Ben and Abhishek - Iceland
Constiution
Clara and Dennis Sell - Ushahidi
Chris - StackOverflow

Quality Control - part 2

**Crowdsourcing and Human Computation
Lecture 13**

**Instructor: Chris Callison-Burch
TA: Ellie Pavlick**

Website: crowdsourcing-class.org

Different Mechanisms for Quality Control

- Aggregation and redundancy

- Embedded gold standard data

- **Economic incentives**

- Reputation systems

- Statistical models

Does pay impact quality?

- Economic theory holds that workers are rational actors

- Will choose to improve their performance in response to a scheme that rewards improvements with financial gain

- Example: executive compensation tied to stock price

Different pay schemes

- Lazear studied of workers who installed windshields on a production line

- Switched from pay per hour to pay per unit during a year and a half

- Individual productivity for workers who started in the hourly rate and switched to the per-unit scheme increased by 20%

- Conclusion: performance-based pay schemes can elicit improved performance

Is that the whole story?

- Sometimes financial incentives can undermine "intrinsic motivation". This can lead to poorer outcomes.

- For complex tasks, performance pay can encourage workers to focus only on the aspects of their jobs that are actively measured

- Can also lead to employees avoid taking risks, thereby hampering innovation

Financial Incentives and the "Performance of Crowds"

- Experiment with economic incentives on Amazon Mechanical Turk

- An exciting tool for behavioral research, since you can recruit thousands of participants from a real labor market

Impact of compensation

- Does compensation change the quantity of work performed (output)?

