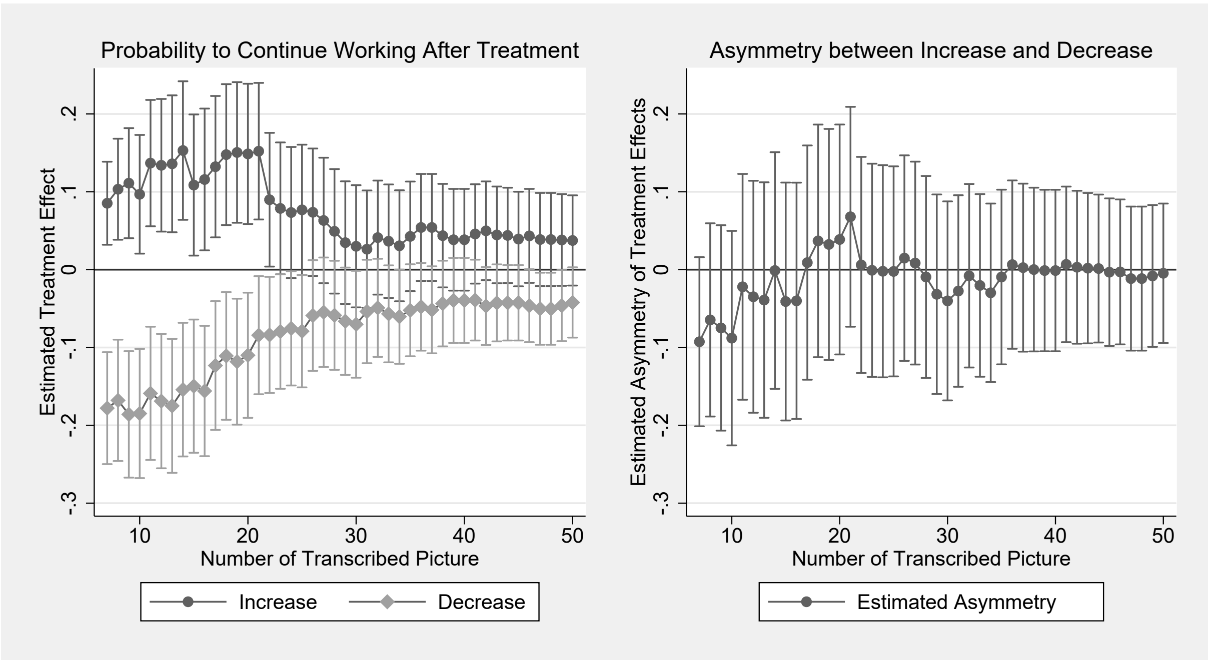
Notes: The Figure depicts the share of subjects in each period, relative to the total number of subjects who initially started the task, by experimental group. For example, the value in period 10 indicates the share of subjects who complete 10 pictures in a given experimental group as a fraction of all subjects who started working on the labor task in this experimental group. Treatment occurred after period 6. The sample includes all participants who started working on the task. Number of observations is 1152.

Figure 6: Extensive Margin by Treatment Group



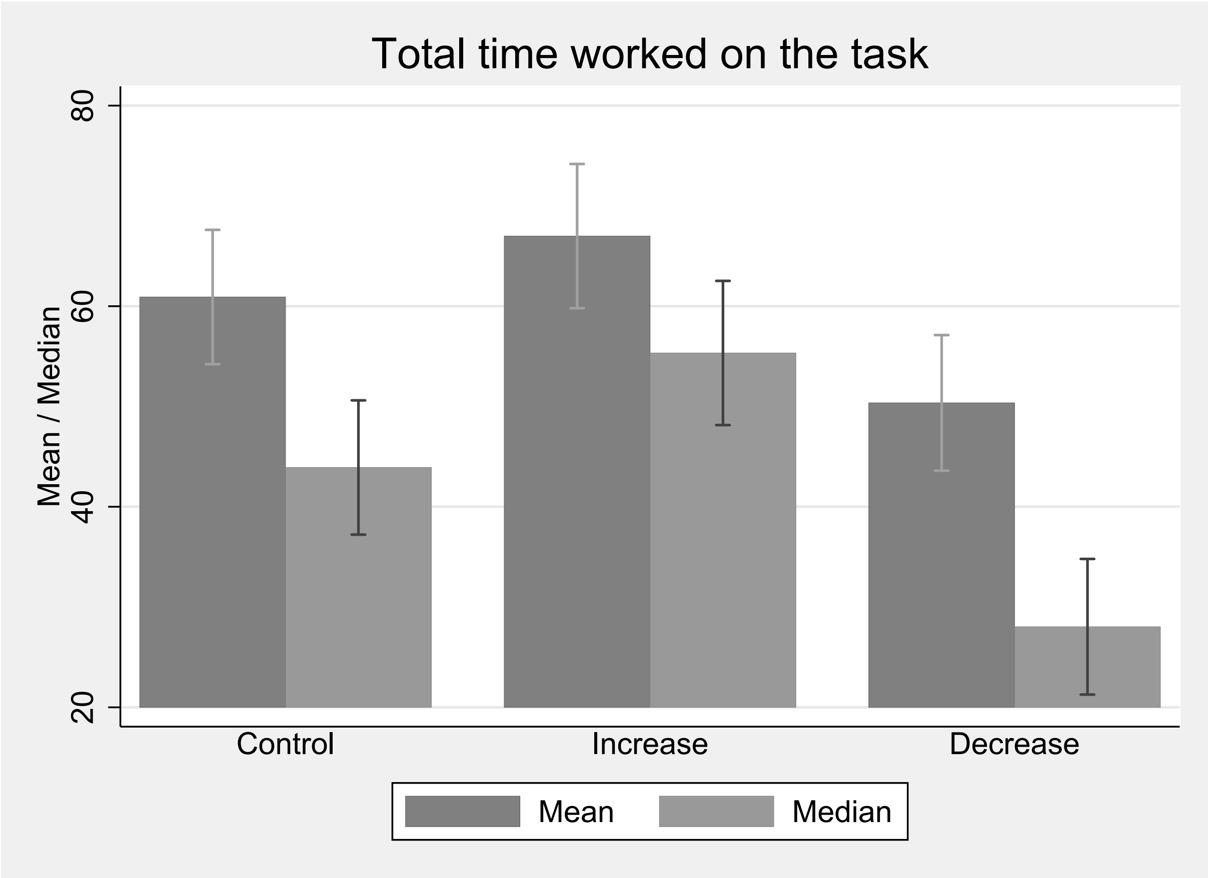
Notes: The Figure depicts the share of subjects in each group who quit the labor task immediately after seeing the treatment notification (i.e., share of subjects who transcribed six pictures but not a seventh one), along with 95% confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

