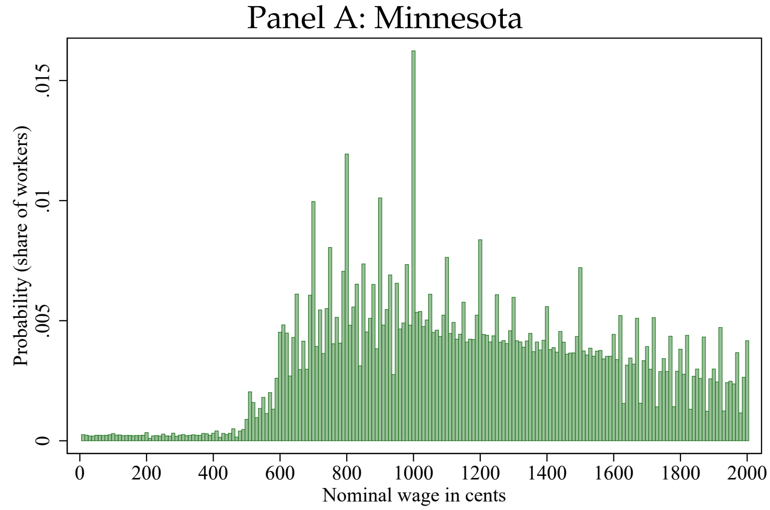
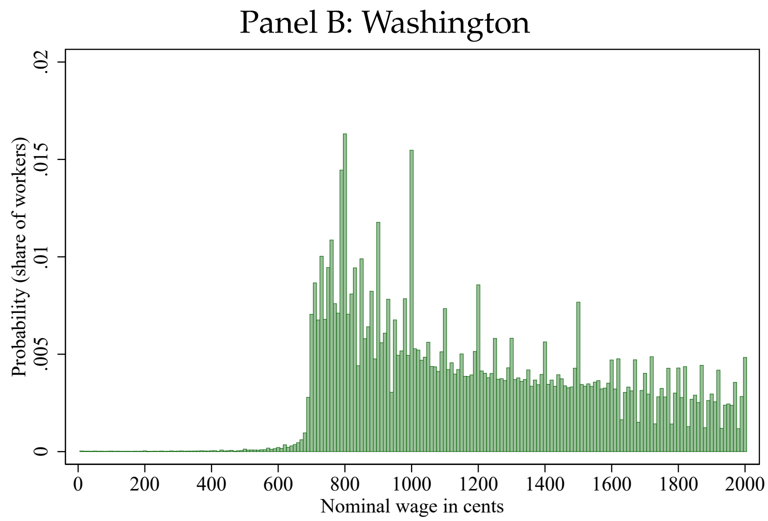
Online Appendix A

Additional Figures

Appendix Figure A.1 plots the histograms of hourly wages in (nominal) $0.10 bins using administrative data separately for the states of Minnesota (panel A) and Washington (panel B). Both are based on hourly wage data from UI records from 2003-2007. Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts are normalized by dividing by total employment in that state, averaged over the sample period. The figure shows very clear bunching at multiples of $1 in both states, especially at $10. Appendix Figure A.2 plots the overlaid histograms of hourly wages, pooled across both MN and WA, in real $0.10 bins from 2003q4 and 2007q4, and shows that the nominal bunching at $10.00 occurs at different places in the real wage distribution in 2003 and 2007.

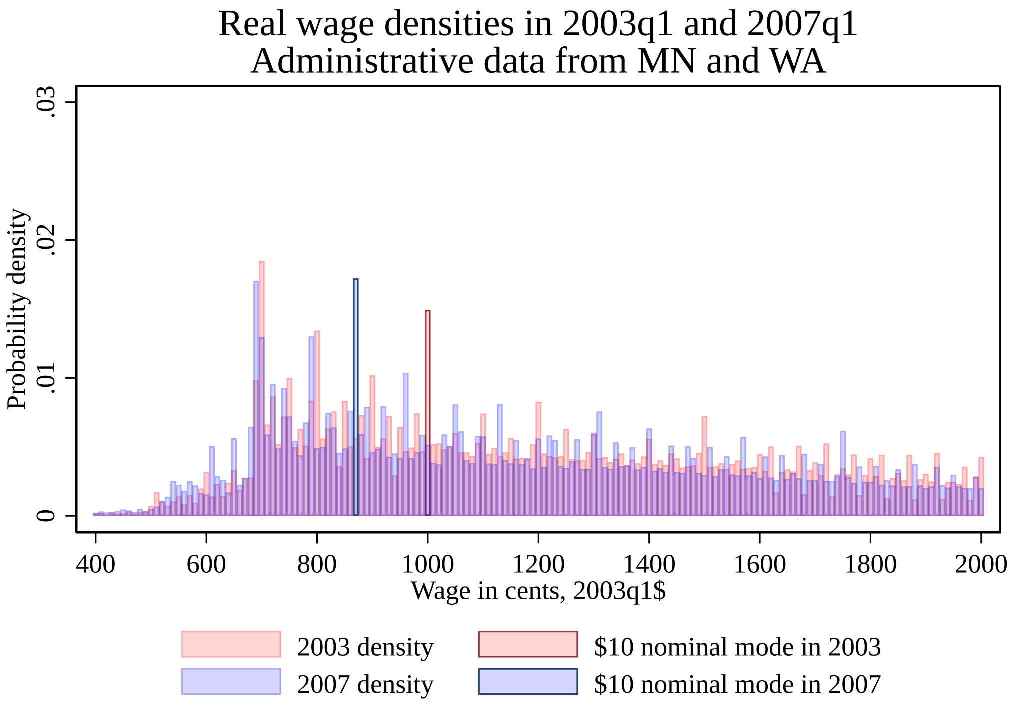
Figure A.1: **Histograms of Hourly Wages In Administrative Payroll Data from Minnesota and Washington, 2003-2007**





