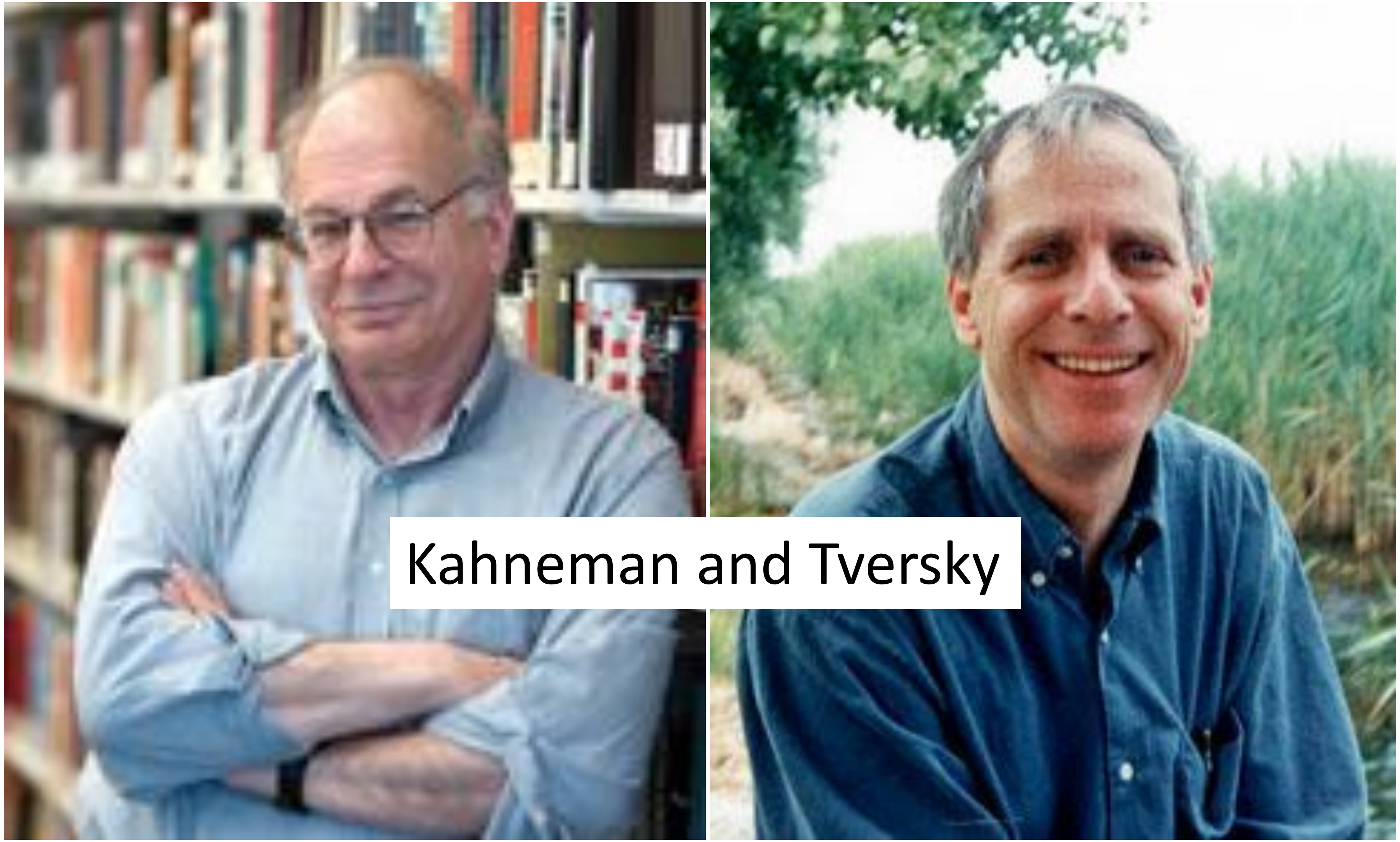
Lecture 7:
CS 5306 / INFO 5306:
Crowdsourcing and
Human Computation

What Do We Know about Amazon Mechanical Turk?

What Do We Know about People?



Kahneman and Tversky

Judgment under Uncertainty: Heuristics and Biases

Biases in judgments reveal some heuristics of thinking under uncertainty.