Figure 7: Dynamics of Treatment Effects



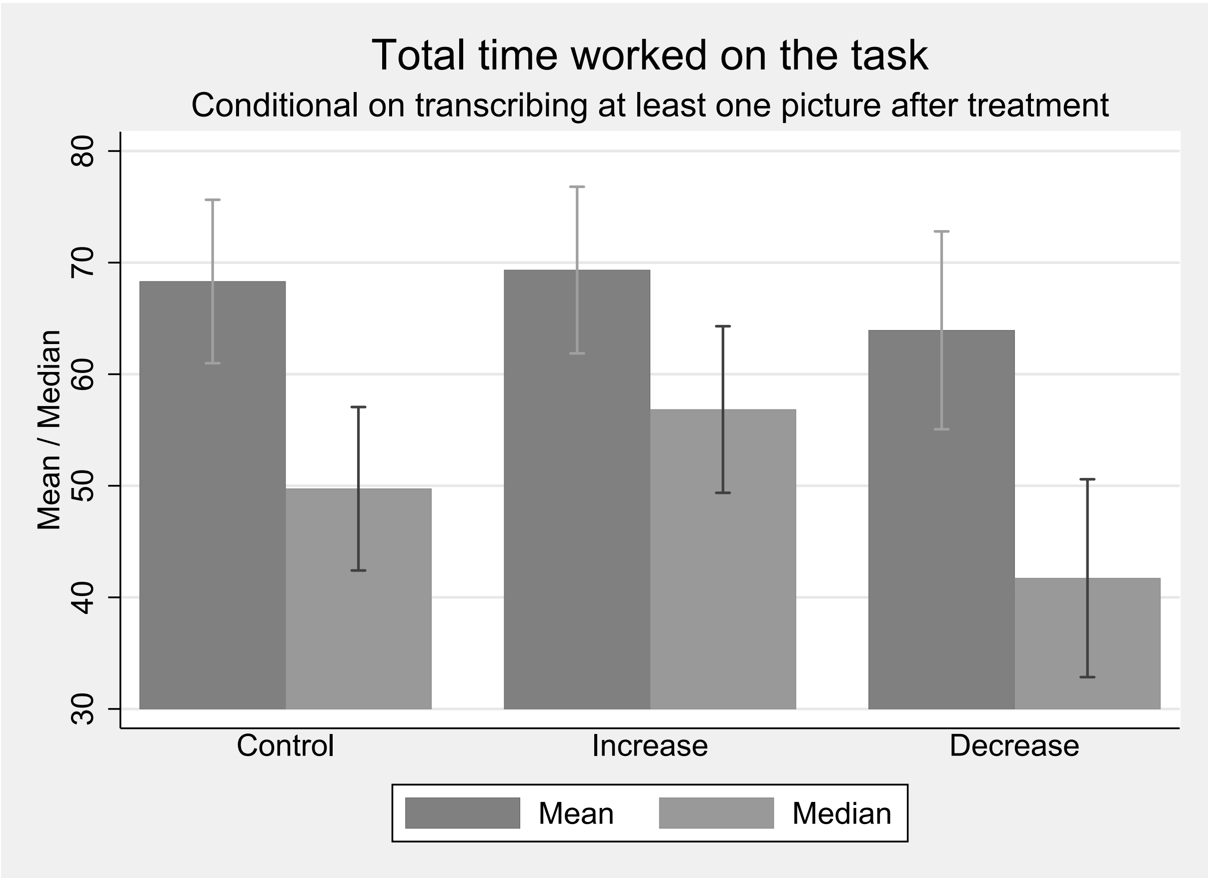
Notes: **Left panel** shows coefficients that are based on regressions of the probability of working in a given period on a full interaction of an increase-group dummy and period dummies as well as a decrease-group dummy and period dummies (see equation 4 in the main body of the paper). The outcome variable is a dummy indicating if an individual works in the respective period (for example, this dummy takes value 1 in period 10 if the respective individual transcribed the 10th picture). All effects are relative to the control group without wage change. **Right panel** shows the difference in left-panel regression coefficients between the *increase* and *decrease* treatment groups. All coefficients are shown along with 95% confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

Figure 8: Total Time Worked by Treatment Group



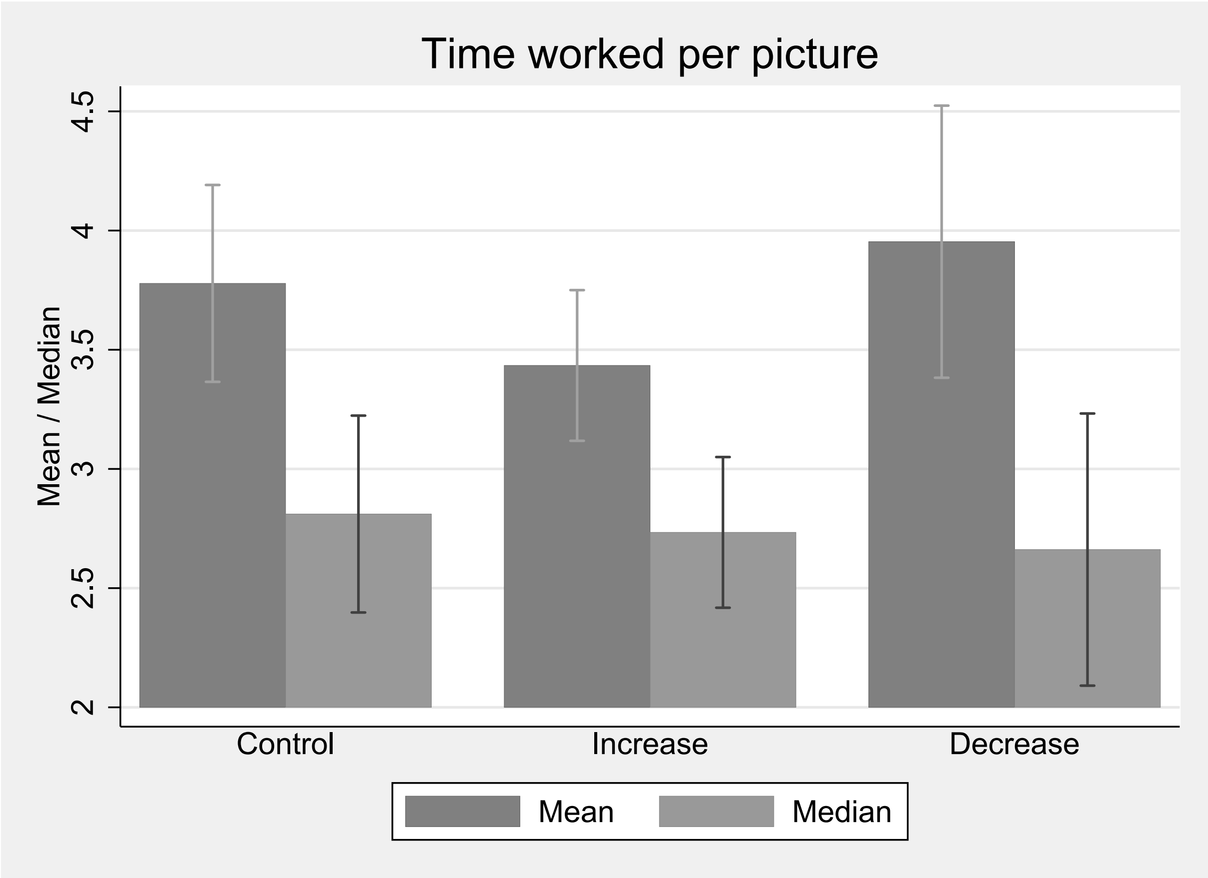
Notes: The Figure depicts for each group the average and median time (in minutes) that subjects totally spent on working on the labor task, along with 95% confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

Figure 9: Total Time Worked by Treatment Group: Intensive margin



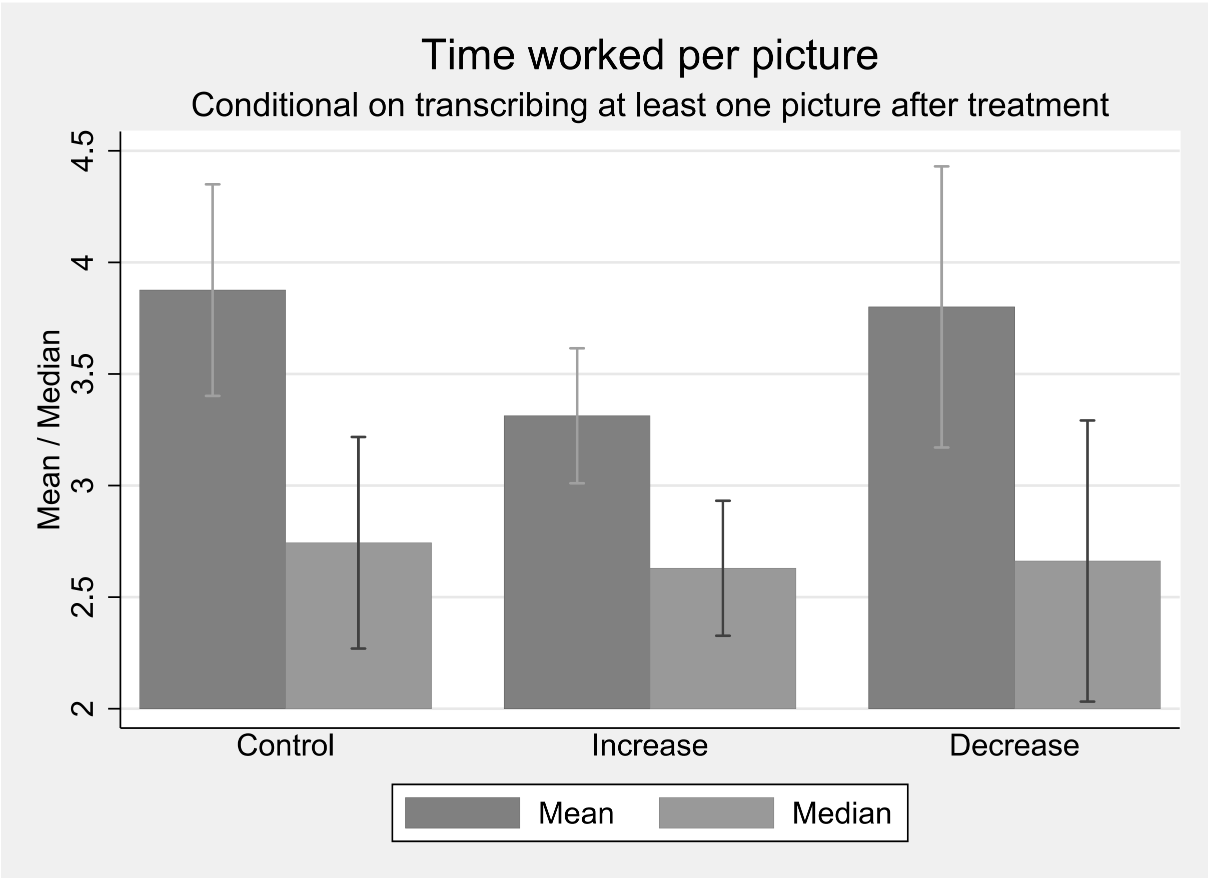
Notes: The Figure depicts for each group the average and median time (in minutes) that subjects totally spent on working on the labor task, along with 95% confidence bands. The underlying sample is restricted to subjects who did *not* quit the labor task immediately after seeing the treatment notification (i.e., restricted to subjects who have transcribed at least seven images). The number of observations is 591 with 213 subjects in the control group, 203 subjects in the ”wage increase” group and 175 subjects in the ”wage decrease” group. All 591 subjects in the sample have transcribed at least seven images.