Notes. The figure shows histograms of hourly wages in $0.10 (nominal) wage bins, averaged over 2003q1 to 2007q4, using administrative Unemployment Insurance payroll records from the states of Minnesota (Panel A) and Washington (Panel B). Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts in each bin are normalized by dividing by total employment in that state, averaged over the sample period. The UI payroll records cover over 95% of all wage and salary civilian employment in the states. The counts here exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators as having substantial reporting errors.

Figure A.2: **Histograms of Real Hourly Wages In Administrative Payroll Data from Minnesota and Washington, 2003-2007**



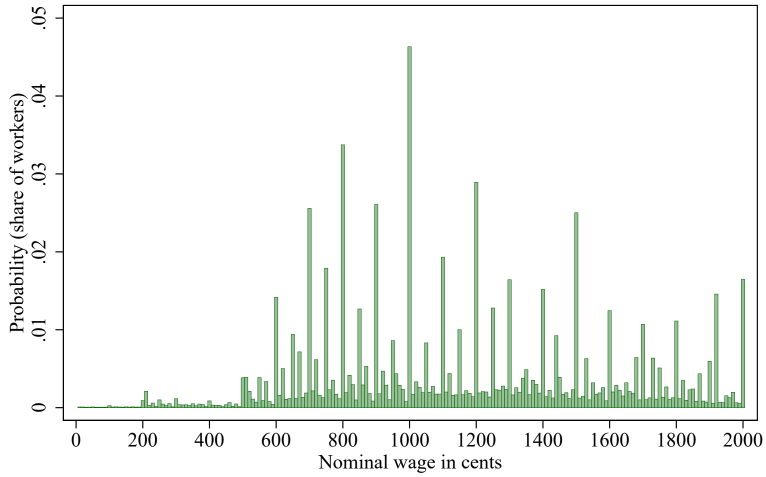
Notes. The figure shows a histogram of hourly wages in $0.10 real wage bins (2003q1 dollars) for 2003q1 and 2007q1, using pooled administrative Unemployment Insurance payroll records from the states of Minnesota and Washington. The nominal $10 bin is outlined in dark for each year—highlighting the fact that this nominal mode is at substantially different part of the real wage distributions in these two periods. Hourly wages are constructed by dividing quarterly earnings by the total number of hours worked in the quarter. The counts in each bin are normalized by dividing by total employment in that state for that quarter. The UI payroll records cover over 95% of all wage and salary civilian employment in the states. The counts here exclude NAICS 6241 and 814, home-health and household sectors, which were identified by the state data administrators as having substantial reporting errors.

Online Appendix B Bunching in Hourly Wage Data from Current Population Survey and Supplement

For comparison, we next show an analogous histogram of hourly nominal wage data using the national CPS data. In Figure B.3, we plot the nominal wage distribution in U.S. in 2003 to 2007 in $0.10 bins. There are notable spikes in the wage distribution at $10, $7.20 (the bin with the federal minimum wage), $12, $15, along with other whole numbers. At the same time, the spike at $10.00 is substantially larger than in the administrative data (exceeding 0.045), indicating rounding error in reporting may be a serious issue in using the CPS to accurately characterize the size of the bunching.

