

CSE 518

Financial Incentives

Presenters:

Rajagopal Venkatesaramani (Raj)
Michael Guo

Goals as a Requester

- To collect data with least error (if quantifiable).
- Motivate/**incentivize** workers to put reasonable effort.

Worker Incentives

- Joy!
- Greater Good!

\$\$\$\$\$ (really, \$)

Designing Incentives

- Pay More
 - Raise the base payment
 - Bonuses for higher quality work
- Pay to Quit

Financial Incentives and the “Performance of Crowds”

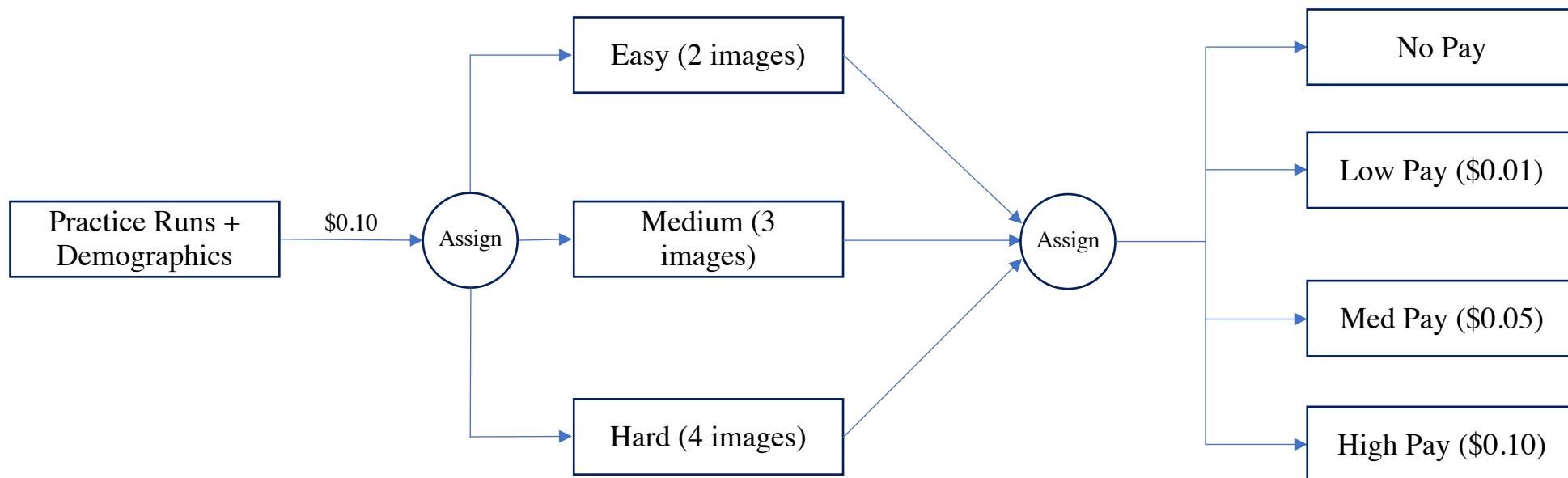
Winter Mason and Duncan J. Watts, KDD-HCOMP 2009

Aim : To study the relationship between financial incentives and performance.

Experiments :

1. Chronologically order up to 99 sets of images from traffic cameras.
2. Solve up to 24 word puzzles.

Traffic Images - Setting



Interface

Figure 1. Screenshot of the description of Study 1, which participants saw when deciding to participate in the experiment.

Instructions

At the beginning of a task, you will be presented with a list of images taken from traffic cameras. An example list is shown below.



1. CAM-047 A / 2. CAM-047 A / 3. CAM-047 A /

Your goal is to reorder the list chronologically from left to right and top to bottom. The sorted list is shown below.



1. CAM-047 A / 2. CAM-047 A / 3. CAM-047 A /

Notice that in the sorted row, the truck on the right moves away from the camera, and the blue cab on the left approaches the camera. To correctly sort the photos, you need to determine the flow of the traffic.

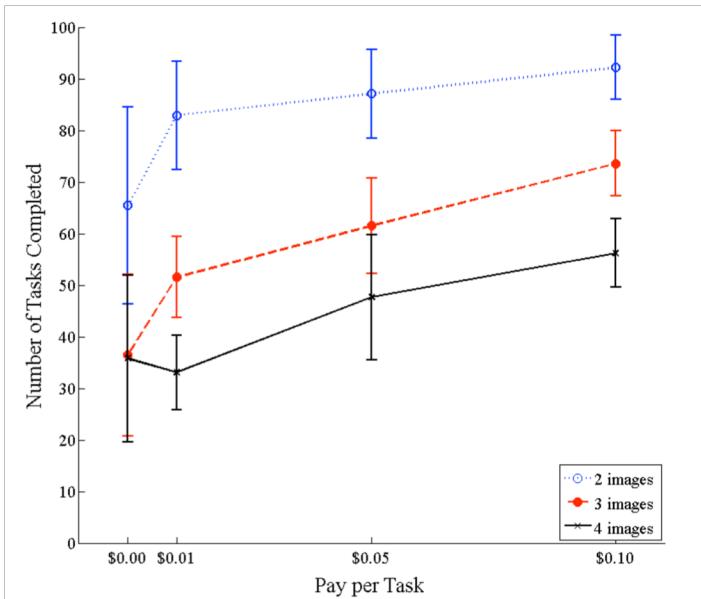
To reorder a list, click and drag a photo to the position it belongs. The other photos will move accordingly. Once you believe the list is in the correct order, click on the "Submit" button at the top of the page.

If you do not want to complete any more tasks, click on the "Finished" button at the bottom of the page.
(This button will not be available in the next 3 practice examples)

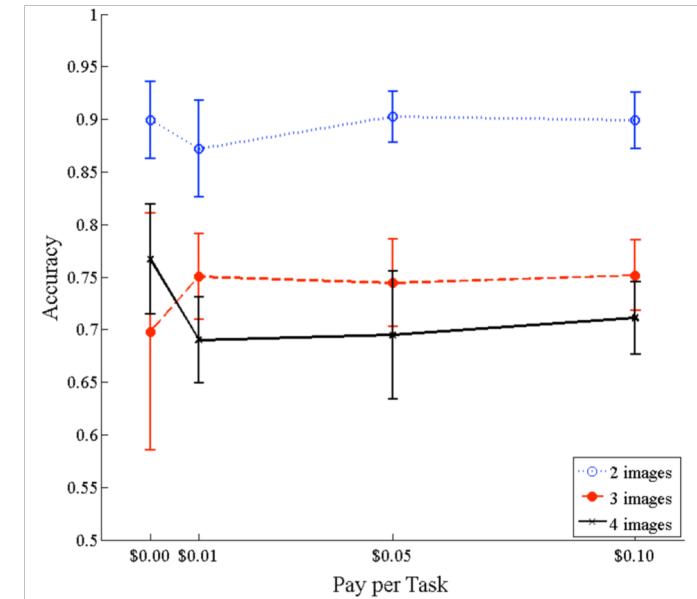
Click here to practice: [Practice](#)

Images provided by Trafficland.com

Results



Compensation correlated with number of tasks completed.