Figure 10: Avg. Time per Transcription by Treatment Group



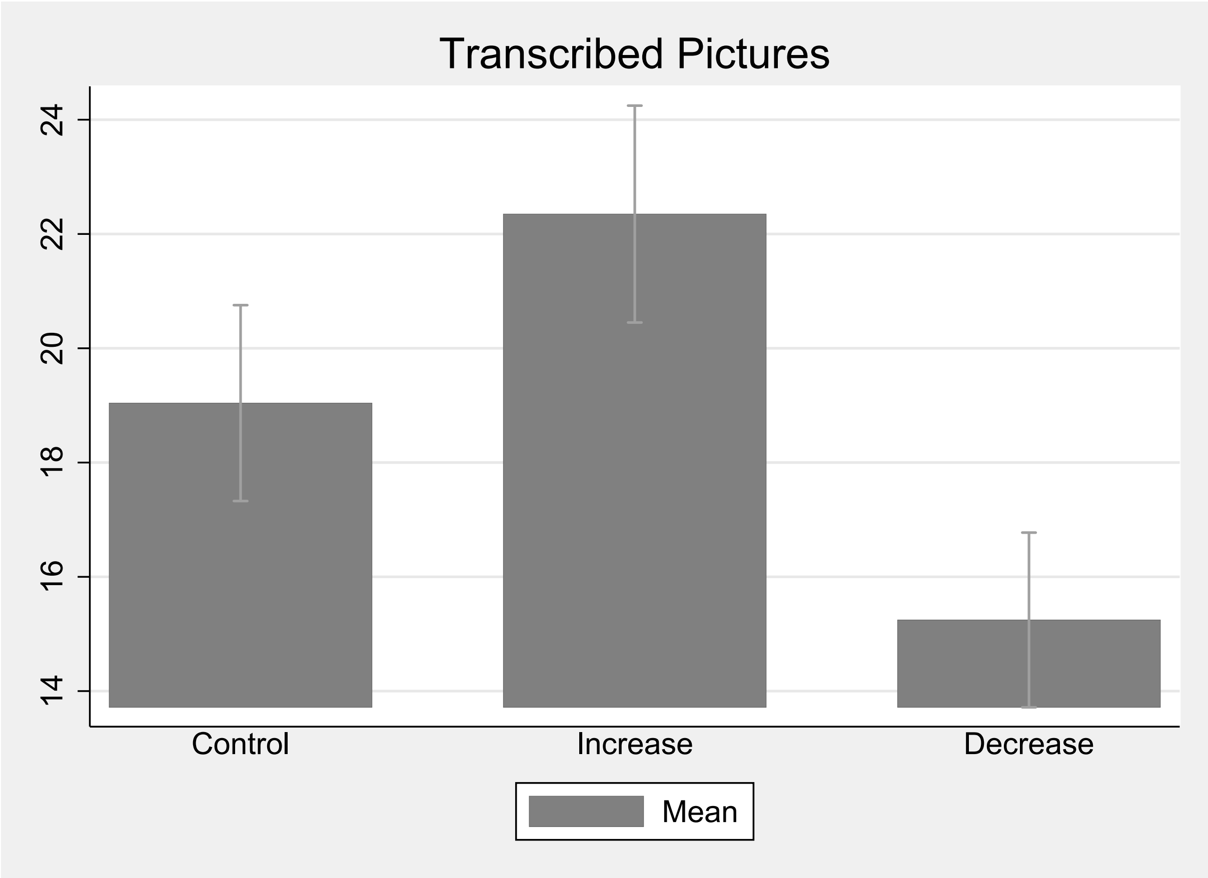
Notes: The Figure depicts for each group the average and median time (in minutes) that subjects spent working on one image, along with 95% confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the "wage increase" group and 257 subjects in the "wage decrease" group. All 720 subjects in the sample have transcribed at least six images.

Figure 11: Avg. Time per Transcription by Treatment Group: Intensive margin



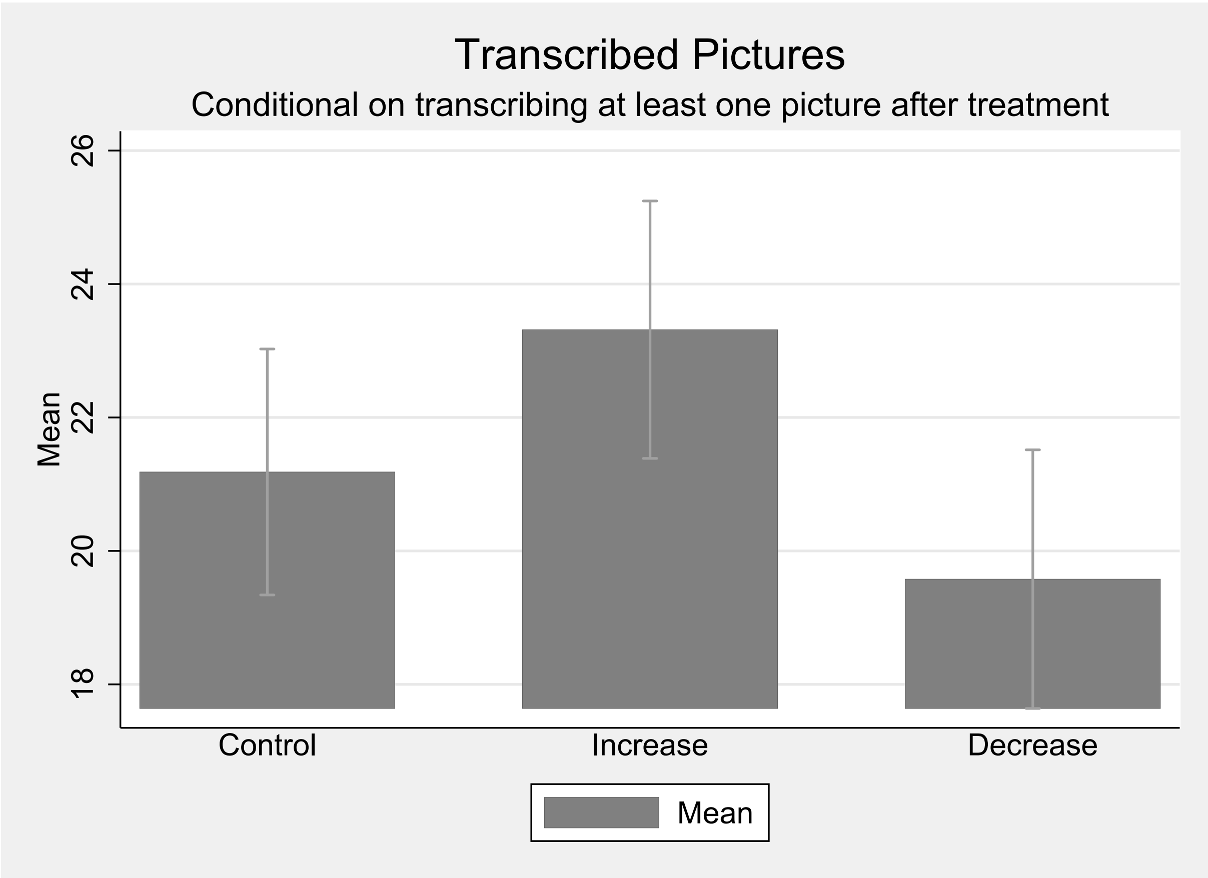
Notes: The Figure depicts for each group the average and median time (in minutes) that subjects spent working on one image, along with 95% confidence bands. The underlying sample is restricted to subjects who did *not* quit the labor task immediately after seeing the treatment notification (i.e., restricted to subjects who have transcribed at least seven images). The number of observations is 591 with 213 subjects in the control group, 203 subjects in the "wage increase" group and 175 subjects in the "wage decrease" group. All 591 subjects in the sample have transcribed at least seven images.