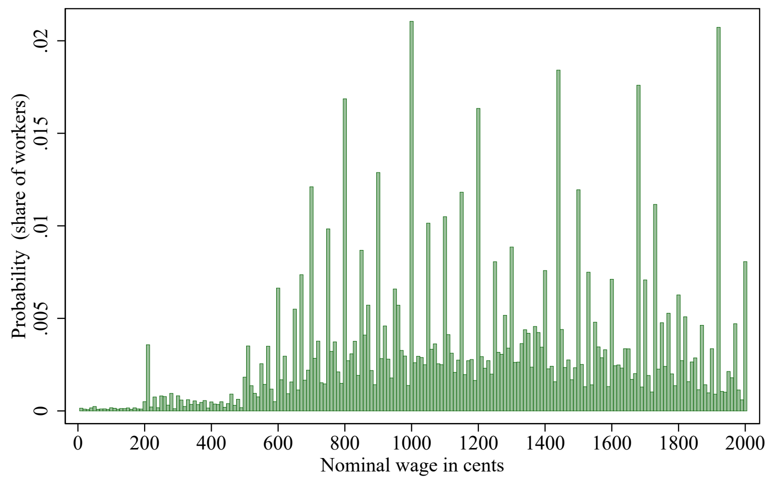
We also use a 1977 CPS supplement, which matches employer and employee reported hourly wages, to correct for possible reporting errors in the CPS data. We re-weight wages by the relative incidence of employer versus employee reporting, based on the two ending digits in cents (e.g., 01, 02, ... , 98, 99). As can be seen in Figure B.4, the measurement error correction produces some reduction in the extent of visible bunching, which nonetheless continues to be substantial. For comparison, the probability mass at $10.00 is around 0.02, which is closer to the mass in the administrative data than in the raw CPS. This is re-assuring as it suggests that a variety of ways of correcting for respondent rounding produce estimates suggesting a similar and substantial amount of bunching in the wage distribution.

Figure B.3: **Histogram of Hourly Wages in National CPS data, 2003-2007**



Notes. The figure shows a histogram of hourly wages by $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4, using CPS MORG files. Hourly wages are constructed by average weekly earnings by usual hours worked. The sample is restricted to those without imputed earnings. The counts here exclude NAICS 6241 and 814, home-health and household sectors. The histogram reports normalized counts in $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4. The counts in each bin are normalized by dividing by total employment, averaged over the sample period.

Figure B.4: **Wage Bunching in CPS data, 2003-2007, Corrected for Reporting Error Using 1977 CPS supplement**



Notes. The figure shows a histogram of hourly wages by $0.10 (nominal) wage bins, averaged over 2003q1 to 2007q4, using CPS MORG files, where individual observations were re-weighted to correct for overreporting of wages ending in particular two-digit cents using the 1977 CPS supplement. Hourly wages are constructed by dividing average weekly earnings by usual hours worked. The sample is restricted to those without imputed earnings. The counts here exclude NAICS 6241 and 814, home-health and household sectors. The histogram reports normalized counts in $0.10 (nominal) wage bins, averaged over 2003q1 and 2007q4. The counts in each bin are normalized by dividing by total employment, averaged over the sample period.

Online Appendix C

Testing Discontinuous Labor Supply on Amazon Mechanical Turk Observational Data

Our Amazon Mechanical Turk experiment focused on discontinuities at 10 cents, while our bunching estimator used the excess mass at \$1.00. In this appendix we present evidence from observational data scraped from Amazon Mechanical Turk to show that there is also no evidence of a discontinuity in worker response to rewards at \$1.00. Our primary source of data was collected by Panos Ipseiros between January 2014 and February 2016, and, in principle, kept track of all HITs posted in this period.

We keep the discussion of the data and estimation details brief, as interested readers can see details in Dube et al. (2018). Dube et al. (2018) combines a meta-analysis of experimental estimates of the elasticity of labor supply facing requesters on Amazon Mechanical Turk with Double-ML estimators applied to observational data.. That paper does not look at discontinuities in the labor supply at round numbers.

Following Dube et al. (2018) we use the observed duration of a batch posting as a measure of how attractive a given task is as a function of observed rewards and observed characteristics. We calculate the duration of the task as the difference between the first time it appears and the last time it appears, treating those that are present for the whole period as missing values. We convert the reward into cents. We are interested in the labor supply curve facing a requester. Unfortunately, we do not see individual Turkers in this data. Instead we calculate the time until the task disappears from our sample as a function of the wage. Tasks disappear once they are accepted. While tasks may disappear due to requesters canceling them rather than being filled, this is rare. Therefore, we take the time until the task disappears to be the duration of the posting—i.e., the time it takes for the task to be accepted by a Turker. The elasticity of this duration with respect to the wage will be equivalent to the elasticity of labor supply when offer arrival rates are constant and reservation wages have an exponential (constant hazard) distribution. We estimate regressions of the form:

$$\ln(duration_h) = \beta \times \ln(reward_h) + \delta_{requester} + \delta_{hourposted} + \epsilon$$

Where h indexes HIT batches,and the specification includes requester fixed effects and fixed effects for the hour the HIT batch is first posted. We also show specifications that add keyword combination fixed effects (the keywords allow Turkers to look for particular tasks), log of the initial number of HITs in the batch, and log of time allotted by the requester. This will almost always be an overestimate of the actual time taken to complete the task, but is likely correlated with it. Note that time allotted is also how much time a Turker has to do the task, and if the task is too long relative to the time allotted, it may expire before the Turker can complete the task. Hence short time allotted does not necessarily imply the task is shorter, and Turkers may be averse to tasks that have too little time allotted.