Compensation **not** correlated with accuracy.

Really?

“One possible explanation for the absence of an effect of wages on accuracy is that subjects simply assumed they would be paid regardless of performance.”

Additional Experiment : Fixed rate \$0.01. Some participants were given the same instructions as before while others were told that one out of every four image sets was a test image set used to gauge their accuracy.

Four settings:

- Participants informed accuracy is measured.
- Feedback on accuracy after every fourth image.
- Participants explicitly told that payment would be contingent on accuracy.
- Participants given feedback AND told payment is contingent.

No significant difference.

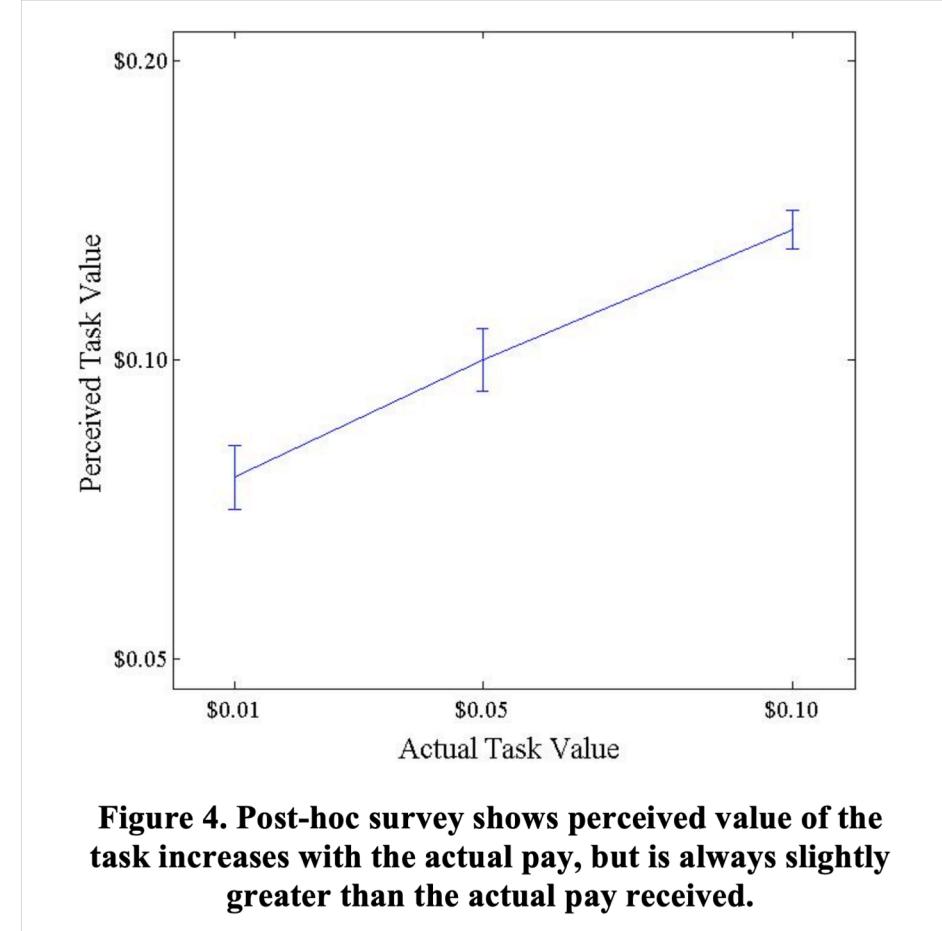
Discussion

- Why did increased payments not help with accuracy?
 - Not much room for improvement.
 - “Anchoring Effect” – wait for it...
- Can the experiment design be improved to better account for these factors?

Anchoring Effect

The workers perceive the value of their work to be higher than the offered payments.

Higher payments thus may not in themselves be sufficient motivation.



Word Puzzles

- Workers shown a list of 15 words possible in the grid (but not necessarily. 7-14 words present. Median 11).
- Number of words found indicates quality, number of puzzles completed indicates quantity.
- Yet again, various settings.

To select a word, first click on the first letter of the word, then click on the last letter of the word. If you are correct, it will turn red and the word will appear to the right of the puzzle.

For each puzzle you will see a set of possible words and their category. Not all of the words listed are in the puzzle! In addition, the number of words in each puzzle changes. The list of possible words follows:
ACHIEVE, ATTAIN, BUILDING, CHAIR, COMPETE, GREEN, LAMP, MASTER, MUSIC, PLANT, STAPLE, STEREO, STRIVE, SUCCEED, TURTLE
For this practice puzzle, you will have to find at least 8 words to continue.