Figure 12: Number of Transcribed Pics by Treatment Group



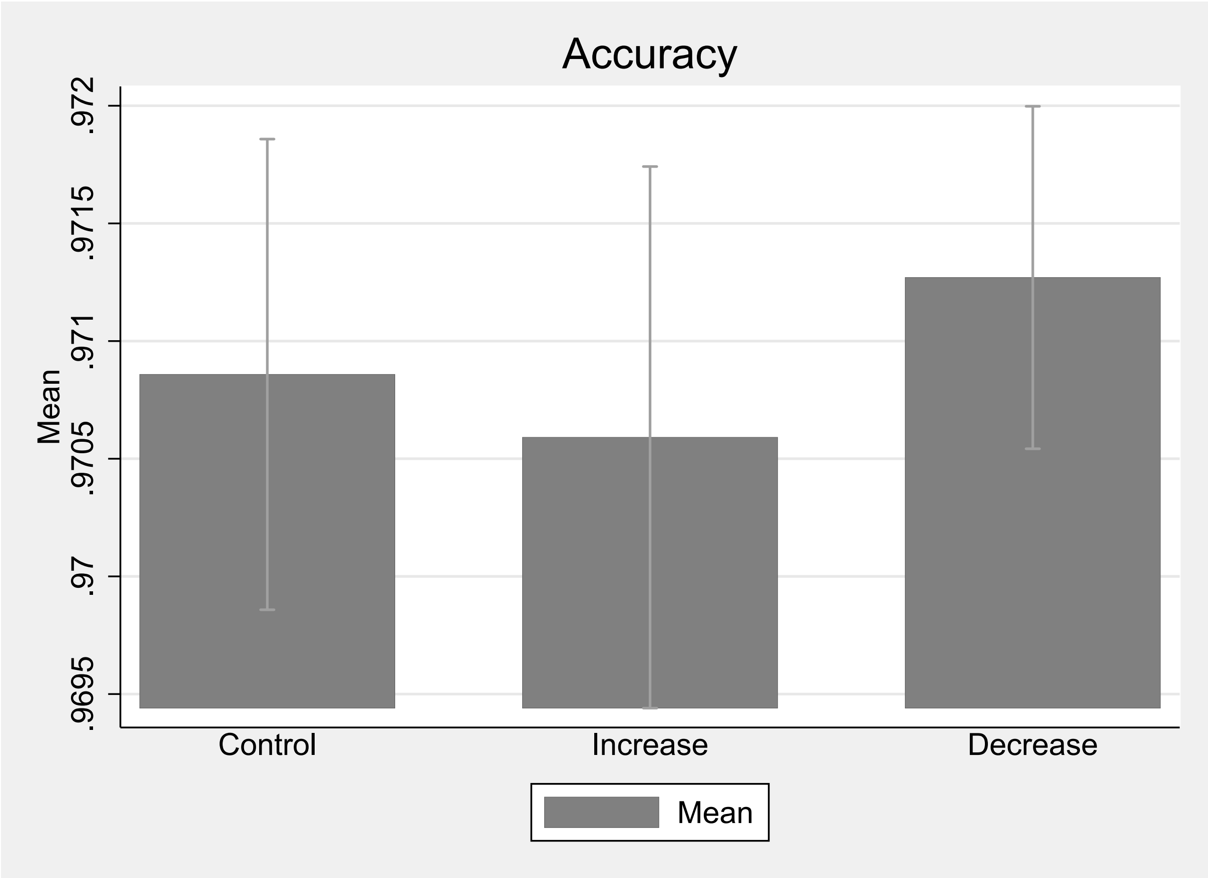
Notes: The Figure depicts for each group the average number of images that subjects transcribed, along with 95% confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the ”wage increase” group and 257 subjects in the ”wage decrease” group. All 720 subjects in the sample have transcribed at least six images.

Figure 13: Number of transcribed pics conditional on workers who completed at least one pic after the treatment notification



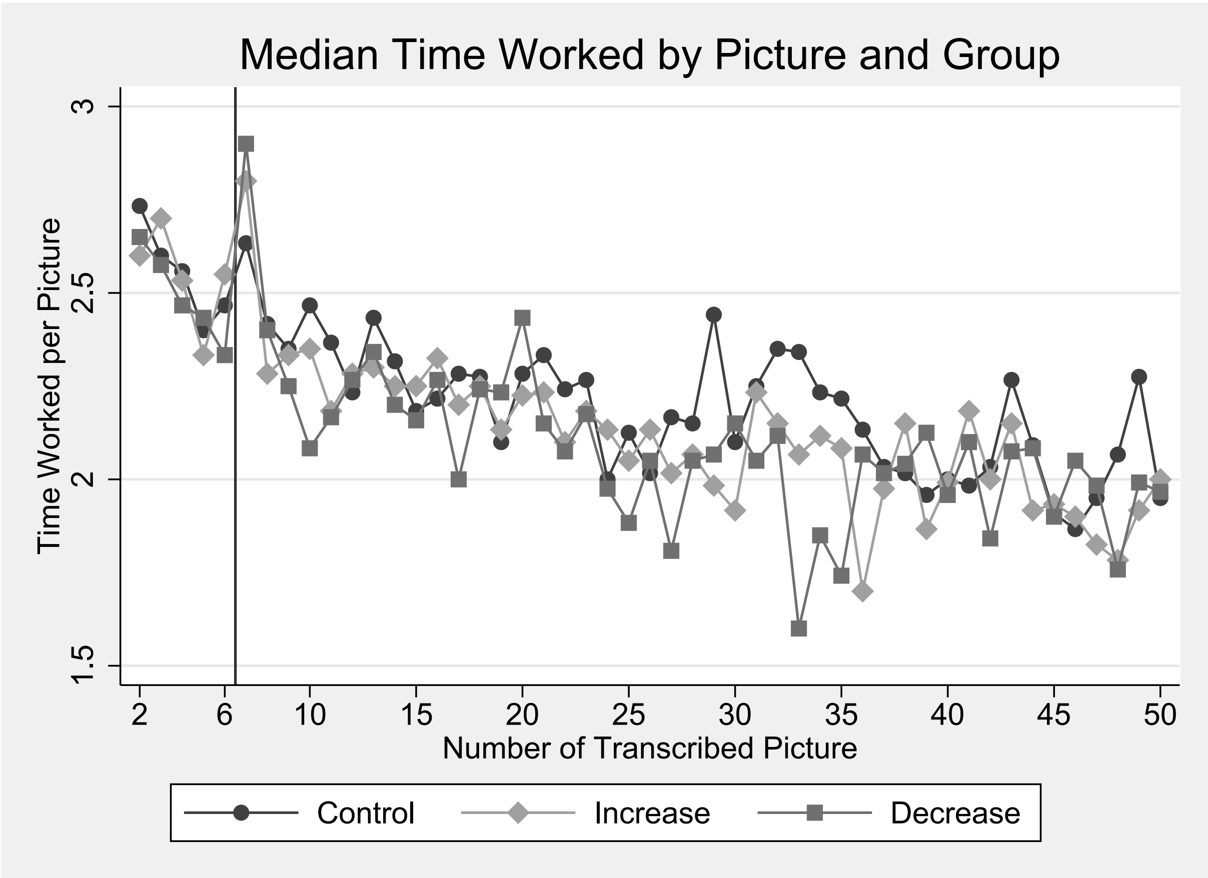
Notes: The Figure depicts for each group the average number of images that subjects transcribed, along with 95% confidence bands. The underlying sample is restricted to subjects who did *not* quit the labor task immediately after seeing the treatment notification (i.e., restricted to subjects who have transcribed at least seven images). The number of observations is 591 with 213 subjects in the control group, 203 subjects in the ”wage increase” group and 175 subjects in the ”wage decrease” group. All 591 subjects in the sample have transcribed at least seven images.