Results are shown in Table C.1. There is a clear negative relationship between rewards and duration. If the distribution of reservation wages has a constant hazard and the rate at which offers are received is constant, this implies an upward sloping labor supply curve with a very low elasticity (< 1), but still considerably larger than our experimental estimate on MTurk.[20] We also show analogues of our experimental specifications from our pre-analysis plan. The first approach tests for a discontinuity by adding an indicator for rewards greater than or equal to 100 ("Jump at 100"). This level discontinuity is tested in specifications 3 and 4, and there is no evidence of log durations becoming discontinously larger above $1.00. The second approach tests for a slope break at $1.00 by estimating a knotted spline that allows the elasticity to vary between 51 and 95 cents, 95 cents and $1.00,

[20]In Dube et al. (2018) we implement a more comprehensive adjustment for unobserved heterogeneity using a double-machine-learning estimator proposed by Chernozhukov et al. (2017); this yields much smaller labor supply elasticities relative to the fixed-effects specifications, and very close, not only to our experimental estimates presented above, but also to the precision-weighted mean calculated from a number of other experimentally estimated elasticities.

and then greater than \$1.00. The slope break specification is tested in specifications 5 and 6, where we report the change in slopes at \$1.00 ("Spline"). Again, there is no evidence of a change in the relationship between log duration and log reward between \$0.95 and \$1.00 versus greater than \$1.00.

Table C.1: **Duration of Task Posting by Log Reward and Jump at $1.00**

	(1)	(2)	(3)	(4)	(5)	(6)
Log Reward	-0.663***	-0.842***	-0.689**	-0.938***	-0.632*	-0.976**
	(0.171)	(0.210)	(0.274)	(0.338)	(0.329)	(0.405)
Jump at 100			0.015	0.058		
			(0.116)	(0.165)		
Spline					-0.243	0.287
					(2.361)	(3.347)
Additional controls:						
Requester x Source FE	Y	Y	Y	Y	Y	Y
Hour Posted FE	Y	Y	Y	Y	Y	Y
Keyword FE	N	Y	N	Y	N	Y
Log Initial HITs	N	Y	N	Y	N	Y
Log Time Alloted	N	Y	N	Y	N	Y
Sample size	22,097	15,684	22,097	15,684	22,097	15,684

Notes. Sample is restricted to HIT batches with rewards between 51 and 149 cents. Columns 3, 4 and 8 estimate a specification testing for a discontinuity in the duration at $1.00, as in our pre-analysis plan, while columns 5 and 6 estimate the spline specification testing for a change in the slope of the log duration log reward relationship at $1.00, also from the pre-analysis plan. Significance levels are * 0.10, ** 0.05, *** 0.01.