RANDOM WORDS

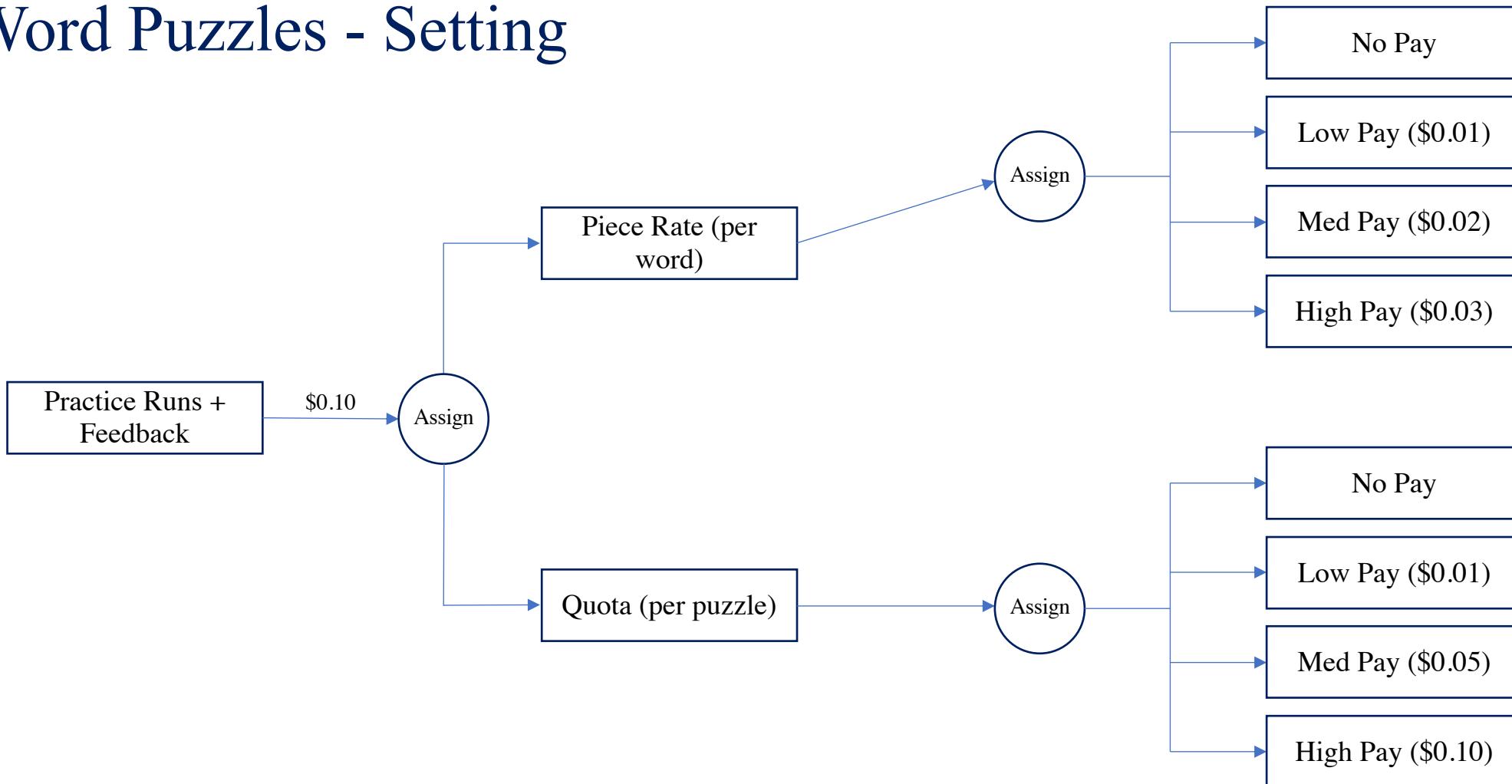
SUCCEED
BUILDING
ACHIEVE

Submit Puzzle

M	A	S	T	E	R	M	O	Z	Q	K	K	W	L	F
R	G	T	R	D	B	U	I	L	D	I	N	G	O	T
D	S	U	C	U	W	R	J	B	M	Q	P	G	L	C
L	Q	R	P	E	T	E	P	M	O	C	A	E	P	F
K	Z	T	C	V	F	T	B	X	W	Q	V	Q	A	I
S	U	L	F	J	D	G	W	U	W	I	C	G	Z	O
W	O	E	O	M	P	M	A	L	R	E	S	U	Z	L
X	V	R	Q	X	O	N	T	T	L	N	U	F	N	W
Y	H	B	A	I	E	N	S	P	L	I	C	N	E	F
E	H	L	D	T	A	V	A	W	I	S	C	E	K	U
N	K	G	K	L	T	T	E	F	Y	B	E	E	C	M
O	P	R	P	B	S	A	L	I	P	J	E	R	N	X
U	V	B	A	Z	J	M	I	D	H	F	D	G	U	V
Q	P	I	A	W	T	F	U	N	U	C	H	G	I	X
Y	K	I	O	T	W	L	O	R	N	O	A	Q	W	A

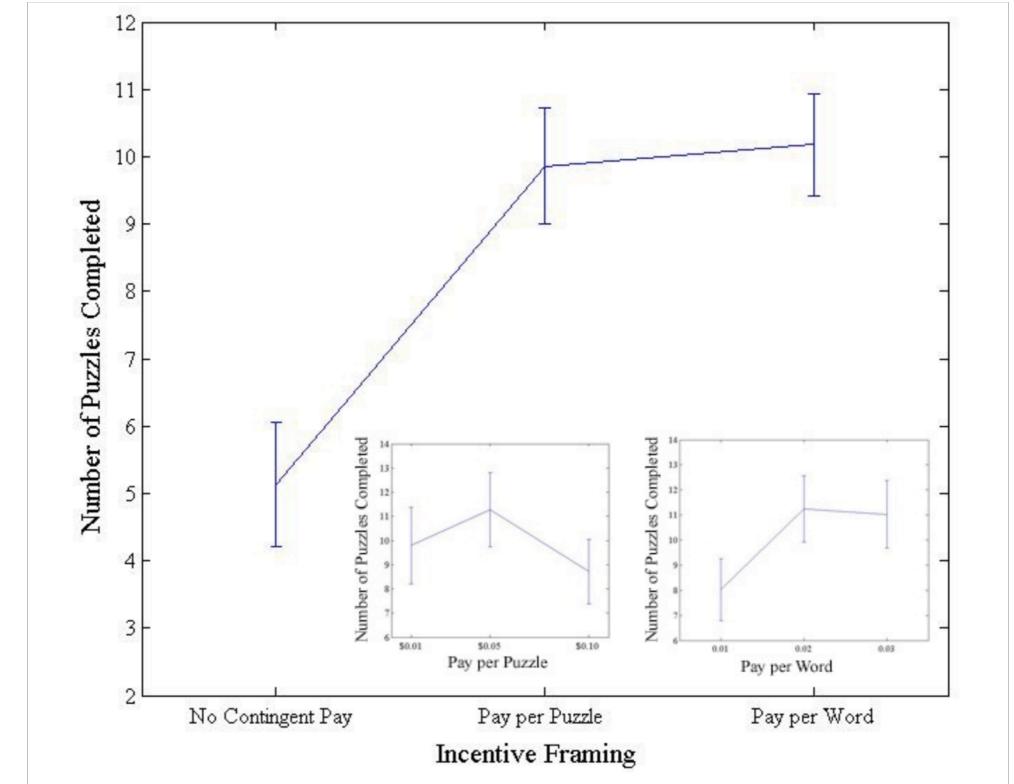
Figure 5. Screenshot of Study 2. Participants found words hidden in a grid of letters.

Word Puzzles - Setting



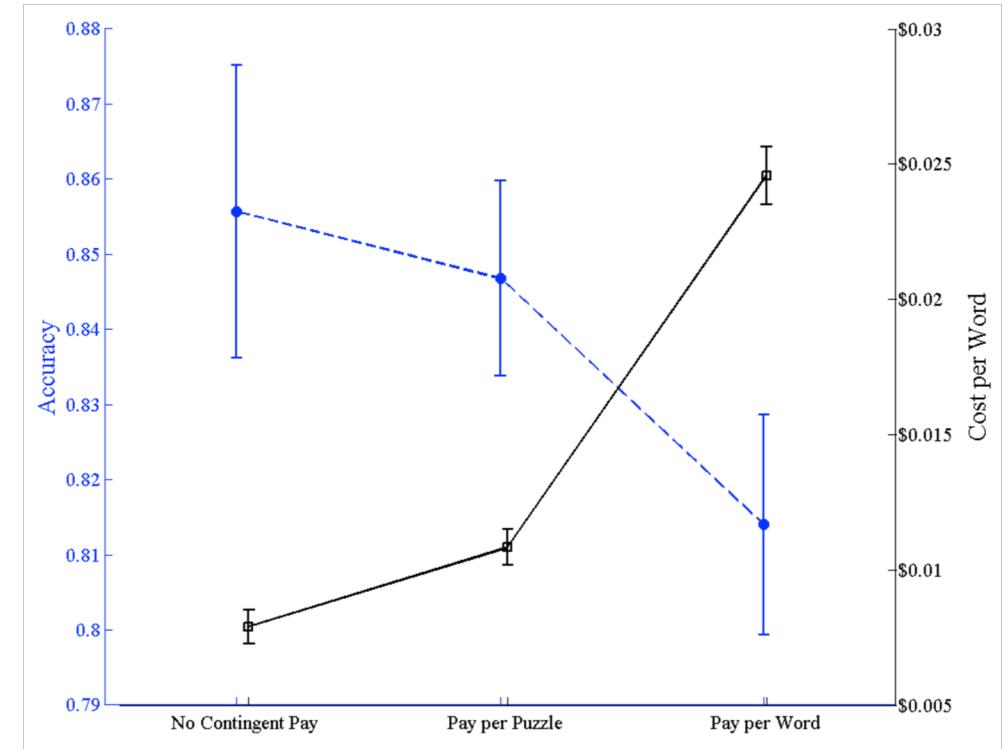
Results

Figure 6. Participants who were paid for each puzzle completed or word found completed significantly more puzzles than those who did not receive contingent pay. Insets show number of puzzles completed did not differ by pay, within per-puzzle (left) or per-word (right) schemes.

