

Do People Confuse Average and Marginal Tax Rates? Evidence From a Field Experiment

Daniel L. Chen¹

Abstract

The purpose of this experiment is to see if people react to labor market schedules in a manner that suggests confusion between average and marginal tax rates. I present individuals with identical payment schedules for data entry of a series of paragraphs, making it arguably harder for individuals to be confused by average or marginal payment schemes.

If people “iron”, then individuals presented with payment schedules displaying average payment rates should do more work than individuals presented with payment schedules displaying marginal payment rates. My first experiment demonstrates this. A natural follow-up question is whether individuals still “iron” when they are not presented with either average or marginal payment schedules. This may more accurately reflect an actual real-world setting where schedules are often hidden and difficult to compute. My second experiment finds that when workers are not primed with per paragraph payments, they enter significantly more paragraphs than when they are shown either average or marginal payment schedules.

Methodology

I recruited experimental subjects from an online labor market, created by a labor market intermediary (LMI). Through an interface provided by the LMI, registered users perform tasks (posted by buyers) for money. The tasks are generally simple for humans to do yet difficult for computers---common tasks are captioning photographs, extracting data from scanned documents and transcribing audio clips. Buyers control the features and contract terms of the tasks they post: they choose the design, piece-rate, time allowed per task, how long each task will be available and how many times they want a task completed.²

Workers, who are identified to buyers only by a unique string of letters and numbers, can inspect tasks and the offered terms before deciding whether to complete them. Buyers can require workers to have certain qualifications, but the default is that workers can “accept” a task immediately and begin work. Once the worker “submits” their work, buyers can approve or reject their submission. If the buyer approves, the LMI pays the worker with buyer-provided escrow funds; if the buyer rejects, the worker is paid nothing. The buyer can also grant bonuses, which is useful for our purposes since we can create complex contracts rather than use the default piece-rate format.

Although most buyers post tasks directly on the LMI website, it is possible to host tasks on an external site that workers reach by following a link. I used this external hosting method; I posted a single placeholder task containing a description of the work and a link to follow if subjects

¹ Daniel Chen, Toulouse Institute for Advanced Studies, daniel.li.chen@gmail.com. I thank the Berkman Center for Internet and Society, the John M. Olin Fellowship in Law and Economics, and Institute for Humane Studies for research support.

² Tasks are often done multiple times by different workers for quality-control purposes.

wanted to participate. The subjects would then be randomized (stratified in the order that they arrived at the job) to one of several treatment conditions.

For the real-effort task, subjects transcribed (not translated) paragraph-sized chunks of Adam Smith's *The Wealth of Nations*. A sample paragraph is shown in Figure 1. This task was sufficiently tedious that no one was likely to do it “for fun” and it was sufficiently simple that all market participants could do the task. The source text was machine translated into Dutch, which increased the error rate of the transcriptions, thereby providing a more informative measure of work quality. Translating the text also prevented subjects from finding the text elsewhere on the Internet.³

I advertised the data transcription job as paying at least 11 cents with a bonus available depending on the number of paragraphs transcribed. I displayed the entire schedule of payments for the paragraphs along with each transcription, as shown in Figure 1. The schedule consisted of one column indicating the number of paragraphs completed, a second column indicating the per-paragraph payment, and a third column indicating the total payment inclusive of completing the current paragraph.

The experimental manipulation involved changing whether the per-paragraph payment was shown in average terms or in marginal terms. In other words, individuals saw either the average payment per paragraph or they saw the marginal payment for the current paragraph. Everyone received the same underlying schedule, however. I paid subjects payments for each of the first 10 paragraphs but did not pay for any of the paragraphs thereafter. A homo economicus decision-maker should stop at paragraph 10, since they see the table of payments for all paragraphs and know they will no longer receive any payments. In a second experiment, I manipulated whether individuals saw any per-paragraph information or not. My hypothesis was that showing the marginal or average payment scheme may prime individuals to behave differently than individuals who do not see any priming whatsoever, which might more closely approximate a real tax situation.