Figure 14: Accuracy by Treatment Group



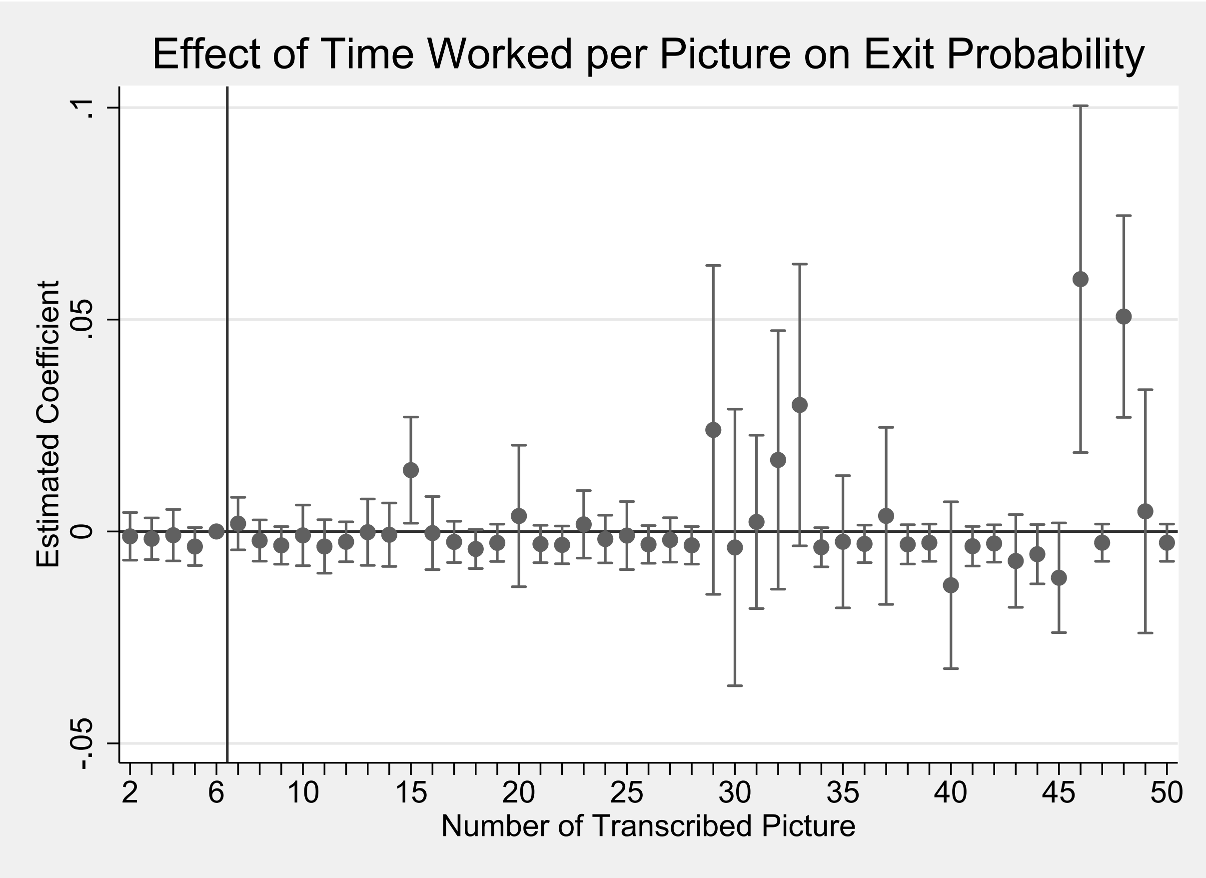
Notes: The Figure depicts for each group the average transcription accuracy, i.e., the average share of characters in each image that is transcribed correctly, along with 95% confidence bands. The number of observations is 720 with 248 subjects in the control group, 215 subjects in the ”wage increase” group and 257 subjects in the ”wage decrease” group. All 720 subjects in the sample have transcribed at least six images.

Figure 15: Time worked per pictures over periods, by experimental group



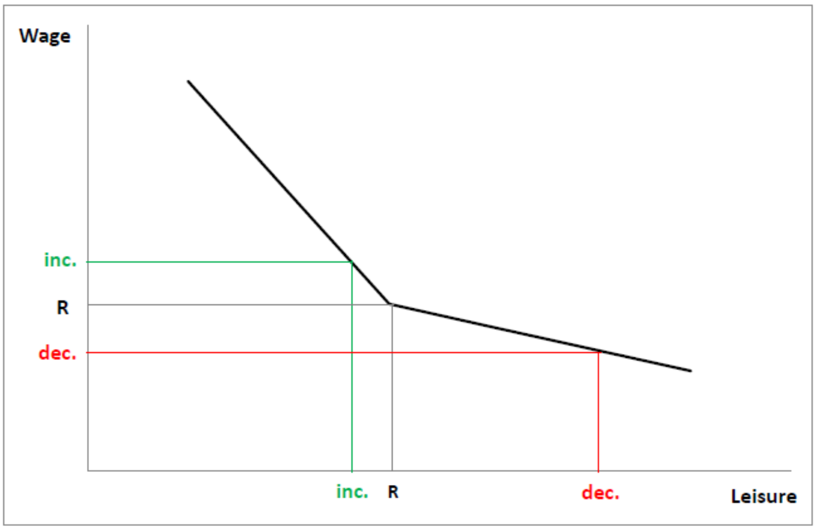
Notes: The Figure depicts the median time it takes to transcribe a picture in each period by experimental group. For example, the value in period 10 indicates the median time it took subjects in a given experimental group to transcribe the 10th picture. Treatment occurred after period 6. The sample includes all participants who started working on the task. Number of observations is 1152.

Figure 16: Effect of Performance on Probability to Quit



Notes: The Figure depicts coefficients which are based on a regression of a 'Last-picture-transcribed' dummy on a full interaction of period dummies and the time used to transcribe the respective picture. 'Last-picture-transcribed' dummy is '1' if the picture in this respective round is the last transcribed picture of the respective participant. Coefficients are shown along with 95% confidence bars. Standard error clustered by individual. Treatment occurred after period 6 and we normalize the coefficients to the pre-treatment period. The sample includes all participants who started working on the task. Number of observations is 1152.

Figure 17: Prediction: Labor Supply under Loss Aversion



Notes: The Figure displays the relationship between leisure and wages under loss aversion. The curve is kinked at the reference wage R. Individuals who currently face the reference wage respond stronger to a wage decrease (by supplying less labor/more leisure) than to a wage increase of equal magnitude (by supplying more labor/less leisure).

Tables

Table 1: Summary statistics: Pictures transcribed and Accuracy

variable	**N**	**mean**	**sd**	**p10**	**p50**	**p90**
Pics transcribed	1152	12.81	13.23	2.00	7.00	33.00
Total time	1152	39.79	50.25	3.17	19.64	104.35
Accuracy	1151	0.97	0.02	0.96	.97	0.98

Notes: Summary statistics for outcome variables. The sample is all subjects who started working on the task (i.e., including those who did not necessarily get to see the treatment notification after 6 transcribed pictures). *Pics transcribed* is the average number of images that subjects transcribed. *Total time* is the average time (in minutes) that subjects totally spent on working on the labor task. *Accuracy* the average share of characters that is transcribed correctly. N is the number of observations. sd is the standard deviation. pX indicates the X-th percentile.

Table 2: Number of Observations

	Seen Treatment		
Group	No	Yes	Total
Control	143	248	391
Increase	149	215	364
Decrease	140	257	397
Total	432	720	1152

Notes: Number of observations by treatment group who (i) started working on the task but did not see the treatment notification, i.e., they transcribed five images or less (Column *No*) and (ii) who transcribed at least six pictures and therefore saw the treatment notification (Column *Yes*).