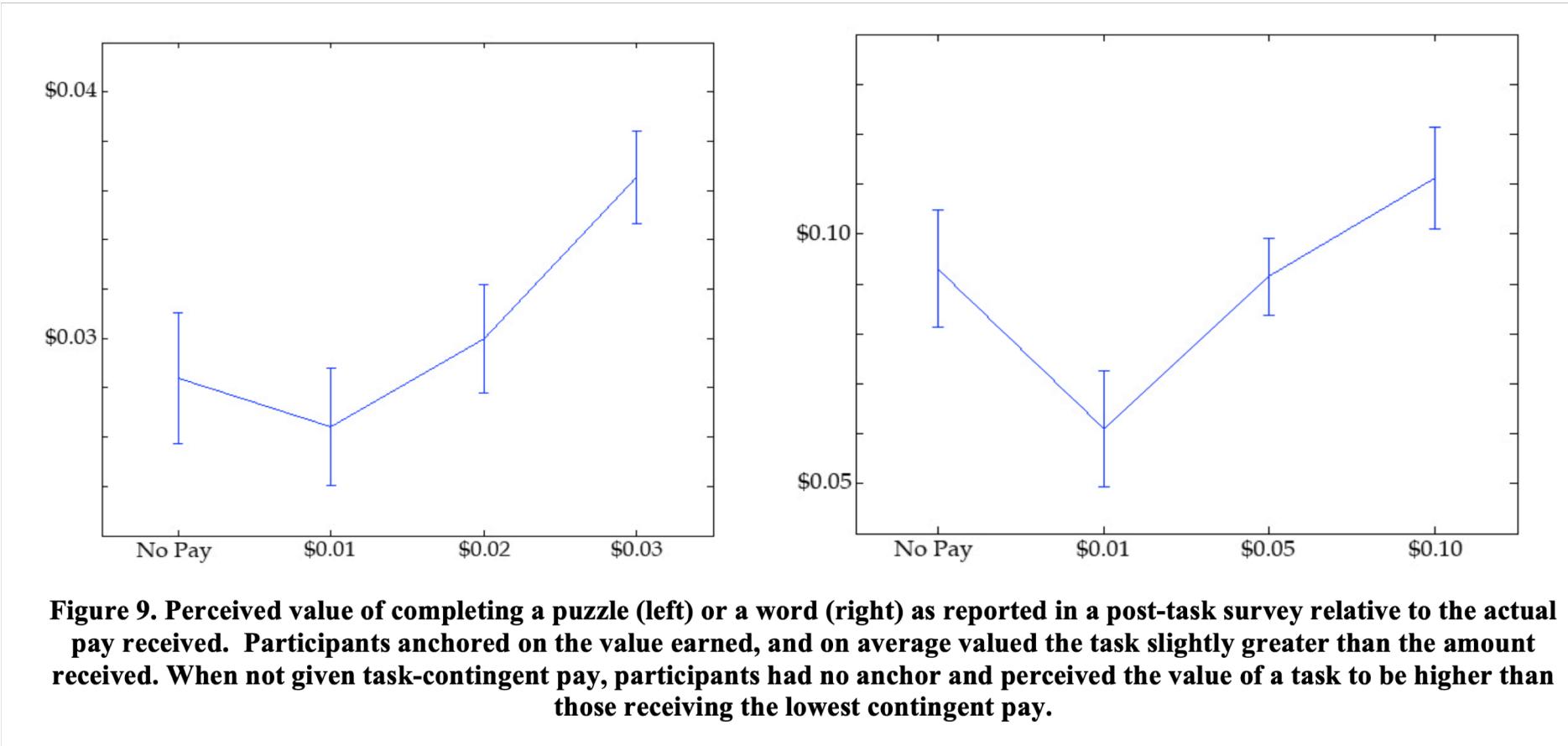


Results

Figure 7. Participants were paid more in total (solid line) but found fewer words per puzzle (dashed line) in the pay per-word condition than in the pay per-puzzle condition.



Anchoring Effect



Designing Incentives

- Pay More
 - Raise the base payment
 - Bonuses for higher quality work
- Pay to Quit

Incentivizing High Quality Crowdwork

Chien-Ju Ho et al., WWW 2015

Aim : To study the effectiveness of *Performance-Based-Payments* (PBP).

Experiments :

1. Find (20) typos in a given document – with variations.
2. Find differences in pairs of images.
3. Handwriting Recognition and Audio Transcription.

Incentivizing High Quality Crowdwork

Chien-Ju Ho et al., WWW 2015

Definition : An *effort-responsive* task is one where a worker can improve the quality of the submission by putting in a higher effort.

Finding Typos – Does PBP work?

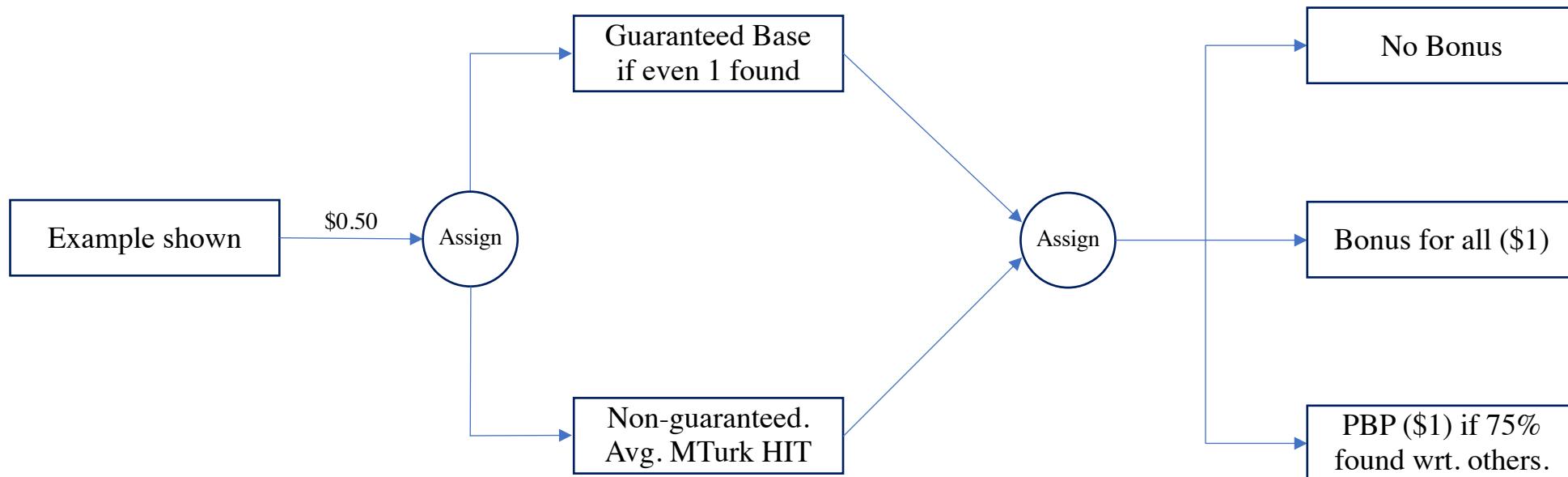
- 20 typos (from a list of common typos) were inserted into articles with 500-700 words.
- Workers were asked to list the line number, the incorrect word and the correct spelling.
- Restricted to US workers.