Results

I find that the workers who saw the average payment schedule do significantly more paragraphs than the workers who saw the marginal payment schedule. When workers are not primed with per paragraph payments, they enter significantly more paragraphs than when they are shown either average or marginal payment schedules.

Average vs. Marginal Payment Rates

In my first experiment, 86% of participants did more than 1 paragraph (Table 1 Panel A). 28% of participants did more than 10 paragraphs. The average number of paragraphs submitted was

³ Since the text was presented as images, subjects were unable to simply copy and paste the text into the text box on the form. It is irrelevant whether or not any subjects actually knew Dutch since, if anything, knowledge of the language might make the task more difficult given the poor quality of the translation. One subject that apparently did speak Dutch sent an email warning us that the work was “grammatical gibberish.”

9.4. 44% were male. They had on average 14.7 years of education. The average age was 32. They spent 12 hours per week doing these tasks. 284 individuals viewed our job task, but only 79 entered any paragraphs; another 108 subjects went straight to the demographic survey in order to still receive the participation payment. There was no difference in selection rates of individuals across average and marginal treatment groups (72% of individuals in the marginal treatment entered zero paragraphs and 73% of individuals in the average treatment entered zero paragraphs. Results available on request). Demographic characteristics (age, education, gender, and hours spent online) were also balanced across treatment groups among those who remained in the experiment.

Subjects who saw the marginal payment rate instead of the average payment rate reduced the number of paragraphs they entered by 3.5 and this is statistically significant at the 10% level, both with and without demographic controls (Columns 1 and 2 of Table 2). This relationship becomes stronger when the outcome is log number of paragraphs with statistical significance at the 5 or 10% level (Columns 3 and 4). The log number of paragraphs is computed as the natural logarithm of the number of paragraphs. When I include attriters (those who saw the job task but decided not to participate or those who proceeded immediately to the demographic survey) as having entered zero paragraphs, subjects who saw the marginal payment rate instead of the average payment rate reduced the number of paragraphs they entered by 1.7 and this is statistically significant at the 10% level (results available on request).

One of my post-employment survey questions asked subjects to enter how they would choose between several pairs of proposed lotteries. I did this in order to assess risk; while their risk preferences (the number of times they made the riskier choice / higher possible lottery winnings out of ten pairings) were not correlated with the number of paragraphs entered as shown in Table 1, I can use the final lottery pair to assess which subjects were paying attention. Since the final pairing asks subjects to choose between two certain choices, all subjects should pick the higher of two values. When I restrict to these individuals, my results are broadly consistent in magnitude and remains statistically significant at the 10% level for the sample including attriters. The results are slightly weaker, however, which is not surprising since those who are not paying attention are probably the individuals who are more likely to schedule in the first place. As for the other demographic characteristics, not surprisingly, people who spend more time working online enter more paragraphs. People who arrived in the experiment later were much more likely to attrite, but since we stratified subjects to treatment by arrival time, this fact should introduce little bias to my results.

While subjects who saw the marginal payment rate entered fewer paragraphs, and concomitantly spent less time overall, they did not enter them at a speed per paragraph significantly different from subjects who saw the average payment rate. Nor did these subjects have significantly different error rates (results available on request).

Priming Payment Rates vs. No Priming

In my second experiment, 83% of participants did more than 1 paragraph (Table 1 Panel B). 23% of participants did more than 10 paragraphs. The average number of paragraphs submitted was 8.3. 41% were male. They had on average 14.9 years of education. The average age was

30. They spent 11 hours per week doing these tasks. 702 individuals viewed our job task, but only 341 entered any paragraphs. There was no difference in selection rates of individuals across average and marginal treatment groups (49% of individuals who were not primed completed zero paragraphs and 53% of individuals who were primed completed zero paragraphs. Results available on request). Demographic characteristics (age, education, gender, and hours spent online) were balanced across treatment groups among those who remained in the experiment except for age; those who were not primed and remained in the experiment were slightly older (statistically significant at the 10% level).