Table 3: Regression estimates and implied elasticities

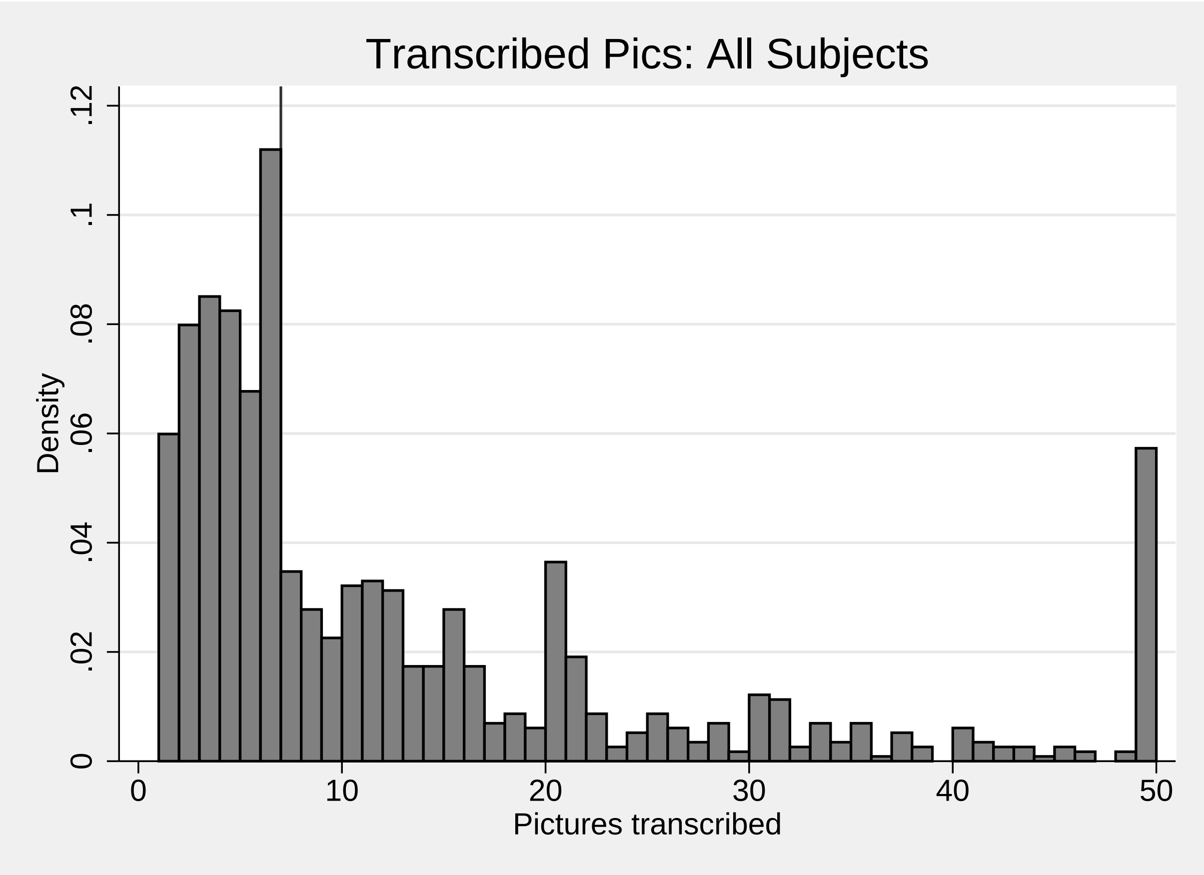
	Dependent Variable				
	I	II	III	IV	V
	Ext. margin	Pics	Total Time	Mean Time	Accuracy
Reference group:	Control				
Increase	-0.085***	3.309**	6.078	-0.344	-0.000
	(0.027)	(1.298)	(4.985)	(0.264)	(0.001)
Decrease	0.178***	-3.795***	-10.556**	0.175	0.000
	(0.037)	(1.166)	(4.832)	(0.358)	(0.001)
constant	0.141***	19.040***	60.916***	3.778***	0.971***
	(0.022)	(0.870)	(3.399)	(0.210)	(0.001)
N	720	720	720	720	719
R2	0.082	0.044	0.016	0.003	0.001
p($Inc = -Dec$)	0.094	0.820	0.596	0.752	0.906
elast increase	-3.02	0.87	0.50	-0.45	0
elast decrease	-6.30	0.99	0.87	-0.23	0

Notes: OLS regressions based on equation 1. Robust standard errors in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%. The explanatory variables of interest are dummies indicating the *Increase* and *Decrease* group, respectively. The coefficients are relative to the omitted *Control* group. The outcome variables in columns (I) to (V) are: (I) *Ext. margin* is the extensive margin measured as a dummy variable indicating whether a subject quit the task immediately after seeing the treatment notification. (II) *Pics* is the number of images transcribed. (III) *Total time* is the time (in minutes) that a subject totally spent on working on the labor task. (IV) *Mean Time* is the time (in minutes) that a subject spent working on one image. (V) *Accuracy* is the share of characters that is transcribed correctly. N is the number of observations. *R2* is R-squared. p($Inc = -Dec$) is the p-value from a t-test testing whether the coefficients from the coefficients for the Increase and Decrease group add up to zero. *elast increase* and *elast decrease* are the elasticities in the treatment group that indicate how the respective outcome variable responds to the wage change, using the control group as the counterfactual (see equation 3 in section 3.3).

Appendix

A Distribution of pictures for all workers

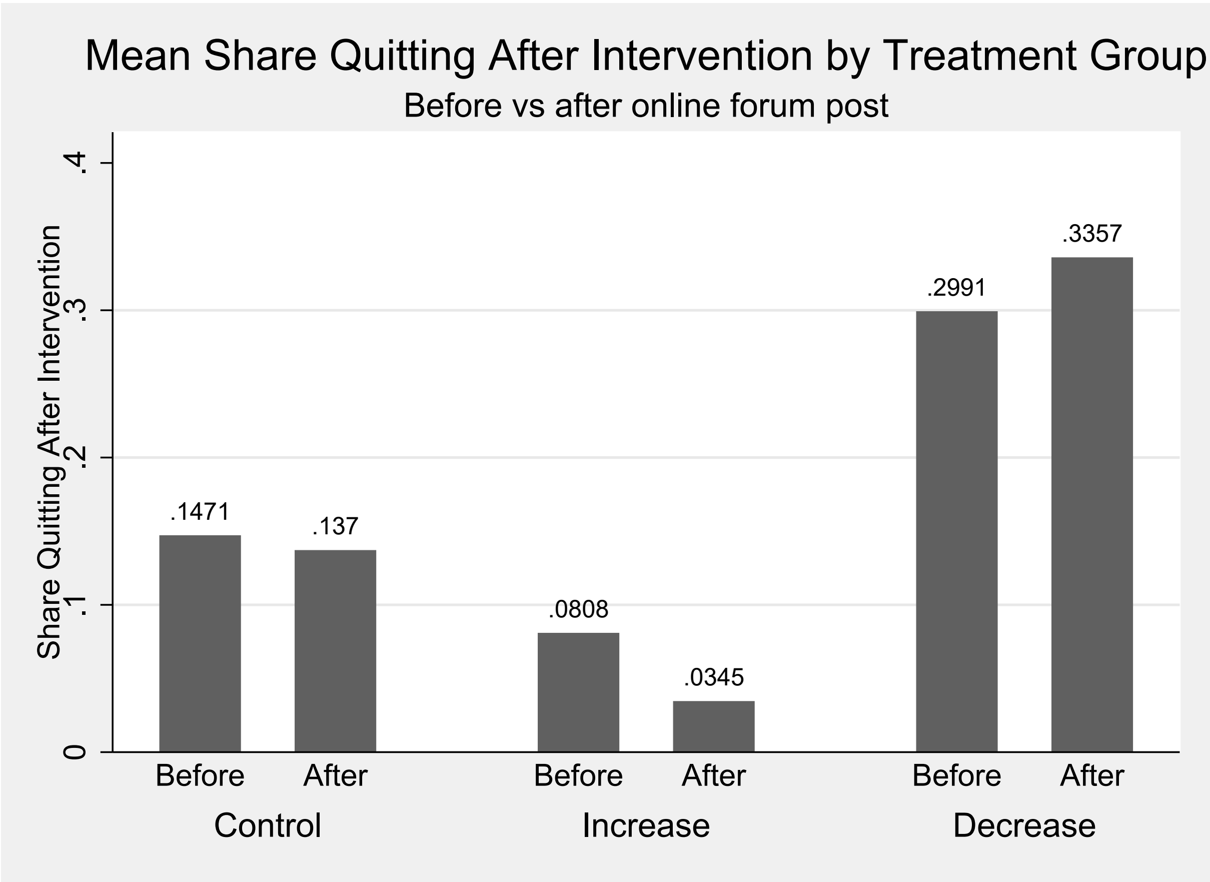
Figure 18: Histogramm of transcribed pictures



Notes: Histogram of pictures described for all workers who worked on the task. The number of observations is 1152. Subjects saw the treatment notification after transcribing 6 pictures (indicated by the vertical line).