DISCUSSION:

Why is this a good experiment design?
You're allowed to disagree of course.

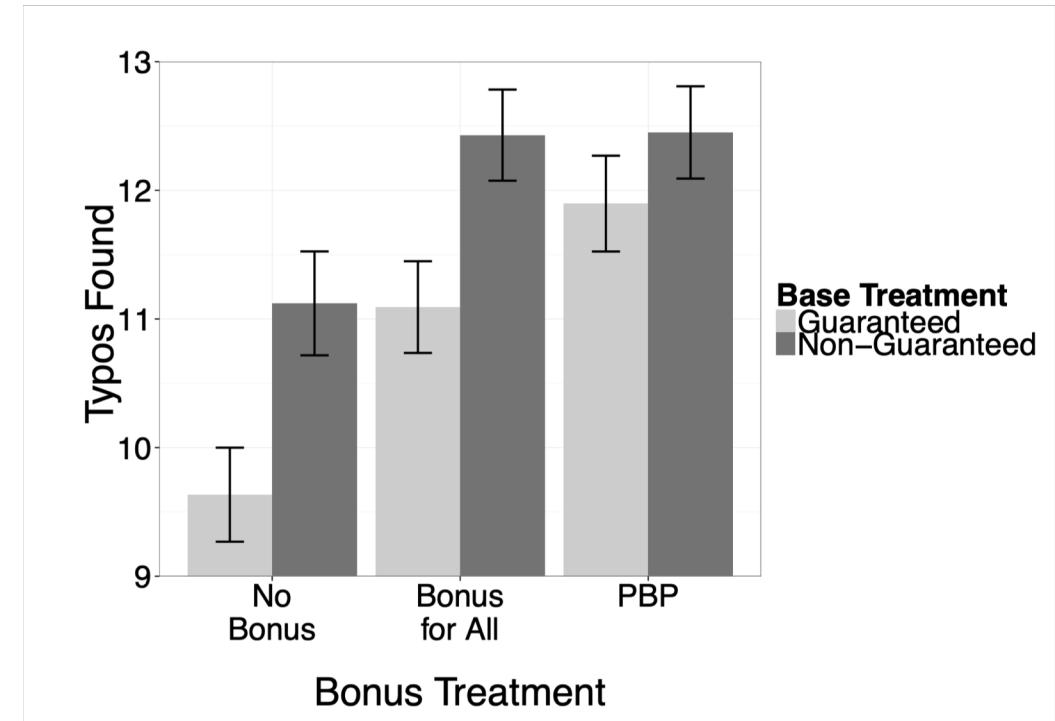
Reminder: It is CJ's paper, so be nice.

Typos - Setting



Results

- PBPs improve quality.
- Simply paying more improves quality.
- PBPs can save money compared to high unconditional payment.
- All payments may be implicitly performance based.



Finding Typos – *When* does PBP work?

(a) Varying Bonus Threshold

- Find 5 typos.
- Find 25% of typos that others found.
- Find 50% of typos that others found.
- Find 75% of typos that others found.

(b) Varying Bonus Amounts

- No Bonus (Control)
- \$0.05
- \$0.5
- \$1

if found 75% of other workers' typos.