Subjects who were not primed with per-paragraph payment rates increased the number of paragraphs they entered by 3.1 and this is statistically significant at the 5% level, both with and without demographic controls (Columns 1 and 2 of Table 3). This relationship is not significant when the outcome is log number of paragraphs (Columns 3 and 4). The log number of paragraphs is computed as the natural logarithm of the number of paragraphs. When I only include individuals who completed at least ten paragraphs, those who are not primed are entering 19 more paragraphs and this is statistically significant at the 1% level, both with and without controls (Columns 1 and 2 of Table 4). The relationship remains statistically significant at the 1% level when the outcome is log number of paragraphs (Columns 3 and 4).

When I include attriters (those who saw the job task but decided not to participate) as having entered zero paragraphs, subjects who were not primed with per-paragraph payment rates increased the number of paragraphs they entered by 1.9 and this is statistically significant at the 1% level (results available on request). Since no individuals who entered zero paragraphs entered demographic information, including the controls results in dropping the attriters.

When I restrict to individuals who were paying attention in the risk lottery question, our results are broadly consistent in magnitude and statistical significance, but again, are slightly weaker, consistent with the notion that individuals who are not paying attention are the ones more likely to schedule (results available on request). As for the other demographic characteristics, not surprisingly, people who spend more time working online enter more paragraphs.

While subjects who were not primed entered more paragraphs, and concomitantly spent more time overall, they did not enter them at a speed per paragraph significantly different from subjects who saw the average payment rate. Nor did these subjects have significantly different error rates (results available on request).

Summary

I offered a kinked wage schedule and show that labor is curtailed when workers see the marginal payment schedule as opposed to the average payment schedule. When workers were not primed with average or marginal payment rates, they did significantly more work. These results suggest that workers may be less responsive to marginal rates than previously thought, eliminating some of the loss in labor supply from high marginal taxes.

If you do this paragraph: Your total amount earned will be: 11 cents Your average earnings per paragraph will be: 11 cents

Paragraph Number : 1

Number of Paragraphs Completed	Average Payment Per Paragraph	Total Payment
1	11	11
2	11	22
3	11	33
4	11	44
5	11	55
6	11	66
7	11	77
8	11	88
9	11	99
10	11	110
11	10	110
12	9.17	110
13	8.46	110
14	7.86	110
15	7.33	110
16	6.88	110
17	6.47	110
18	6.11	110
19	5.79	110
20	5.50	110
21	5.24	110
22	5	110
23	4.78	110
24	4.58	110
25	4.40	110
26	4.23	110
27	4.07	110
28	3.93	110
29	3.79	110
30	3.67	110
31	3.55	110
32	3.44	110
33	3.33	110
34	3.24	110
35	3.14	110

De jaarlijkse arbeid van elk volk is het fonds die oorspronkelijk levert hij met alle benodigdheden en conveniencias van het leven die het jaarlijks verbruikt, en die altijd bestaan, hetzij in de onmiddellijke

Enter transcription here:

Payment for next task:

- Do another
- Finish
- Cancel this paragraph and Finish

Submit Work

Table 1: Summary Statistics

Panel A		Experiment 1 -- Average vs. Marginal Payment Rates							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Age	Education	Male	Arrival Time	Hours Online	Attrite	Over_one	Over_ten	Num_para
Mean	31.82 (0.769)	14.74 (0.241)	0.441 (0.0365)	8.293 (.569)	11.95 (1.073)	0.722 (0.0266)	0.861 (0.0392)	0.278 (0.0508)	9.494 (0.928)
N	186	186	186	186	186	284	79	79	79
Panel B		Experiment 2 -- Priming vs. No Priming of Per Paragraph Payment Rates							
Mean	30.34 (0.593)	14.86 (0.185)	0.405 (0.0277)	40.184 (.543)	10.73 (0.636)	0.514 (0.0189)	0.827 (0.0205)	0.232 (0.0229)	8.276 (0.626)
N	316	316	316	316	316	702	341	341	341

Notes: Each column displays a regression of the outcome variable on a constant. Standard errors are in parentheses. Over_one is an indicator for entering more than one paragraph. Over_ten is an indicator for entering more than ten paragraphs. Num_para is the number of paragraphs entered. Attrite is an indicator for not entering any paragraphs. Experiment 1 had subjects who did not enter paragraphs but provided demographic information. Experiment 2 did not have any subjects do that and some subjects who entered paragraphs then chose not to provide demographic information.