- Does it change the quality of the work (accuracy)?
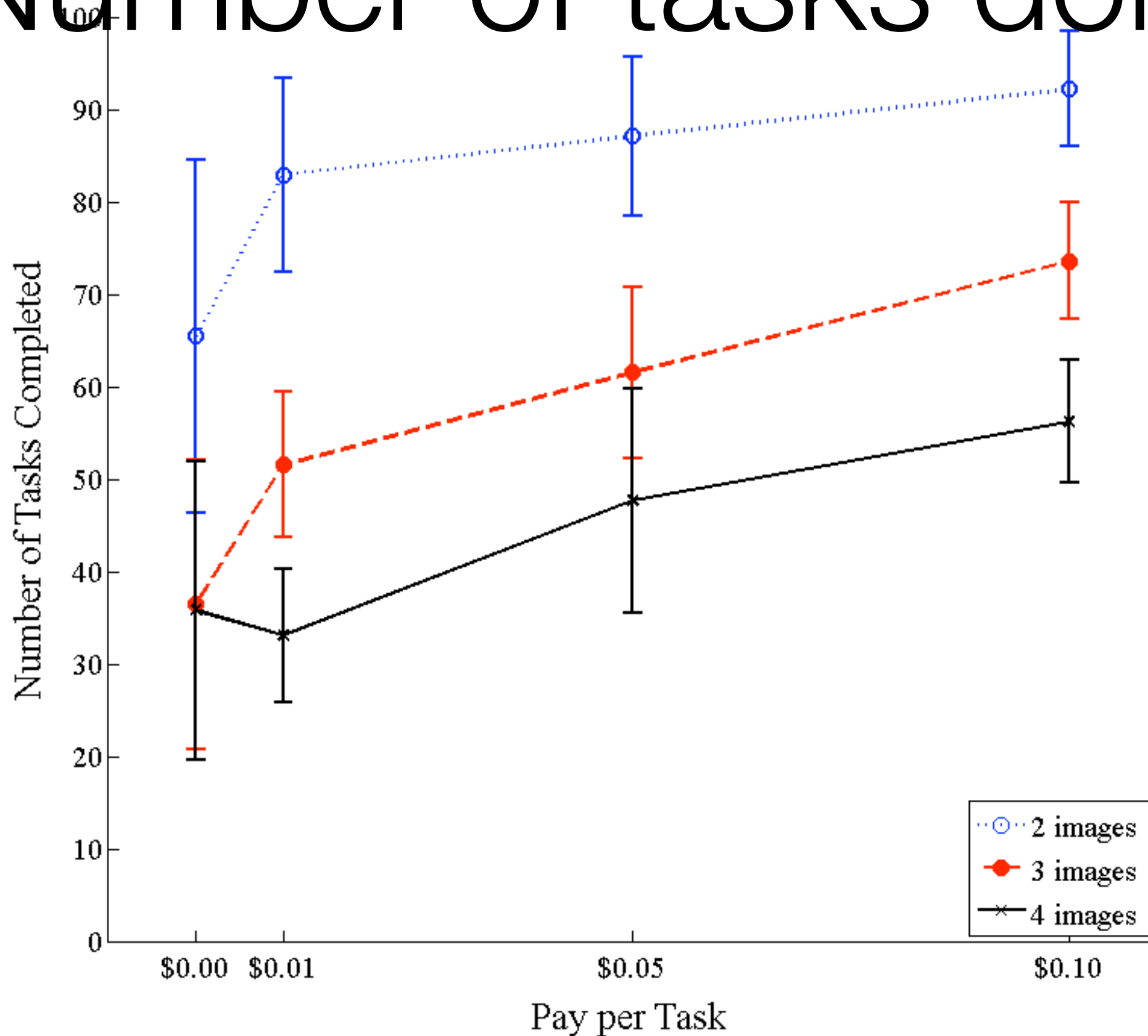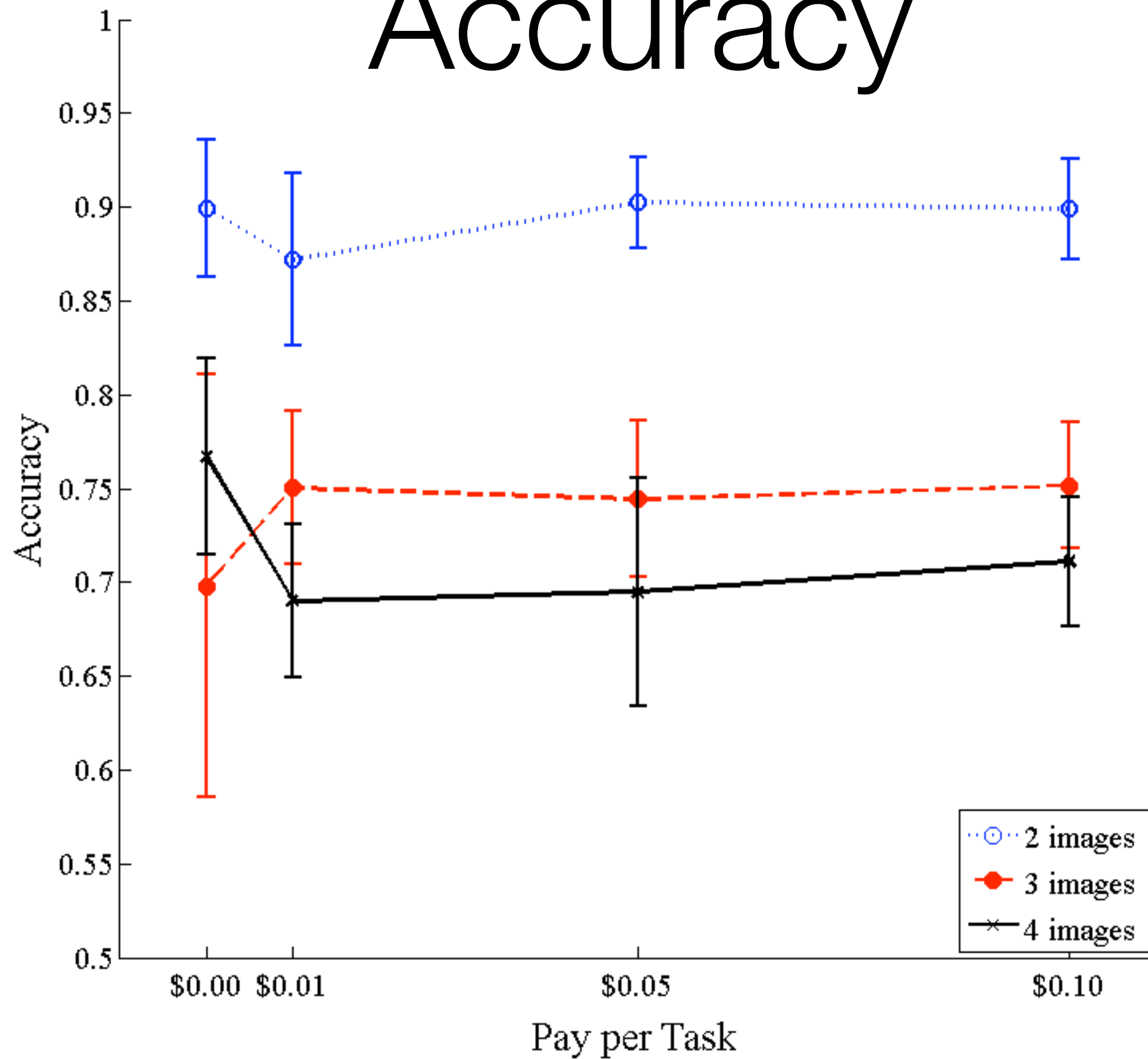
Re-order Traffic Images

Unsorted

Sorted

Payment scheme

- Everyone: $0.10 for doing training examples and filling out a survey

- Payment levels: nothing, 1¢, 5¢, 10¢ per set

- Num images per set (independent of payment): 2, 3, 4

- Each person sorted up to 99 sets of images, could end participation at any point and get paid for what they did