Results

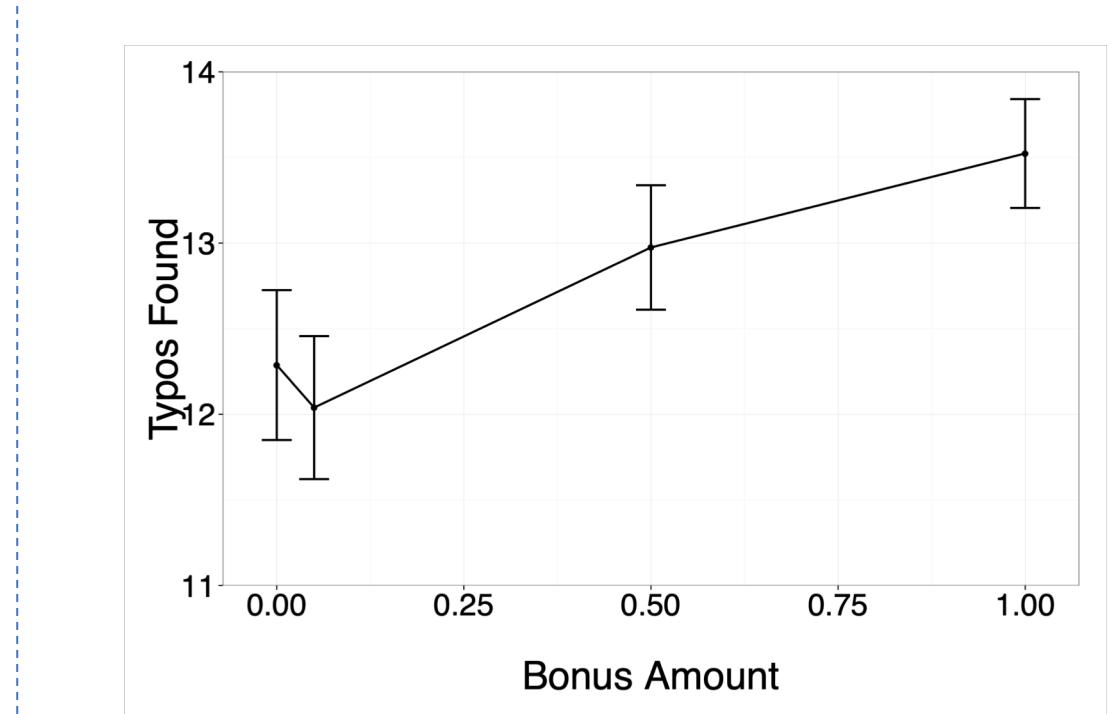
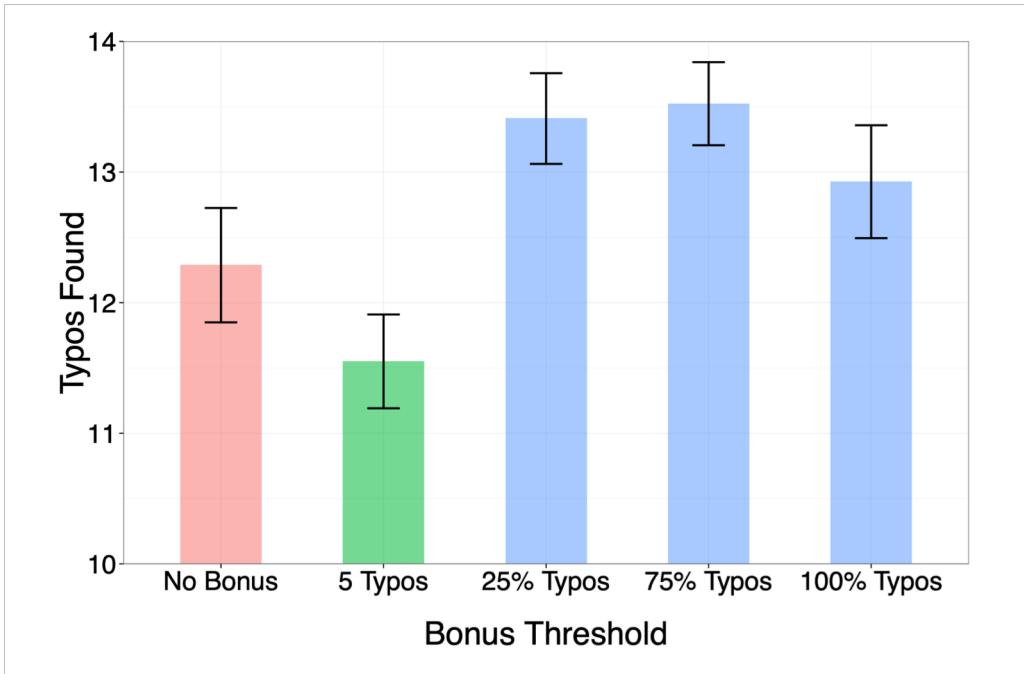
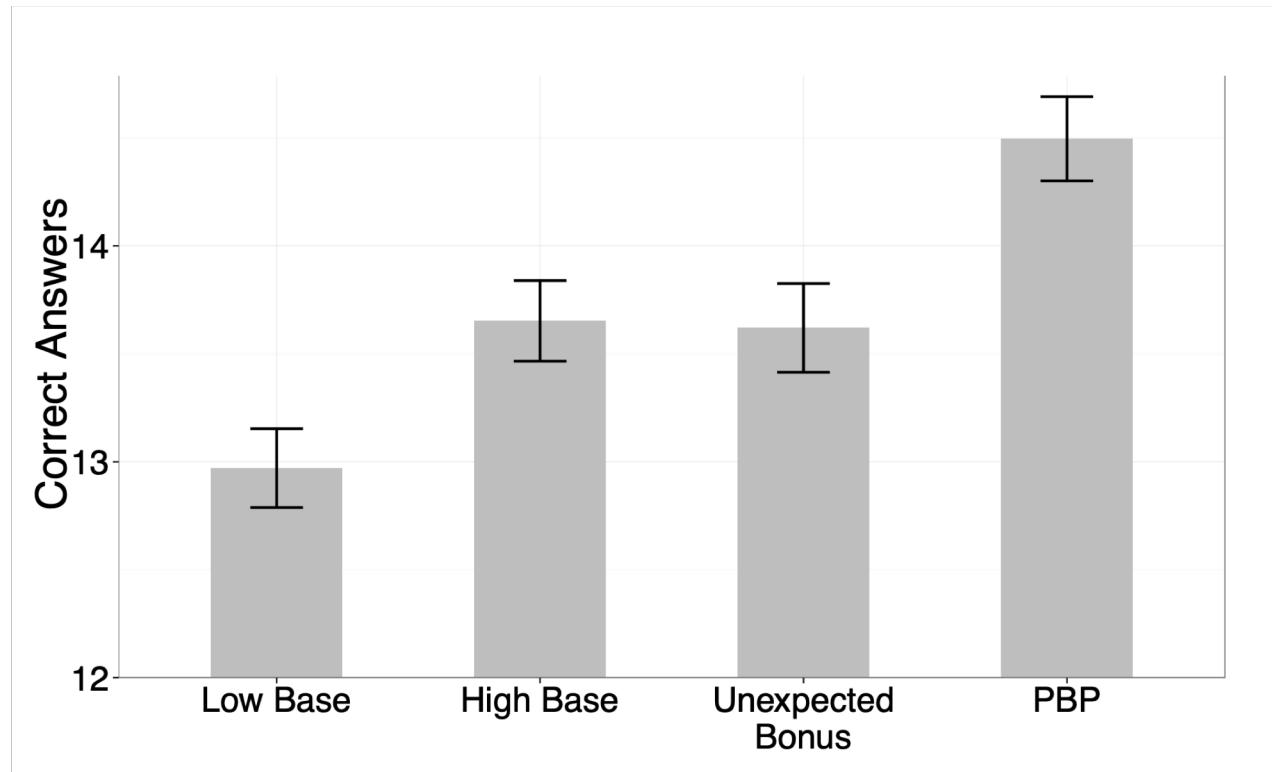


Image Differences – *Why* does PBP work?

- Workers shown 20 pairs of images.
 - 10 are identical.
 - 10 contain minor differences.
- Four payment levels:
 - Low Base (\$0.50)
 - High Base (\$1.50)
 - Unexpected Bonus (\$0.50 + \$1.00)
 - PBP (\$0.50 + \$1.00 if correctly labelled 80% of image-pairs.)

Results

- PBPs work!
- Prior knowledge of bonus doesn't matter.



Where does PBP work?

(a) Handwriting Recognition

- 2 images with 89 and 74 words.
- Control Group – Base Payment only.
- Bonus Group – upon achieving 90% accuracy on gold standard on 1 of the images.

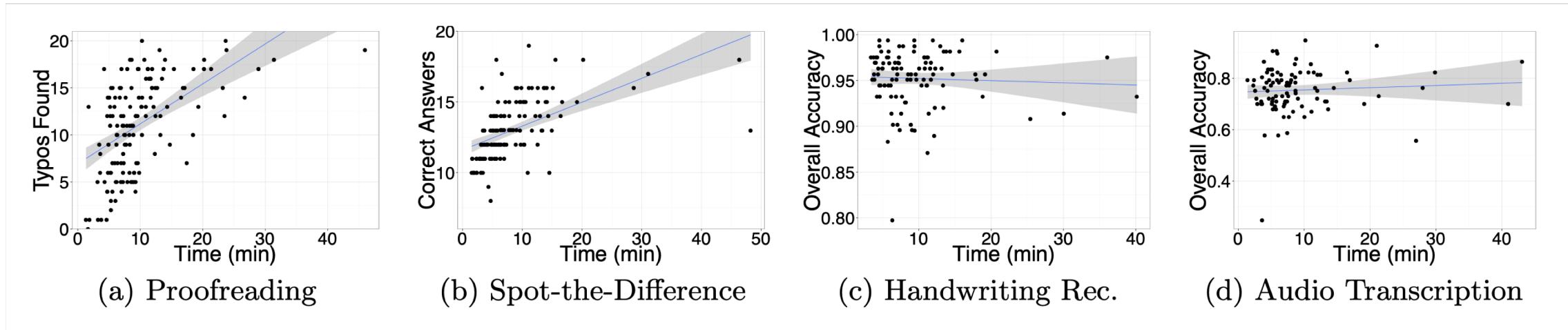
(b) Audio Transcription

- 10 clips – 5 seconds each.
- Heavy accents chosen.
- Workers told 5 out of 10 will be compared to gold standard, and bonuses offered at
 - 80% accuracy
 - 85% accuracy
 - 90% accuracy

DISCUSSION:

What's the difference between these tasks and previous experiments?