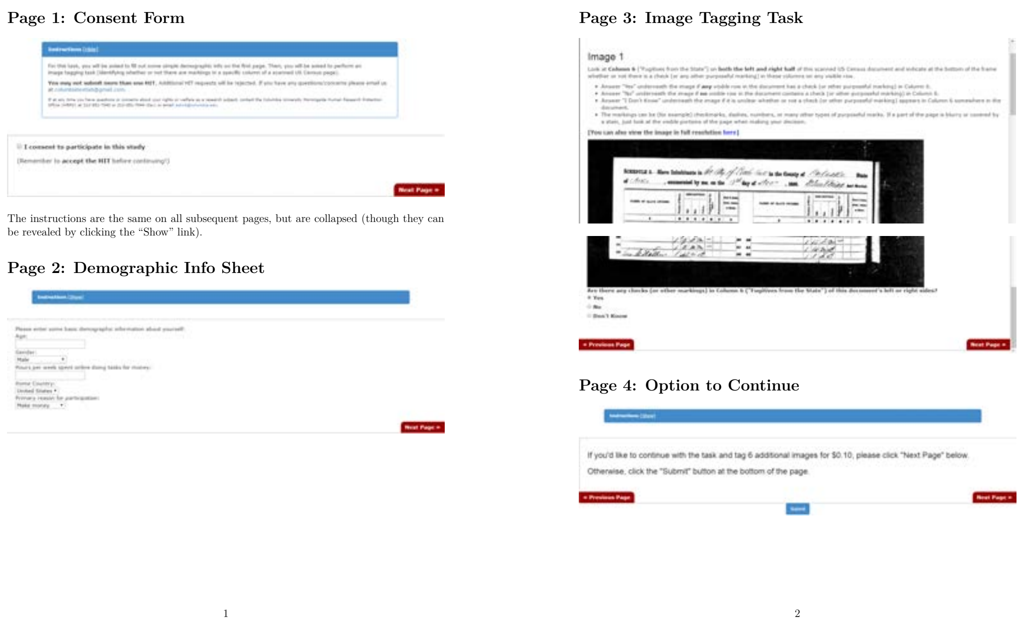
Online Appendix D

Additional Experimental Details and Specifications from Pre-analysis Plan

Figure D.1 shows screenshots from the experimental layout facing MTurk subjects. while D.3 shows specifications parallel to those from the main text, except with the number correct as the outcome, to measure responsiveness of subject effort to incentives. There is no evidence of any effect of higher rewards on the number of images labelled.

In Tables D.1 and D.2 we show specifications from our pre-analysis plan that parallel those in 7 and D.3, respectively. These were linear probability specifications in the level of wages without any controls, instead of the logit specifications with log wages and controls we show in the main text. We also pool the two different task volumes. The initial focus of our experiment was to test for a discontinuity at 10 cents, which is unaffected by our changes in specification. While the elasticity is qualitatively very similar, the logit-log wage specification shown in the text is closer to our model, a variant of the model specified by Card et al. (2016), and improves precision on the elasticity estimate.

Figure D.1: **Online Labor Supply Experiment on MTurk**



Notes. The figure shows the screen shots for the consent form and tasks associated with the online labor supply experiment on MTurk.

Table D.1: **Preanalysis Specifications: Task Acceptance Probability by Offered Task Reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wage	0.004	0.008	0.004		0.001	0.003	-0.003		0.008*	0.013	0.011*	
	(0.003)	(0.006)	(0.004)		(0.004)	(0.008)	(0.005)		(0.004)	(0.009)	(0.006)	
Jump at 10			0.001				0.022				-0.021	
			(0.016)				(0.022)				(0.025)	
Spline		-0.002				0.193				-0.205		
		(0.156)				(0.206)				(0.236)		
Local				0.010				0.015				0.004
				(0.023)				(0.031)				(0.034)
Global				-0.000				0.011				-0.012
				(0.015)				(0.020)				(0.023)
η	0.052	0.090	0.050		0.015	0.035	-0.029		0.095*	0.157	0.140*	
	(0.033)	(0.071)	(0.048)		(0.042)	(0.087)	(0.062)		(0.051)	(0.116)	(0.073)	
Sample	Pooled	Pooled	Pooled	Pooled	6 HITs	6 HITs	6 HITs	6 HITs	12 HITs	12 HITs	12 HITs	12 HITs
Sample Size	5184	5184	5184	5184	2683	2683	2683	2683	2501	2501	2501	2501

Notes. The reported estimates are linear regressions of task acceptance probabilities on log wages, controlling for number of images. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the difference in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-8 repeat 1-4, but restrict the sample to "sophisticates": Turkers who respond that they work more than 10 hours a week and their primary motivation is money. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table D.2: **Preanalysis Specifications: Task Correct Probability by Offered Task Reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Wage	-0.001	0.001	-0.001		0.001	0.006*	0.001		-0.003**	-0.005	-0.003	
	(0.001)	(0.003)	(0.002)		(0.002)	(0.004)	(0.003)		(0.002)	(0.003)	(0.002)	
Jump at 10			-0.000				0.000				-0.002	
			(0.007)				(0.011)				(0.009)	
Spline		-0.012				-0.013				-0.008		
		(0.067)				(0.101)				(0.087)		
Local				0.003				-0.000				0.006
				(0.012)				(0.018)				(0.015)
Global				-0.003				-0.007				0.000
				(0.007)				(0.009)				(0.009)
η	-0.009	0.014	-0.008		0.008	0.060*	0.008		-0.029**	-0.047*	-0.026	
	(0.013)	(0.024)	(0.018)		(0.020)	(0.036)	(0.029)		(0.015)	(0.028)	(0.023)	
Sample	Pooled	Pooled	Pooled	Pooled	6 HITs	6 HITs	6 HITs	6 HITs	12 HITs	12 HITs	12 HITs	12 HITs
Sample Size	5184	5184	5184	5184	2683	2683	2683	2683	2501	2501	2501	2501

Notes. The reported estimates are linear regressions of task acceptance probabilities on log wages, controlling for number of images. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the difference in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-8 repeat 1-4, but restrict the sample to "sophisticates": Turkers who respond that they work more than 10 hours a week and their primary motivation is money. Robust standard errors in parentheses.
* $p < 0.10$, ** $p < 0.5$, *** $p < 0.01$