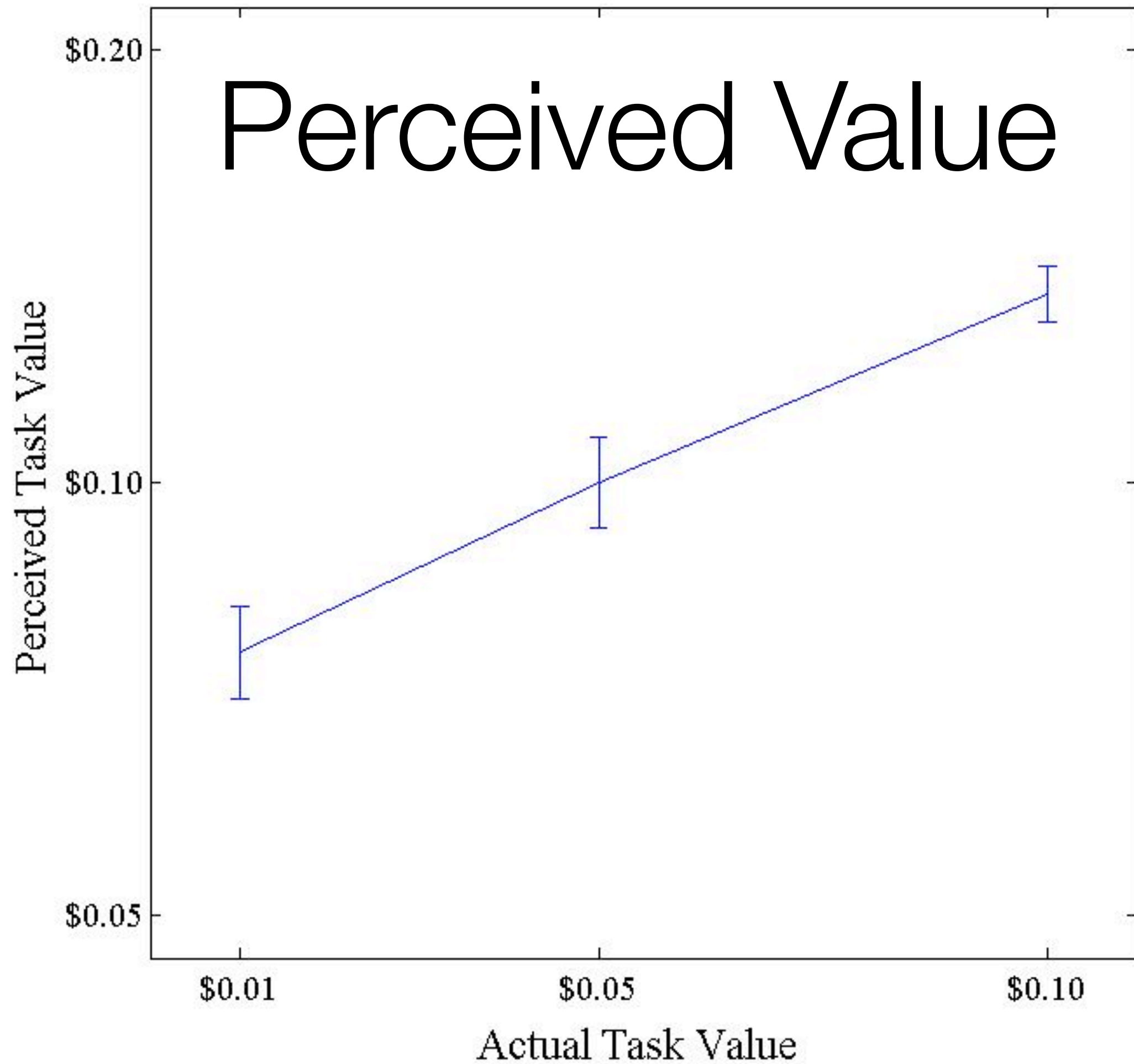- 611 subjects sorted a total of 36,425 image sets

Number of tasks done

Word Jumble Puzzles

| 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	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

- Find as many of the of words in a set as you can:

- ACHIEVE, ATTAIN, BUILDING, CHAIR, COMPLETE, GREEN, LAMP, MASTER, MUSIC, PLANT, STAPLE, STEREO, STRIVE, SUCCEED, TURTLE

- Not all of the words listed are in the puzzle!