Amos Tversky and Daniel Kahneman

Many decisions are based on beliefs concerning the likelihood of uncertain events such as the outcome of an election, the guilt of a defendant, or the future value of the dollar. These beliefs are usually expressed in statements such as "I think that . . . ," "chances are . . . ," "it is unlikely that . . . ," and so forth. Occasionally, beliefs concerning uncertain events are expressed in numerical form as odds or subjective probabilities. What determines such beliefs? How do people assess the probability of an uncertain event or the value of an uncertain quantity? This article shows that people rely on a limited number of heuristic principles which reduce the complex tasks of assessing probabilities and predicting values to simpler judgmental operations. In general, these heuristics are quite useful, but sometimes they lead to severe and systematic errors.

The subjective assessment of probability resembles the subjective assessment of physical quantities such as distance or size. These judgments are all based on data of limited validity, which are processed according to heuristic rules. For example, the apparent distance of an object is determined in part by its clarity. The more sharply the object is seen, the closer it appears to be. This rule has some validity, because in any given scene the more distant objects are seen less sharply than nearer objects. However, the reliance on this rule leads to systematic errors in the estimation of distance. Specifically, distances are often overestimated when visibility is poor because the contours of objects are blurred. On the other hand, distances are often underestimated when visibility is good because the objects are seen sharply. Thus, the reliance on clarity as an indication of distance leads to common biases. Such biases are also found in the intuitive judgment of probability. This article describes three heuristics that are employed to assess probabilities and to predict values. Biases to which these heuristics lead are enumerated, and the applied and theoretical implications of these observations are discussed.

The authors are members of the department of psychology at the Hebrew University, Jerusalem, Israel.

Representativeness

Many of the probabilistic questions with which people are concerned belong to one of the following types: What is the probability that object A belongs to class B? What is the probability that event A originates from process B? What is the probability that process B will generate event A? In answering such questions, people typically rely on the representativeness heuristic, in which probabilities are evaluated by the degree to which A is representative of B, that is, by the degree to which A resembles B. For example, when A is highly representative of B, the probability that A originates from B is judged to be high. On the other hand, if A is not similar to B, the probability that A originates from B is judged to be low.

For an illustration of judgment by representativeness, consider an individual who has been described by a former neighbor as follows: "Steve is very shy and withdrawn, invariably helpful, but with little interest in people, or in the world of reality. A meek and tidy soul, he has a need for order and structure, and a passion for detail." How do people assess the probability that Steve is engaged in a particular occupation from a list of possibilities (for example, farmer, salesman, airline pilot, librarian, or physician)? How do people order these occupations from most to least likely? In the representativeness heuristic, the probability that Steve is a librarian, for example, is assessed by the degree to which he is representative of, or similar to, the stereotype of a librarian. Indeed, research with problems of this type has shown that people order the occupations by probability and by similarity in exactly the same way (*1*). This approach to the judgment of probability leads to serious errors, because similarity, or representativeness, is not influenced by several factors that should affect judgments of probability.

Insensitivity to prior probability of outcomes. One of the factors that have no effect on representativeness but should have a major effect on probability is the prior probability, or base-rate frequency, of the outcomes. In the case of Steve, for example, the fact that there are many more farmers than librarians in the population should enter into any reasonable estimate of the probability that Steve is a librarian rather than a farmer. Considerations of base-rate frequency, however, do not affect the similarity of Steve to the stereotypes of librarians and farmers. If people evaluate probability by representativeness, therefore, prior probabilities will be neglected. This hypothesis was tested in an experiment where prior probabilities were manipulated (*1*). Subjects were shown brief personality descriptions of several individuals, allegedly sampled at random from a group of 100 professionals—engineers and lawyers. The subjects were asked to assess, for each description, the probability that it belonged to an engineer rather than to a lawyer. In one experimental condition, subjects were told that the group from which the descriptions had been drawn consisted of 70 engineers and 30 lawyers. In another condition, subjects were told that the group consisted of 30 engineers and 70 lawyers. The odds that any particular description belongs to an engineer rather than to a lawyer should be higher in the first condition, where there is a majority of engineers, than in the second condition, where there is a majority of lawyers. Specifically, it can be shown by applying Bayes' rule that the ratio of these odds should be $(.7/.3)^2$, or 5.44, for each description. In a sharp violation of Bayes' rule, the subjects in the two conditions produced essen-

Kahneman and Twersky

Humans are subject to “architectural” errors

THE *NEW YORK TIMES* BESTSELLER

THINKING, FAST AND SLOW

DANIEL KAHNEMAN

WINNER OF THE NOBEL PRIZE IN ECONOMICS

"[A] masterpiece . . . This is one of the greatest and most engaging collections of insights into the human mind I have read." —WILLIAM EASTERLY, *Financial Times*

What Do We Know about People?