Results



Effort Responsive

Not Effort Responsive

And of course, we'll now do some math.

Designing Incentives

- Pay More
 - Raise the base payment
 - Bonuses for higher quality work
- Pay to Quit

Q : How/why would pay-to-quit incentives work?

Pay to Quit Incentives

- Companies such as Zappos.com and Amazon.com provide financial incentives for newer employees to quit.
- The premise is that workers who will accept this offer are misaligned with their company culture, which will therefore negatively affect quality over time.

The Effects of Pay-to-Quit Incentives on Crowdworker Task Quality

Christopher G. Harris, CSCW 2015

5 empirical experiments to evaluate different pay-to-quit incentives, and measure effect on accuracy, retention rate and improvement.

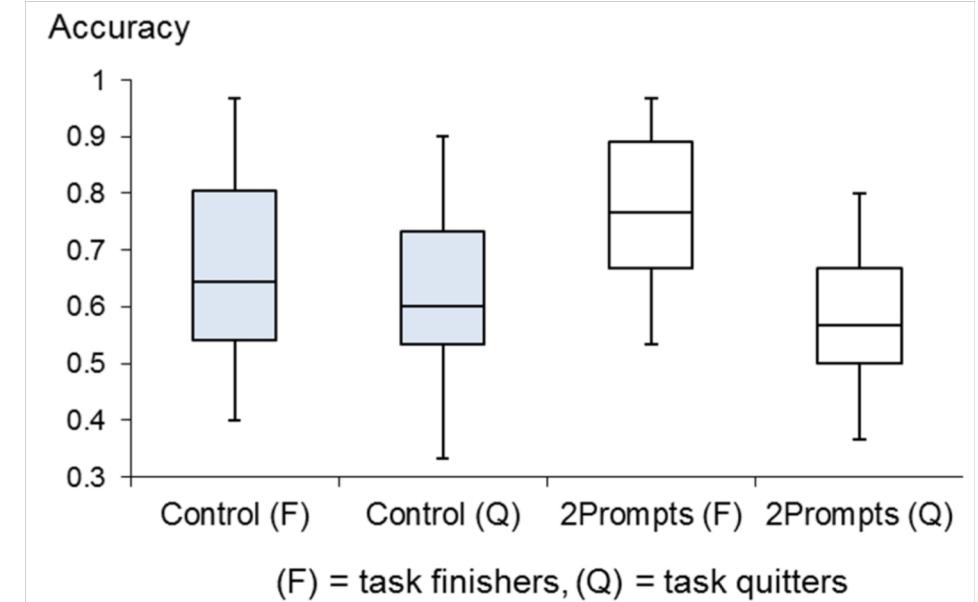
1. The Role of Explicit Pay to Quit Incentives

- Answer 30 questions – choose movie title based on description or image.
- Workers offered incentives (\$0.01 per question) to quit using prompts.
- While users can quit the HIT anytime, they only get paid if they quit on a prompt, except control group.
 - Quit at question 15.
 - Quit at question 10 or 20.
 - Quit at question 7, 14, or 21.

Results

Table 1. Mean accuracy and standard deviations for the control group (N=60) and three treatment groups (each N=60), each offering a different number of inducement prompts.

Group	Control		1 Prompt		2 Prompts		3 Prompts	
	M	SD	M	SD	M	SD	M	SD
All	.647	.175	.631	.187	.708	.135	.612	.165
Finishers	.653	.171	.639	.193	.738	.126	.610	.178
Quitters	.645	.194	.597	.164	.575	.111	.619	.095



2. The Role of Performance Feedback

- Feedback consists of participant's accuracy and average accuracy of other workers.
 - Control group provided feedback at the end.
 - Feedback and quit prompt at questions 10 and 20.
 - Accurate feedback after each question.
 - Accurate accuracy but told they outperformed average on questions 10 and 20.
 - Accurate accuracy but told they performed worse than average on questions 10 and 20.

Results

Table 2. Mean accuracy and standard deviations for the control group (N=60) and two treatment groups (N=60), each offering a different frequency of feedback to the participant.

Group	Control		Each Round		Each Prompt	
	M	SD	M	SD	M	SD
All	.629	.159	.649	.176	.686	.153
Finishers	.636	.157	.668	.177	.712	.151
Quitters	.603	.167	.583	.171	.559	.159

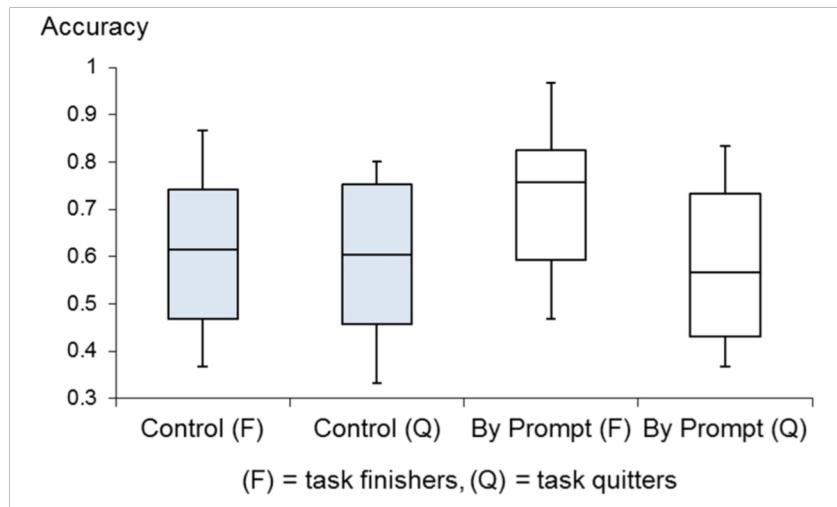
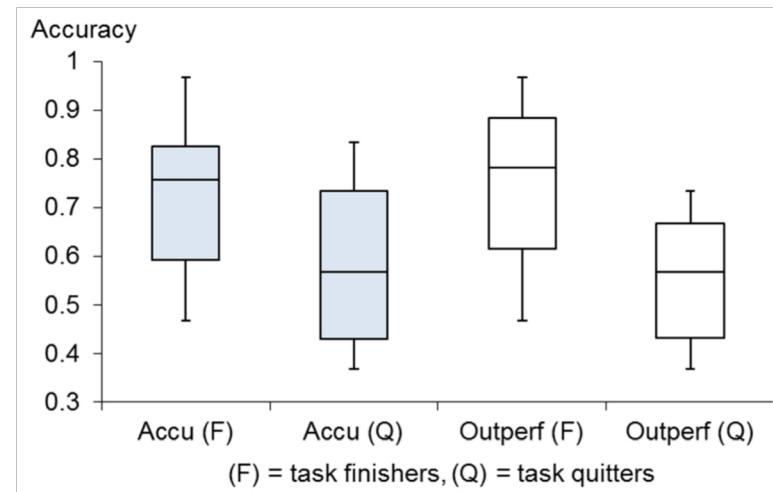


Table 3. Mean accuracy and standard deviations for three treatment groups (each N=60), each providing a different feedback message during the pay-to-quit incentive prompt.