Experimental setup

- Different pay rates (just as before)

- Subjects were told that they would be paid either on a per-grid basis or a per-word basis, or not told anything

- quantity = number of puzzles completed quality = fraction of words found per puzzle

- Participants could do up to 24 puzzles

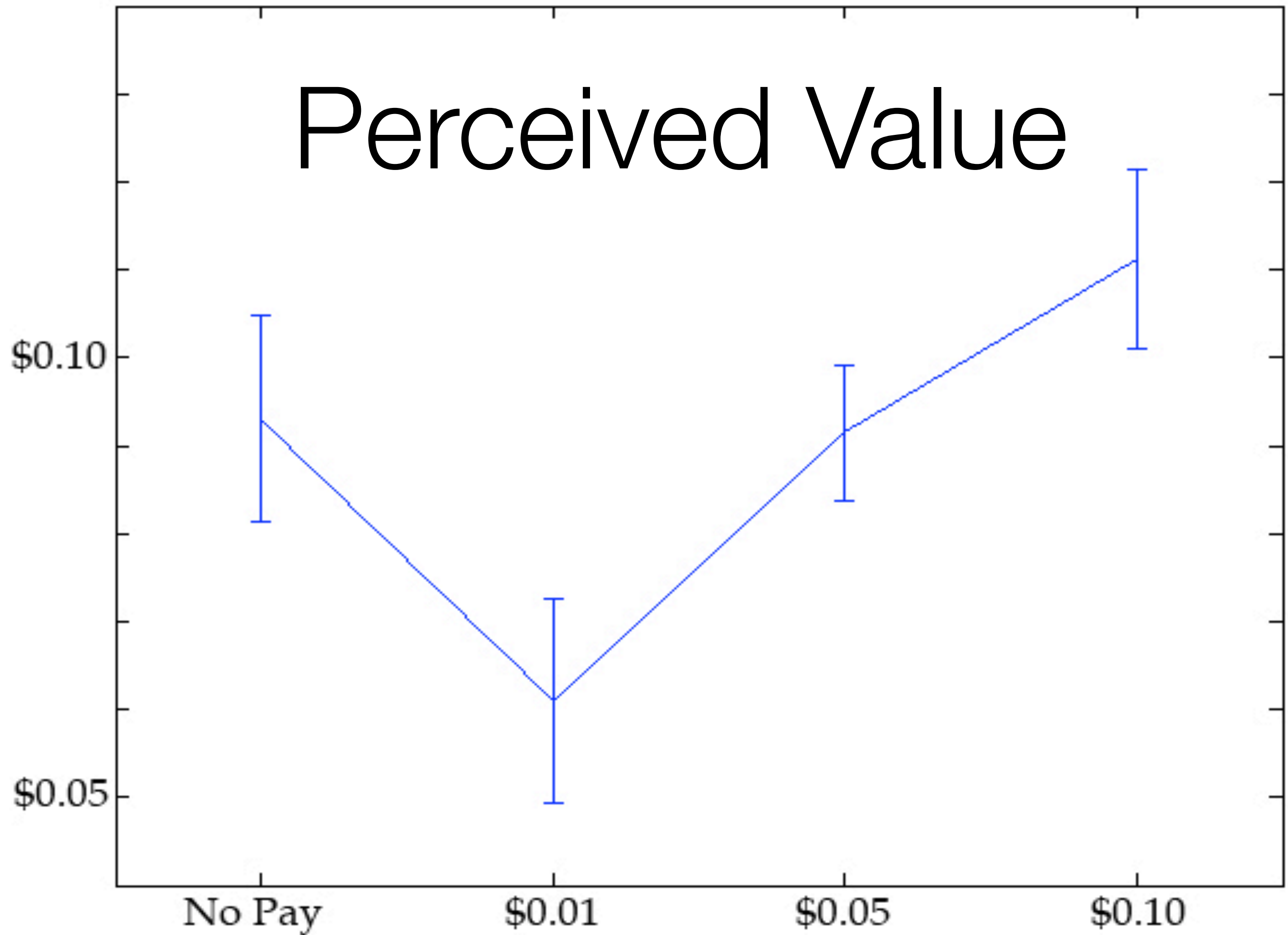- 320 subjects solved 2736 puzzles, finding 23,440 words