Table 2: The Effect of Showing Marginal or Average Payment Rates on the Number of Paragraphs Completed

	Conditional on Completing At Least 1 Paragraph			
	Number of Paragraphs		Log Number of Paragraphs	
	(1)	(2)	(3)	(4)
Marginal (Treatment Indicator)	-3.583*	-3.480*	-0.762**	-0.733*
	(1.822)	(1.935)	(0.376)	(0.410)
Age		0.119		0.0110
		(0.0817)		(0.0173)
Education		-0.0673		0.0149
		(0.294)		(0.0623)
Male		-1.861		0.0563
		(1.946)		(0.412)
Arrival		-0.917		-0.153
		(0.873)		(0.185)
Hours		0.0799		0.0107
		(0.0579)		(0.0123)
Risk_p		0.337		0.0827
		(0.443)		(0.0939)
Constant	11.31***	8.255	5.256***	4.401***
	(1.297)	(5.980)	(0.268)	(1.266)
N	79	76	79	76
R-sq	0.048	0.146	0.051	0.097

Notes: Standard errors in parentheses (* p<0.10 ** p<0.05 *** p<0.01). Log number of paragraphs is computed as ln(number of paragraphs). Demographic controls are age, education (in years), gender, arrival time, hours spent online per week, and risk preferences.

Table 3: The Effect of Priming Payment Rates on the Number of Paragraphs Completed

	Conditional on Completing At Least 1 Paragraph			
	Number of Paragraphs		Log Number of Paragraphs	
	(1)	(2)	(3)	(4)
No_Prime (Treatment Indicator)	3.124** (1.301)	3.098** (1.393)	0.188 (0.230)	0.207 (0.238)
Age		-0.0145 (0.0646)		-0.00638 (0.0110)
Education		0.00454 (0.205)		-0.00323 (0.0349)
Male		0.120 (1.367)		-0.210 (0.233)
Arrival		-0.0520 (0.0696)		-0.0150 (0.0119)
Hours		0.111* (0.0596)		0.0160 (0.0102)
Risk_p		-0.0141 (0.275)		-0.0327 (0.0469)
Constant	7.176*** (0.772)	8.733* (4.815)	4.027*** (0.136)	4.977*** (0.822)
N	341	316	341	316
R-sq	0.017	0.030	0.002	0.021

Notes: Standard errors in parentheses (* p<0.10 ** p<0.05 *** p<0.01). Log number of paragraphs is computed as ln(number of paragraphs). Demographic controls are age, education (in years), gender, arrival time, hours spent online per week, and risk preferences.

Table 4: The Effect of Priming Payment Rates on the Number of Paragraphs Completed

	Conditional on Completing At Least 10 Paragraphs			
	Number of Paragraphs		Log Number of Paragraphs	
	(1)	(2)	(3)	(4)
No_Prime (Treatment Indicator)	18.58*** (3.614)	18.71*** (3.843)	1.403*** (0.271)	1.407*** (0.291)
Age		0.0416 (0.182)		0.00413 (0.0138)
Education		-0.180 (0.523)		-0.0289 (0.0395)
Male		6.201* (3.591)		0.339 (0.272)
Arrival		-0.0387 (0.193)		-0.00117 (0.0146)
Hours		0.162 (0.149)		0.00278 (0.0113)
Risk_p		0.543 (0.771)		0.0530 (0.0583)
Constant	17.29*** (1.950)	14.27 (13.43)	6.477*** (0.146)	6.512*** (1.016)
N	79	76	79	76
R-sq	0.256	0.301	0.258	0.285

Notes: Standard errors in parentheses (* p<0.10 ** p<0.05 *** p<0.01). Log number of paragraphs is computed as ln(number of paragraphs). Demographic controls are age, education (in years), gender, arrival time, hours spent online per week, and risk preferences.