Table D.3: **Task quality by offered task reward on MTurk**

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Log Wage	-0.006	-0.002	0.011		0.001	0.011	0.031	
	(0.012)	(0.017)	(0.017)		(0.022)	(0.033)	(0.034)	
Jump at 10		-0.002				-0.006		
		(0.007)				(0.013)		
Spline			-0.019				-0.052	
			(0.067)				(0.127)	
Local				0.003				0.012
				(0.011)				(0.022)
Global				-0.003				-0.002
				(0.006)				(0.012)
η	-0.006	-0.002	0.011		0.001	0.011	0.032	
	(0.012)	(0.017)	(0.017)		(0.023)	(0.034)	(0.035)	
Sample	Pooled	Pooled	Pooled	Pooled	Sophist.	Sophist.	Sophist.	Sophist.
Sample Size	4073	4073	4073	4031	1407	1407	1407	1396

Notes. The reported estimates are logit regressions of getting at least 1 out of 2 images correctly tagged on log wages (conditional on accepting the task), controlling for number of images done in the task (6 or 12), age, gender, weekly hours worked on MTurk, country (India/US/other), reason for MTurk, and an indicator for HIT accepted after pre-registered close date. Column 1 reports specification 1 that estimates the labor-supply elasticity, without a discontinuity. Column 2 estimates specification 2, which tests for a jump in the probability of acceptance at 10 cents. Column 3 estimates a knotted spline in log wages, with a knot at 10 cents, and reports the difference in elasticities above and below 10 cents. Column 4 estimates specification 4, including indicator variables for every wage and testing whether the different in acceptance probabilities between 10 and 9 cents is different from the average difference between 12 and 8 (local) or the average difference between 5 and 15 (global). Columns 5-8 repeat 1-4, but restrict the sample to "sophisticates": Turkers who respond that they work more than 10 hours a week and their primary motivation is money. Robust standard errors in parentheses.
* *p < 0.10, ** p < 0.5, *** p < 0.01*

Online Appendix E

Theoretical extension: An efficiency wage interpretation where effort depends on wage

In the main paper, we assume that the firm's ability to set wages comes from monopsony power. However, it may be recasted in terms of efficiency wages where wage affects productivity: there, too, the employer will set wages optimally such that the impact of a small change in wages around the optimum is approximately zero. In this section, we show a very similar logic applies in an efficiency wage model with identical observational implications as our monopsony model, with a re-interpretation of the labor supply elasticity η as capturing the rate at which the wage has to increase to ensure that the no-shirking condition holds when the firm wishes to hire more workers. Indeed, the observation that the costs of optimization errors are limited when wages are a choice variable was originally made by Akerlof and Yellen (1985) in the context of an efficiency wage model.

As in Shapiro and Stiglitz (1984), workers choose whether to work or shirk. Working entails an additional effort cost e. Following Rebitzer and Taylor (1995), we allow the detection of shirking, $D(l)$, to fall in the amount of employment $l(w)$.[21] Workers quit with an exogenous rate q. An unemployed worker receives benefit b and finds an offer at rate s. The discount rate is r. All wage offers are assumed to be worth accepting; once we characterize the wage setting mechanism, this implies a bound for the lowest productivity firm. Finally, generalizing both Rebitzer and Taylor (1995) and Shapiro and Stiglitz (1984),

[21] In Shapiro and Stiglitz (1984), the detection probability is exogenously set. This produces some predictions which are rather strong. For example, the model does not predict wages to vary with productivity, as the no shirking condition that pins down the optimal wage does not depend on firm productivity. The same is true for the Solow model, where the Solow condition is independent of firm productivity (see Solow 1979). As a result, those models cannot readily explain wage dispersion that is independent of skill distribution, which makes it less attractive to explain bunching. However, if we generalize the Shapiro-Stiglitz model to allow the detection probability to depend on the size of the workforce as in Rebitzer and Taylor (1995), this produces a link between productivity, firm size and wages. Going beyond Rebitzer and Taylor, we further generalize the model to allow for heterogeneity in firm productivity, which produces a non-degenerate equilbrium offer wage distribution.

we allow the wages offered by firms to vary; indeed our model will predict that higher productivity firms will pay higher wages—leading to equilibrium wage dispersion.