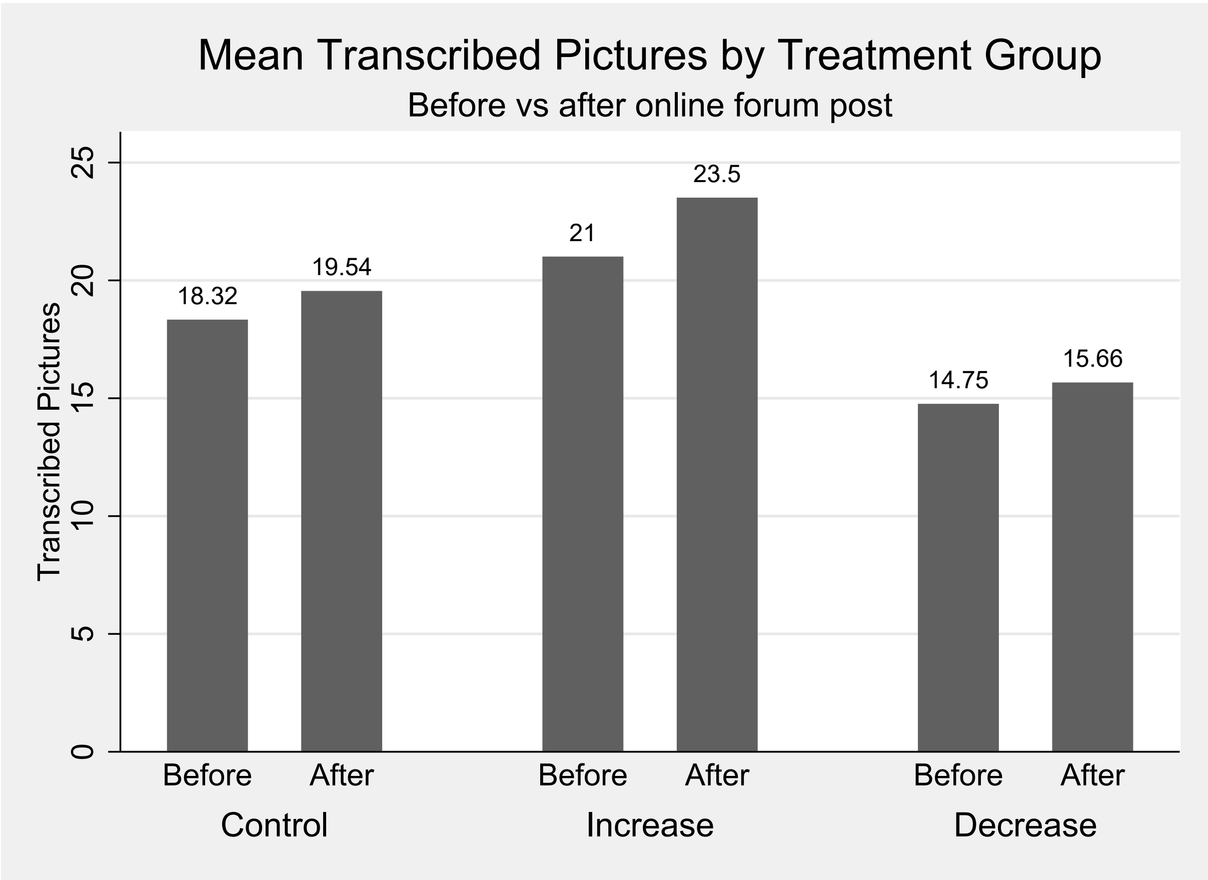
B Robustness: Effect of forum post

Figure 19: Extensive Margin. Before vs after forum post



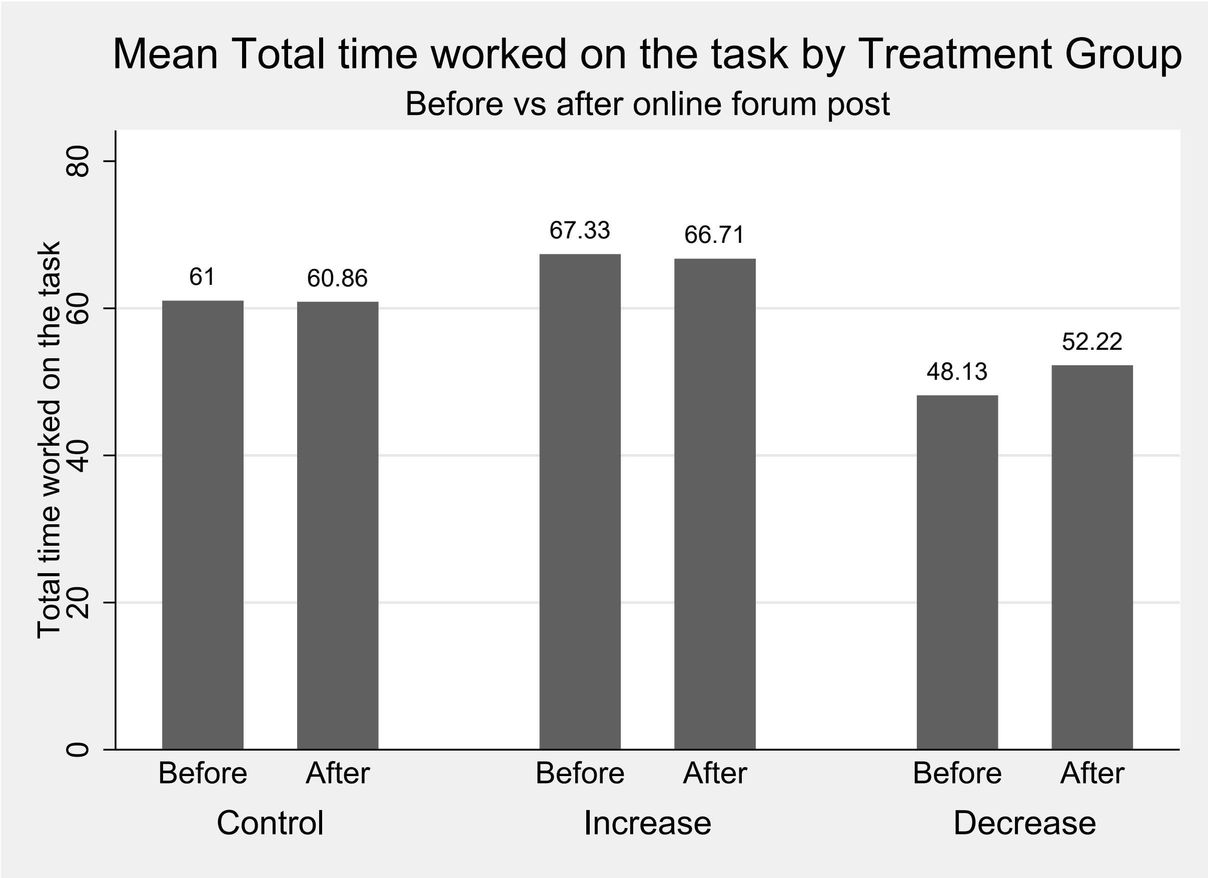
Notes: The Figure depicts the share of subjects in each group who quit the labor task immediately after seeing the treatment notification (i.e., share of subjects who transcribed six pictures but not a seventh one). *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see section3.2.) The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 20: Number of Transcribed Pics. Before vs after forum post



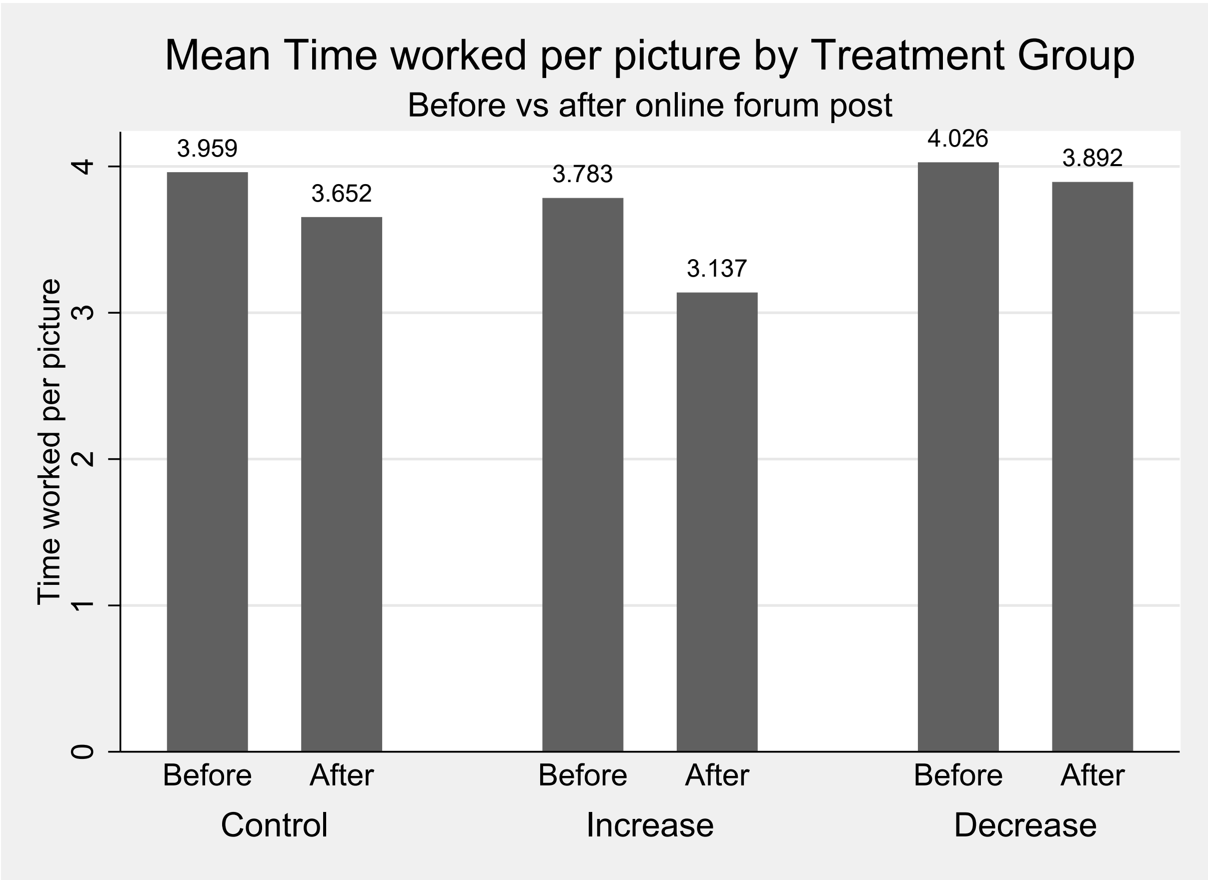
Notes: The Figure depicts for each group the average number of images that subjects transcribed. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see section3.2.) The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 21: Total Time Worked. Before vs after forum post