Fun v. pay

Perceived Value

$0.10

$0.05

No Pay $0.01 $0.05 $0.10
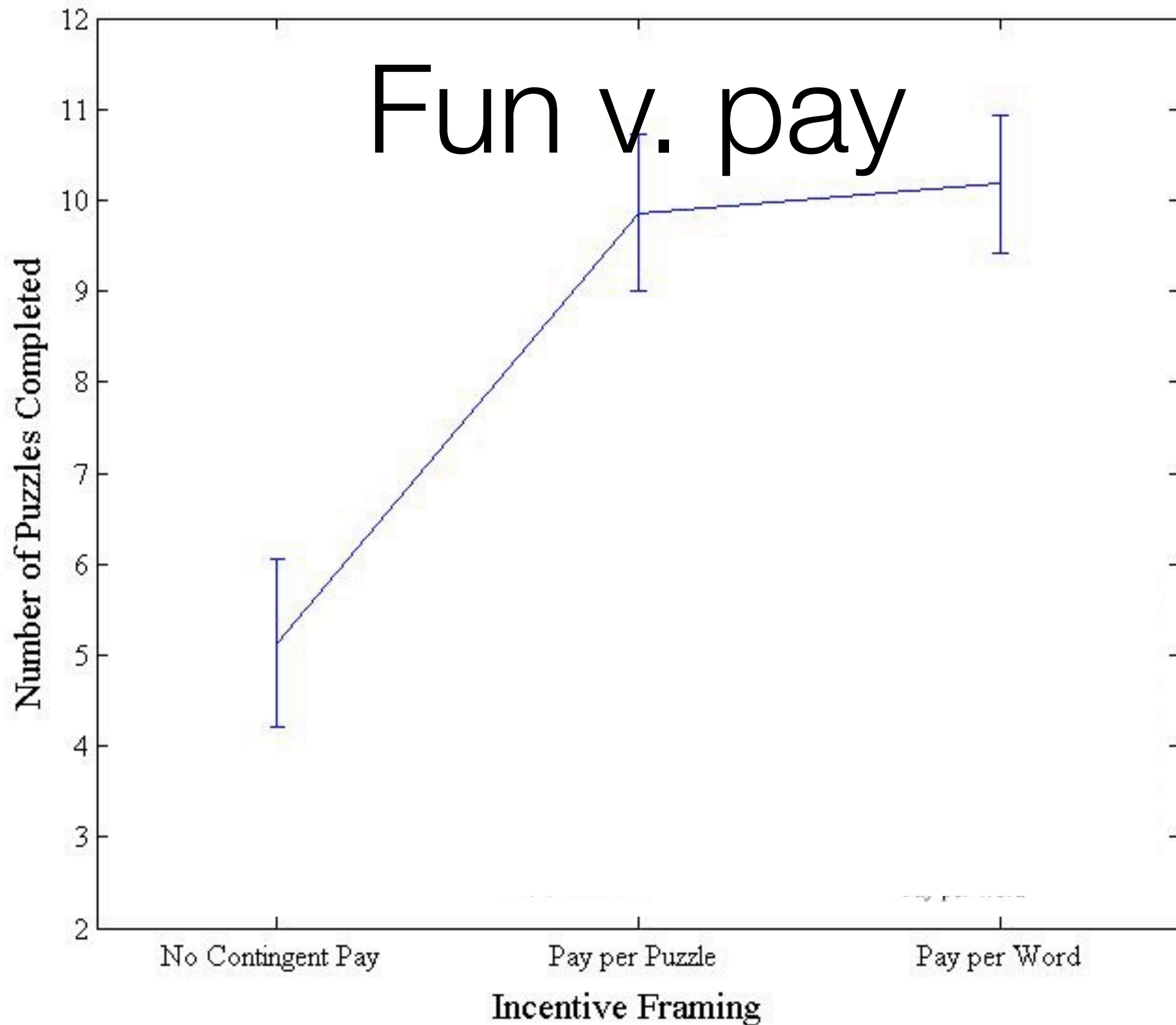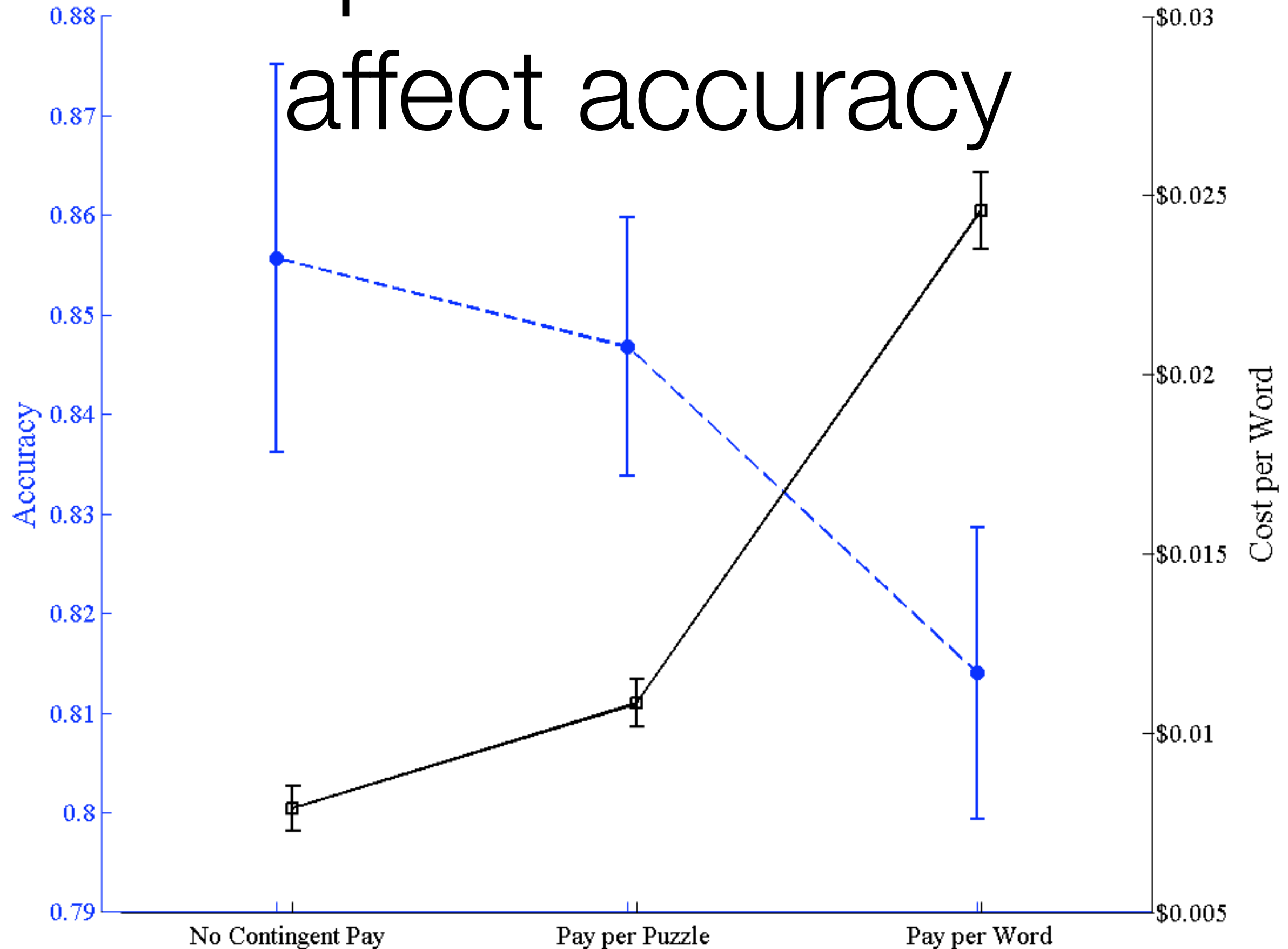
Findings