What Do We Know about Amazon Mechanical Turk?

What Do We Know about People Using Computers?


Cliff Nass

Computers Are Social Actors (CASA)

People interact with computers as if they are social actors

Theories about how people interact can apply
to how people interact with computers

The Media Equation

How People Treat Computers, Television, and New Media Like Real People and Places

Byron Reeves & Clifford Nass

Main Publication Venues

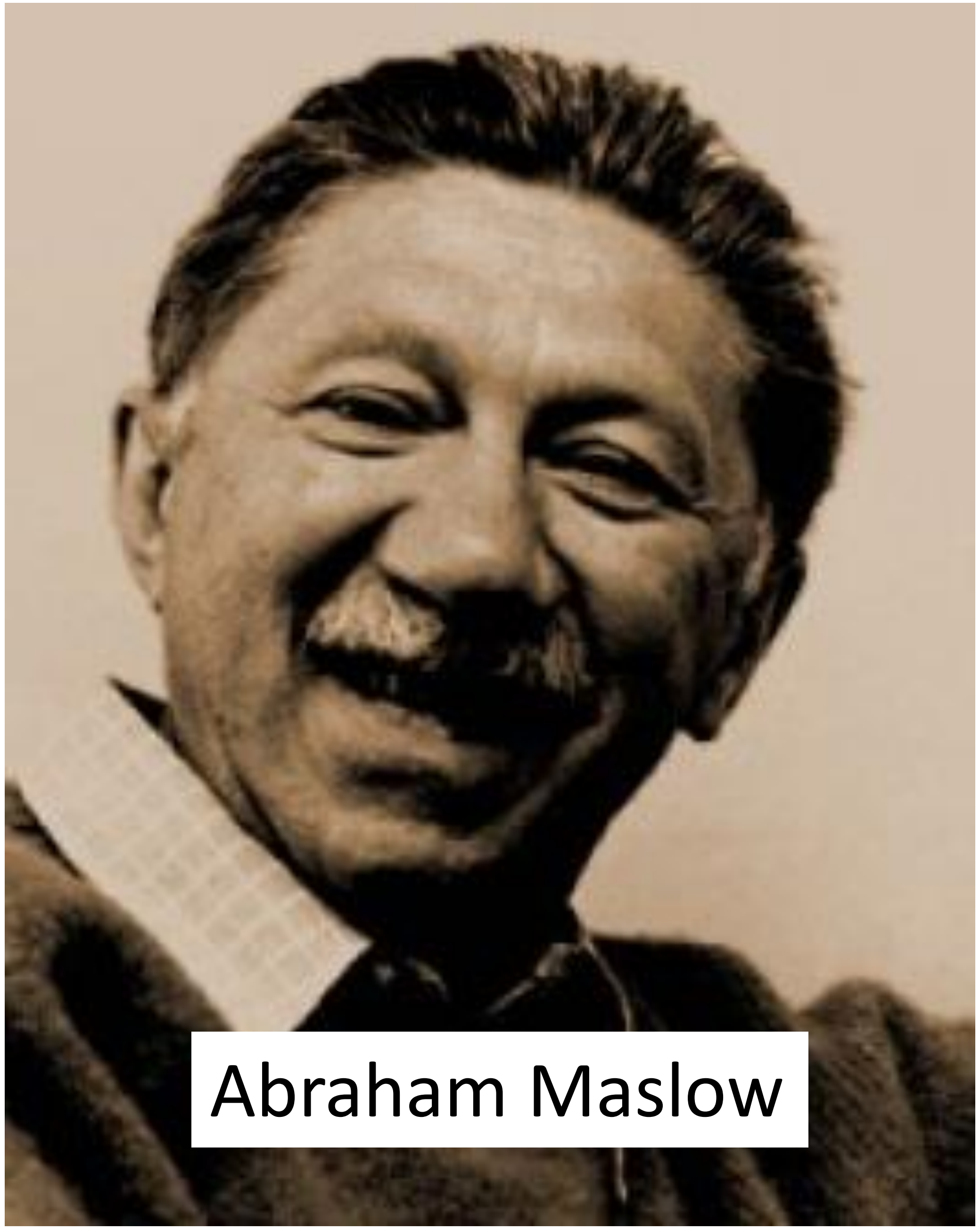
- CHI: ACM Conference on Human Factors in Computing Systems
- CSCW: ACM Conference On Computer-Supported Cooperative Work
- UIST: ACM Symposium on User Interface Software and Technology

What Do We Know about People Using Computers?