We can write the expected value of not shirking as:

$$V^N(w) = w - e + \frac{(1-q)V^N(w)}{1+r} + \frac{qV^U}{1+r}$$

The value of shirking can be written as:

$$V^S(w) = w + \frac{(1-q)(1-D)V^S(w)}{(1+r)} + \frac{(1-(1-q)(1-D))\,V^U}{(1+r)}$$

Finally, the value of being unemployed is:

$$V^U = b + \frac{sEV^N + (1-s)V^U}{(1+r)}$$

The (binding) no shirking condition, NSC, can be written as:

$$V^N(w) = V^S(w)$$

Plugging in the expressions above and simplifying we get the no-shirking condition:

$$w = \frac{r}{1+r}V^U + \frac{e(r+q)}{D(l)(1-q)}$$

We can further express V^U as a function of the expected value of an offer $V^N$and the probability of receiving an offer, s, as well as the unemployment benefit b. However, for our purposes, the key point is that this value is independent of the wage w and is taken to be exogenous by the firm in its wage setting. Since detection probability $D(l)$ is falling in l, we can now write:

$$D(l) = \frac{e\,(r+q)}{\left(w - e + \frac{1}{1+r}V^U\right)(1-q)}$$

This generates a relationship between l and w:

$$l(w) = D^{-1}\left(\frac{e\,(r+q)}{\left(w - e + \frac{1}{1+r}V^U\right)(1-q)}\right) = d\left(\frac{\left(w - e + \frac{1}{1+r}V^U\right)(1-q)}{e\,(r+q)}\right)$$

where $d(x) = D^{-1}(\frac{1}{x})$. Since $D'(x) < 0$, we have $d'(x) > 0$. This is analogous to the labor supply function facing the firm: to attract more workers who will work, one needs to pay a higher wage because detection is decling in employment, $D'(l) < 0$. Therefore, we can write the elasticity of the implicit labor supply function as:

$$\frac{l'(w)w}{l(w)} = \frac{d'(.)w}{d(.)} \times \frac{1-q}{e(r+q)}$$

.

If we assume a constant elasticity $d(x)$ function with elasticity ρ then the implicit "effective labor" supply elasticity is also constant:

$$\eta = \frac{l'(w)w}{l(w)} = \rho \times \frac{1-q}{e(r+q)}$$

The elasticity is falling in effort cost e, exogenous quit rate q, as well as the discount rate, r. It is also rising in the elasticity ρ, since a higher ρ means detection does not fall as rapidly with employment.

The implicit effective labor supply function is then:

$$l(w) = \frac{w^\eta}{C} = \frac{w^{\rho \times \frac{1-q}{e(r+q)}}}{C}$$

which is identical to the monopsony case analyzed in the main text. For a firm with productivity p_i, profit maximization implies setting the marginal cost of labor to the marginal revenue product of labor (p_i), i.e., $w_i = \frac{\eta}{1+\eta}p_i$. [22]

[22] We can also solve for $V^N = \frac{(E(w)-e)(1+r)}{r-b(1+r)} = \frac{\left(\frac{\eta}{1+\eta}E(p)-e\right)(1+r)}{r-b(1+r)}$. This implies we can write the equilibrium value of being unemployed as a function of the primitive parameters as follows: $V^U =$

Finally, we can augment this labor supply function to exhibit left-digit bias. Consider the case where for wage $w \geq w_0$, the wage is perceived to be to equal to $\tilde{w} = w + g$ while under w_0 it is perceived to be $\tilde{w} = w$. Now, the labor supply can be written as:
$l(w) = D^{-1}\left(\frac{e(r+q)}{\left(w-e+\frac{1}{1+r}V^U\right)(1-q)}\right) = d\left(\frac{\left(w-e+\frac{1}{1+r}V^U\right)(1-q)}{e(r+q)}\right)$ for $w < w_0$
$l(w) = D^{-1}\left(\frac{e(r+q)}{\left(w+g-e+\frac{1}{1+r}V^U\right)(1-q)}\right) = d\left(\frac{\left(w+g-e+\frac{1}{1+r}V^U\right)(1-q)}{e(r+q)}\right)$ for $w \geq w_0$
Note that under the condition that $d(x)$ has a constant elasticity, the implicit labor supply elasticity continues to be constant both below and above w_0. However, there is a discontinuous jump up in the $l(w)$ function at w_0. Therefore, we can always appropriately choose a γ such that this implicit labor supply function can be written as:

$$l(w_j, \gamma) = \frac{w^{\eta} \times \gamma^{\mathbb{1}_{w_j \geq w_0}}}{C} = \frac{w^{\rho \times \frac{1-q}{e(r+q)}} \times \gamma^{\mathbb{1}_{w_j \geq w_0}}}{C}$$

Facing this implicit labor supply condition, firms will optimize:

$$\Pi(p, w) = (p - w)l(w, \gamma) + D(p)\mathbf{1}_{w=w_0}$$

With a distribution of productivity, p, higher productivity firms will choose to pay more, as the marginal cost of labor implied by the implicit labor supply function is equated with the marginal revenue product of labor at a higher wage. Intuitively, higher productivity firms want to hire more workers. But since detection of shirking falls with size, this requires them to pay a higher wage to ensure that the no shirking condition holds. Similarly, all of the analysis of firm-side optimization frictions goes through here as well. A low η due to (say) high cost of effort now implies that a large amount of bunching at w_0 can be consistent with a small amount of optimization frictions, δ.