- Paying subjects elicited higher output than gamification, and increasing pay rate yielded even higher output

- However, paying subjects did not affect their accuracy

- Anchoring effects are significant – the reward you set impacts perceived value

Implications for your tasks?

- When you can use non-financial rewards, like intrinsic motivation, do so, since the quality of work will be the same

- When you can't use intrinsic motivation, it might be in your best interest to pay as little as possible. Your work will be done slower, but quality will be similar.

- Is this fair to workers?

What do you think?

- Is studying workers on Mechanical Turk a valid way of studying other labor markets?

- What possible confounds are there?

- What could we do to control for them?

Different Mechanisms for Quality Control

- Aggregation and redundancy

- Embedded gold standard data

- Economic incentives

- **Reputation systems**

- Statistical models

Reputation systems

- Mechanical Turk uses a reputation system

- Each Turker has a small number of variables associated with them, that are exposed to Requesters

- Past approval rate

- Number of HITs approved

- Has masters qualification (photo moderation/categorization master)

Pros and Cons of MTurk's reputation system

Pros	Cons
Gives one bit information about what other Requesters thought of a Worker	Reasons for rejections not shared; Weights all Requesters equally
Allows you to select Amazon's master's qualification, which is given to experienced Workers	It is not clear who gets the master's qual. No way to share other qualifications.
	Asymmetric: applies only to Workers, with no way to rate Requesters

Confederated Trust

- Acceptance rate doesn't show how good a worker is at a particular task

- Qualifications may like the "photo moderation master's" show this

- However, there is no way to share this information with other requesters

- Lots of reinventing the wheel

Confederated Trust

- Do you think it would be useful to share qualifications among requesters?

- How would you do it?

Asymmetric reputation systems

- No way for Turkers to rate requesters, and see beforehand who is scrupulous

- Turkers have built their own external tools for this like TurkOpticon

- No way to see whether a Turkers high rating comes from good Requesters

Choose the best category for this government project (good english important)

Requester: ▼ The Public Group

HIT Expiration Date: Sep 10, 2013 (6 days

communicativity: ▮▮ 1.17 / 5

generosity : ▮▮▮ 1.73 / 5

fairness : ▮▮▮ 1.39 / 5

promptness : ▮▮▮ 1.86 / 5

What do these scores mean?