Group	Accurate		Outperform		Underperform	
	M	SD	M	SD	M	SD
All	.658	.153	.717	.157	.627	.151
Finishers	.712	.151	.727	.155	.695	.155
Quitters	.559	.159	.698	.167	.502	.143



3. The Role of Incentive Type

- Control Group - \$0.01 per question at 2 prompts.
- Treatment 1 – Increasing Incentive.
 - \$0.01 per question + \$0.01 for each correct at 2 pay-to-quit prompts.
- Treatment 2 – Decreasing Incentive.
 - $n = c * (30 + \#incorrect - \#answered)$; $c = 0.5$

Results

Table 4. Mean accuracy and standard deviations for a control group (N=60) and two treatment groups (each N=60), each offering a different incentive type.

Group	Control		Increasing		Decreasing	
	M	SD	M	SD	M	SD
All	.648	.156	.682	.171	.653	.151
Finishers	.656	.152	.707	.172	.699	.149
Quitters	.618	.170	.573	.164	.525	.158

What does this set of results tell us?

Absolutely nothing at all. Moving on...

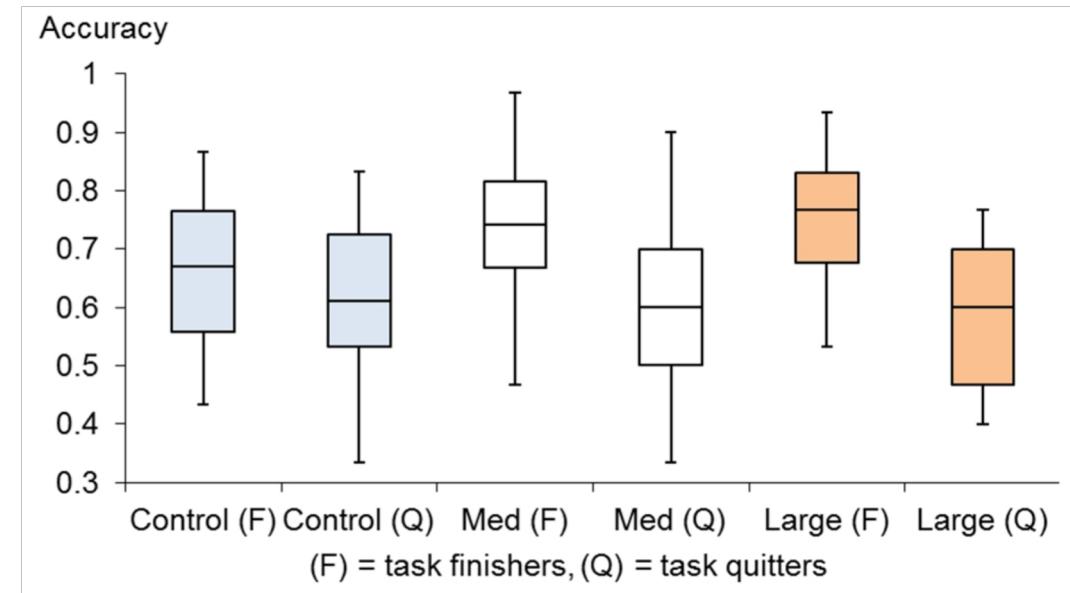
4. The Role of Incentive Size

- Control Group - \$0.01 per question at 2 prompts.
- Treatment 1 – Small decreasing incentive ($c = 0.5$)
- Treatment 2 – Medium decreasing incentive ($c = 0.75$)
- Treatment 3 – Large decreasing incentive ($c = 1$)

Results

Table 5. Mean accuracy and standard deviations for the control group (N=60) and two treatment groups, (each N=60), each offering a different incentive size.

Group	Control		Small (0.5)		Med (0.75)		Large (1.0)	
	M	SD	M	SD	M	SD	M	SD
All	.649	.154	.673	.159	.679	.141	.681	.139
Finishers	.651	.156	.691	.161	.719	.144	.722	.138
Quitters	.621	.146	.599	.149	.568	.131	.569	.142



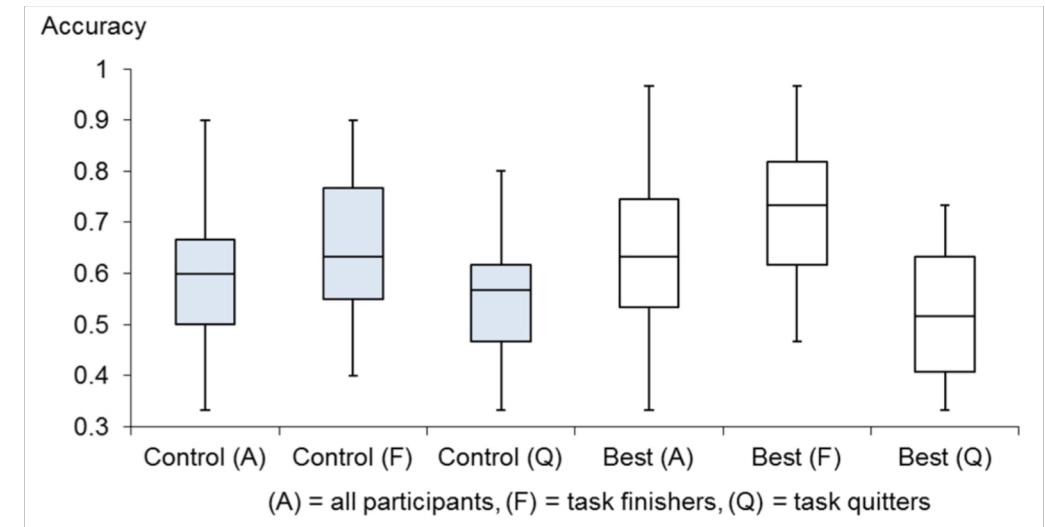
5. Putting It All Together

- Single Best Model.
- Two prompts – questions 10 and 20.
- Accurate feedback on prompts.
- Moderate Decreasing Incentive ($c = 0.75$).

Results

Table 6. Mean accuracy and standard deviations comparing a control group ($N=120$) and a treatment group ($N=120$) comprised of the best strategies from Experiments 1 through 4

Group	Control		Single Best Strategy	
	M	SD	M	SD
All	.637	.146	.687	.133
Finishers	.642	.145	.717	.133
Quitters	.618	.151	.578	.131



73% retention rate on single best model.

27% of workers took incentive to quit.

Thank You!

Questions?

Email us if you're shy!

Raj: rajagopal@wustl.edu
Michael: guominzhe@wustl.edu