One consequence of this observational equivalence is that we cannot distinguish between efficiency wages and monopsony in our observational analysis. However, in our experimental analysis, we find that the evidence from online labor markets is more consis-

$(1+r)\left[\frac{b}{r+s} - \frac{e}{1-b(1+r} + \frac{\eta E(p)}{(1+\eta)(r-b(1+r)}\right]$

tent with a monopsony interpretation than an effort one, due to the absence of any effect of wages on the number of images tagged correctly. At the same time, it is useful to note that many of the implications from this efficiency wage model are quite similar to a monopsony one: for instance, both imply that minimum wages may increase employment in equilibrium, as Rebitzer and Taylor show. Therefore, while understanding the importance of specific channels is useful, the practical consequences may be less than what may appear at first blush.

Online Appendix F

Deconvolution estimator

In this appendix, we describe the deconvolution estimator we use to estimate the distributions of the the elasticity η and δ. Recall that if we condition on $\delta > 0$,we can take logs of equation 15 to obtain:

$$2\ln(\omega) = -\ln(\eta(1+\eta) + \ln(\delta) = -\ln(\eta(1+\eta)) + E[\ln(\delta)\,|\delta > 0] + \ln(\delta_{res})$$

We make the assumption that δ_{res} is lognormally distributed, so that $\ln(\delta_{res}) \sim N(0, \sigma_\delta^2)$, and we fix $E[\ln(\delta)\,|\delta > 0] = \ln(E(\delta|\delta > 0)) + \frac{1}{2}\sigma_\delta^2$. We can use the fact that the cumulative distribution function of $2\ln(\omega)$ is given by $1 - \hat{\phi}(\exp\{2\ln(\omega)\})$ to numerically obtain a density for $2\ln(\omega)$, where $\hat{\phi}$ is empirically estimated from the shape of the missing mass. This then becomes a well-known deconvolution problem, as the density of $-\ln(\eta(1+\eta))$ is the deconvolution of the density of $2\ln(\omega)$ by the Normal density we have imposed on $\ln(\delta_{res})$. We can then recover the distribution of η,$H(\eta)$, from the estimated density of $-\ln(\eta(1+\eta))$.

To see this, consider the general case of when the observed signal (W) is the sum of the true signal (X) and noise (U). (In our case $W = 2\ln(\omega) - E[\ln(\delta)\,|\delta > 0]$ and $U = \ln(\delta_{res})$.)

$$W = X + U$$

Manipulation of characteristic functions implies that the density of W is $f_W(x) = (f_X * f_U)(x) = \int f_X(x-y) f_U(y) dy$ where $*$ is the convolution operator. Let W_j be the observed sample from W.

Taking the Fourier transform (denoted by $\sim$) , we get that $\tilde{f_W} = \int f_W(x) e^{itx} dx = \tilde{f_X} \times \tilde{f_U}$. To recover the distribution of X, in principle it is enough to take the inverse Fourier transform of $\frac{\tilde{f_W}}{\tilde{f_U}}$. This produces a "naive" estimator $\widehat{f_X} = \frac{1}{2\pi} \int e^{-itx} \frac{\sum_{j=1}^{N} \frac{e^{itWj}}{N}}{\phi(t)} dt$, but unfortunately this is not guaranteed to converge to a well-behaved density function. To obtain such a density, some smoothing is needed, suggesting the following deconvolution estimator:

$$\widehat{f_X} = \frac{1}{2\pi} \int e^{-itx} K(th) \frac{\sum_{j=1}^{N} \frac{e^{itW_j}}{N}}{\phi(t)} dt$$

where K is a suitably chosen kernel function (whose Fourier transform is bounded and compactly supported). The finite sample properties of this estimator depend on the choice of f_U. If $\tilde{f_U}$ decays quickly (exponentially) with t (e.g. U is normal), then convergence occurs much more slowly than if $\tilde{f_U}$ decays slowly (i.e. polynomially) with t (e.g. U is Laplacian). Note that once we recover the density for $X = \ln(\eta(1 + \eta))$, we can easily recover the density for η.

For normal $U = \ln(\delta_{res})$, Delaigle and Gijbels (2004) suggest a kernel of the form:

$$K(x) = 48 \frac{cos(x)}{\pi x^4}(1 - \frac{15}{x^2}) - 144 \frac{sin(x)}{\pi x^5}(1 - \frac{5}{x^2})$$

This estimator also requires a choice of bandwidth which is a function of sample size. Delaigle and Gijbels (2004) suggest a bootstrap-based bandwidth that minimizes the mean-integral squared error, which is implemented by Wang and Wang (2011) in the R package `decon`, and we use that method here.