Scores based on **81 reviews**
Terms of Service violation flags: 0
Report your experience with this requester »

me Allotted: 60 minutes

tact Us | Careers at Amazon | Developers |
©2005-2013 Amazon.com, Inc. or its Affilia

Took a leap of faith on this requester and was rewarded with a %50 reject rate and a broken search feature and no feedback. Would not recommend, even if you have thousands of HITs under your belt to cushion the inevitable rejections.

Aug 29 2013 | KBH19 | flag | comment
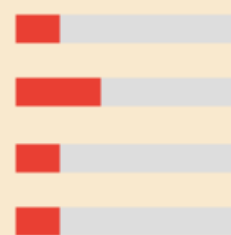
Arbitrarily rejected over half of the hits I submitted, and then banned me from submitting any more hits for them. I suppose that's a blessing in disguise though, as I had no intention of doing any for them again after the first batch of rejections.

Aug 21 2013 | bour...@g... | flag | comment
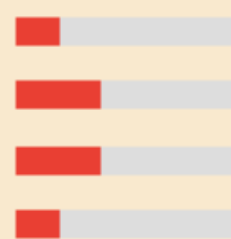
Their HIT is very unclear. There is an option to browse for the result, but it does not work.

Aug 20 2013 | jeff...@g... | flag | comment

qualitative v quantitative

TurkOpticon's qualitative attributes	CrowdWorker's quantitative equivalents
promptness: How promptly has this requester approved your work and paid?	**Expected time to payment:** On average, how much time elapses between submitting work to this Requester and receiving payment?
generosity: How well has this requester paid for the amount of time their HITs take?	**Average hourly rate:** What is the average hourly rate that other Turker make when they do this requester's HITs?
fairness: How fair has this requester been in approving or rejecting your work?	**Approval/rejection rates:** What percent of assignments does this Requester approve? What percent of first-time Workers get any work rejected?
communicativity: How responsive has this requester been to communications or concerns you have raised?	**Reasons for rejection:** Archive of all of the reasons for Workers being rejected or blocked by this Requester.

Amazon's other reputation system

- Amazon has another reputation system in place for its online stores

- Amazon allows anyone to list and sell items through its site, and to set their own prices

- These can be individuals selling used goods, or independent 3rd party sellers who use Amazon to reach a larger customer base

- How does Amazon ensure good customer experience?

Feedback from buyers

- How satisfied were you with how your order was packaged and shipped?

- If you contacted the third-party seller, did you get good customer service and prompt resolution?

- Would you buy from this third-party seller again?

Westinghouse Lighting 7214100 Harmony Two-Light 48-Inch Two-Blade Indoor Ceiling Fan, Brushed Nickel with Opal Frosted Glass

by [Westinghouse](#)

★★★★☆ | ([42 customer reviews](#)) | [11 answered questions](#)

Price + Shipping	Condition	Seller Information	Buying Options
$128.69 ✓ Prime	**New**	amazon.com. In Stock. • Free Two-day Shipping: Get it Wednesday, October 2 (order within) Domestic shipping rates and return policy.	🛒 **Add to cart** or Turn on 1-Click to use your Amazon Prime benefits.
$128.69 + $24.32 shipping	**New**	**PlumberSurplus** Point. Click. Plumb.™ ⭐⭐⭐⭐ **91% positive** over the past 12 months. (12,817 total ratings) Ships in 1-2 business days. Expedited shipping available. Domestic shipping rates and return policy.	🛒 **Add to cart** or Sign in to turn on 1-Click ordering.
$128.69 + $24.32 shipping	**New**	**Remodelr** ⭐⭐⭐⭐ **90% positive** over the past 12 months. (1,855 total ratings) Ships in 1-2 business days. Domestic shipping rates and return policy.	🛒 **Add to cart** or Sign in to turn on 1-Click ordering.
$148.99 + $24.19 shipping	**New**	UNBEATABLE **sale** ⭐⭐⭐⭐⭐ **97% positive** over the past 12 months. (163,508 total ratings) Usually ships within 3 - 4 business days. Domestic shipping rates and return policy.	🛒 **Add to cart** or Sign in to turn on 1-Click ordering.
$202.90 FREE Shipping	**New** New. 90 Day Money Back Guarantee, Caring Customer Service & Accurate Ship Times.	**DEL MAR** Fans & Lighting ⭐⭐⭐⭐⭐ **98% positive** over the past 12 months. (7,007 total ratings) Usually ships within 4 - 5 business days. Domestic shipping rates and return policy.	🛒 **Add to cart** or Sign in to turn on 1-Click ordering.