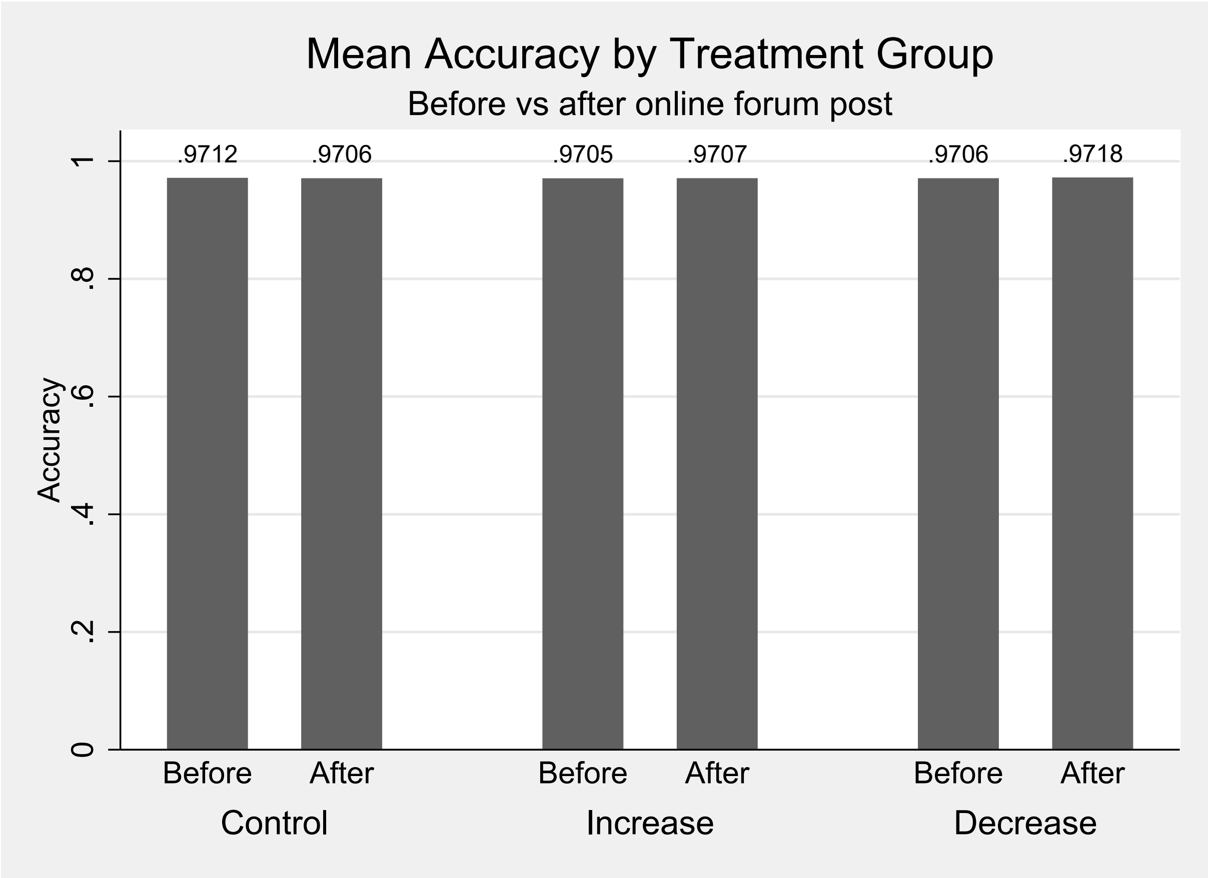
Notes:The Figure depicts for each group the average time (in minutes) that subjects totally spent on working on the labor task. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see section3.2.) The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 22: Avg. Time per Hit. Before vs after forum post



Notes: The Figure depicts for each group the average time (in minutes) that subjects spent working on one image. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see section3.2.) The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

Figure 23: Accuracy. Before vs after forum post



Notes: The Figure depicts for each group the average transcription accuracy, i.e., the average share of characters in each image that is transcribed correctly. *Before* and *After* indicate whether the observation was sampled before or after the task was discussed online (see section3.2.) The number of observations sampled before the forum post is 318 with 102 subjects in the control group, 99 subjects in the "wage increase" group and 117 subjects in the "wage decrease" group. The number of observations sampled after the forum post is 402 with 146 subjects in the control group, 116 subjects in the "wage increase" group and 140 subjects in the "wage decrease" group. All subjects in the sample have transcribed at least six images.

C A Model of Labor Supply under Loss Aversion

This section presents a theoretical framework that allows us to predict the impact of wage increases and decreases on labor-supply. The framework is informed by Ahrens et al. (2014) who incorporate loss aversion into a standard labor-supply model.

Ahrens et al. (2014) develop a model where workers with reference-dependent preferences maximize the following utility function:

$$U(C, L) = U^C(C) - \theta_i \frac{L^{\vartheta_i}}{\vartheta_i},$$

where C is consumption, L is labor supply (hours worked or effort), and θ_i is a parameter to ensure preference continuity at the reference wage. $U^C(C)$ is utility from consumption and the term $\frac{L^{\vartheta_i}}{\vartheta_i}$ indicates disutility from working. ϑ_i is a measure of loss aversion, which is characterized by the following piece-wise function:

$$\vartheta_i = \begin{cases} \vartheta_1 & \text{if } w > w^r \\ \vartheta_2 & \text{if } w < w^r. \end{cases}$$

In this equation, w is the current wage (per unit of L supplied) and w^r is the reference wage.[31] If w is above the reference wage, the worker is in the so-called gain domain, and if w is below the reference wage, she is in the loss domain. A subject is loss averse if $\vartheta_1 > \vartheta_2$, implying that the marginal utility loss from working is higher in the gain domain than in the loss domain. This means that workers are less willing to supply an additional unit of labor when the wage is above the reference wage than when it is below. Maximizing with respect to the budget constraint $C = wL$ gives the following kinked labor-supply curve:[32]

$$L = \begin{cases} (\frac{w}{\theta_1})^{\frac{1}{\vartheta_1 - 1}} & \text{if } w > w^r \\ (\frac{w}{\theta_2})^{\frac{1}{\vartheta_2 - 1}} & \text{if } w < w^r \end{cases}$$

Because of loss aversion with respect to the reference wage w^r (and hence $\vartheta_1 > \vartheta_2$), we get that $\frac{1}{\vartheta_1 - 1} < \frac{1}{\vartheta_2 - 1}$. This implies that subjects whose current wage is the reference wage w^r are more responsive to wage decreases than to wage increases.[33]

[31] It is plausible to argue that \$0.15 constitutes the reference wage w^r in our empirical set-up (see Section 5.1.).

[32] We only discuss the main implications of the model here since Ahrens et al. (2014) has all of the derivations.

[33] We assume an upward sloping labor supply curve where the substitution effect dominates the income effect. That is, subjects work more when wages go up and they work less when wages fall. This assumption is also supported by our empirical findings.