Recent Feedback: ★★★★☆

4.6 stars over the past 12 months (**573** ratings)

5/5: "good transaction, love the lock"
Sophia D., September 22, 2013

5/5: "Awesome experience "
Yadira Morejon, September 22, 2013

5/5: "Exactly what I needed, especially the color matched perfectly. Thank you."
mufasa, September 20, 2013

5/5: "Item was as described"
Thomas F., September 20, 2013

5/5: "Great seller, great item! Fast service too!!"
DB, September 20, 2013

2/5: "arrived bent"
ATD, September 20, 2013

Seller Response: We were not aware of any issue involving this customer's order. We have reached out to the customer to see how we may assist them in a return for a full refund, or a replacement of the damage item.
Date: September 23, 2013

5/5: "Shipping was a little pricey."

Kylie S., September 19, 2013

5/5: "Good price, high quality."
Joanna wang, September 18, 2013

5/5: "item was as descried seller promt whith sevirce"
Thomas Howell, September 16, 2013

5/5: "works great "
Spencer , September 16, 2013

5/5: "just as described"
Theresa M., September 15, 2013

5/5: "Item was as described, value priced and works great."
Amanda S., September 14, 2013

1/5: "When a seller charges $29.95 in shipping for a package weighing .2 lbs, they are gouging. The sponges cost less than $18. I assume their profit is from the shipping. I could get this shipped for less than $6.00! Never again"
Lana L Miller, September 14, 2013

Seller Response: We apologize the customer is not satisfied with the shipping charges. We have reached out to the customer and offered a discount to them as a one time courtesy.
Date: September 17, 2013

What are the economic implications of poor feedback?

$128.69
+ $24.32 shipping

PlumberSurplus
Point. Click. Plumb.™

91% positive over the past 12 months. (12,817 total ratings)

Ships in 1-2 business days. Expedited shipping available.
Domestic shipping rates and return policy.

$128.69
+ $24.32 shipping

Remodelr

90% positive over the past 12 months. (1,855 total ratings)

Ships in 1-2 business days.
Domestic shipping rates and return policy.

$148.99
+ $24.19 shipping

UNBEATABLE SALE

97% positive over the past 12 months. (163,508 total ratings)

Usually ships within 3 - 4 business days.
Domestic shipping rates and return policy.

$202.90
FREE Shipping

DEL MAR
Fans & Lighting

98% positive over the past 12 months. (7,007 total

Price premium

- Multiple sellers all selling the same item, but at different prices

- Price premium is the difference between a cheaper listing and a more expensive listing

- When someone opts for the more expensive item, even though it is identical, what is the reason for paying the premium?

Data-driven analysis

- Panos Ipeirotis (of mturk-tracker fame) harvested data from Amazon's website

- Gathered **transaction data** by repeatedly visiting listings (every 8 hours) and tracking when one item sold

- Gathered **reputation data** for each merchant. Complete history of numerical scores and text-based feedback

Data-driven analysis

- Data set gathered over half a year period

- Transaction data contains 1,078 merchants, 9,484 unique transactions and 107,922 price premiums

- Reputation data contains an average of 4,932 postings for each merchant

NLP + Economics

- Quantify the economics impact of sentiment of the feedback evaluations

- Using NLP techniques to derive semantic orientation and strength of comments

Method

- Each merchant's reputation is represented using a vector of n-dimensions $X = (X_1, X_2, ... X_N)$

- Dimensions were 150 nouns and verbs, values of dimensions could be one of 140 modifiers

- X_1 is "delivery," X_2 is "packaging," X_2 is "service."

- Feedback 1 "I was impressed by the speedy delivery! Great service!": (speedy ; NULL; great)

- Feedback 2 "The item arrived in awful packaging, and the delivery was slow": (slow ; awful ; NULL)

Method

- Construct a matrix out of all of the feedback for a seller

- Weight the more recent feedback more heavily

- Calculate how the values of each dimension effect the price premium

- Use least-squares regression with fixed effects to predict the price premium

Highest scoring phrases

wonderful experience	$5.86	never received	-$7.56
outstanding seller	$5.76	defective product	-$6.82
excellant service	$5.27	horrible experience	-$6.79
lightning delivery	$4.84	never sent	-$6.69
highly recommended	$4.15	never recieved	-$5.29
best seller	$3.80	bad experience	-$5.26
perfectly packaged	$3.74	cancelled order	-$5.01
excellent condition	$3.53	never responded	-$4.87
excellent purchase	$3.22	wrong product	-$4.39
excellent seller	$2.70	not as advertised	-$3.93
excellent communication	$2.38	poor packaging	-$2.92
perfect item	$1.92	late shipping	-$2.89
terrific condition	$1.87	wrong item	-$2.50
top quality	$1.67	not yet received	-$2.35
awesome service	$1.05	still waiting	-$2.25
A+++ seller	$1.03	wrong address	-$1.54
great merchant	$0.93	never buy	-$1.48

Predicting the merchant who makes the sale