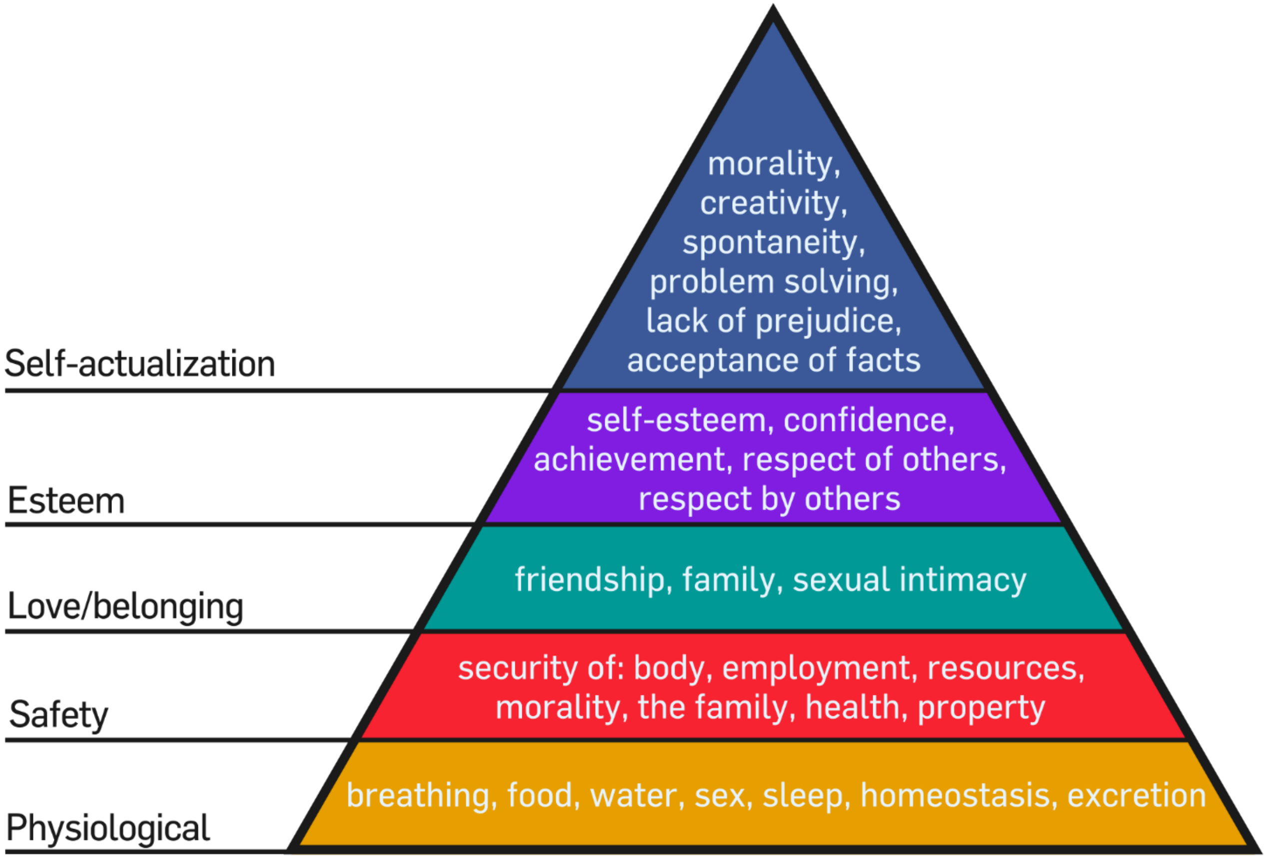
What Do We Know about Amazon Mechanical Turk?

What Do We Know about What Motivates Turkers?

What Do We Know about Motivation?


Abraham Maslow

Maslow’s Hierarchy of Needs



What Do We Know about What Motivates Turkers?

What Do We Know about What Motivates Open Source Programmers?

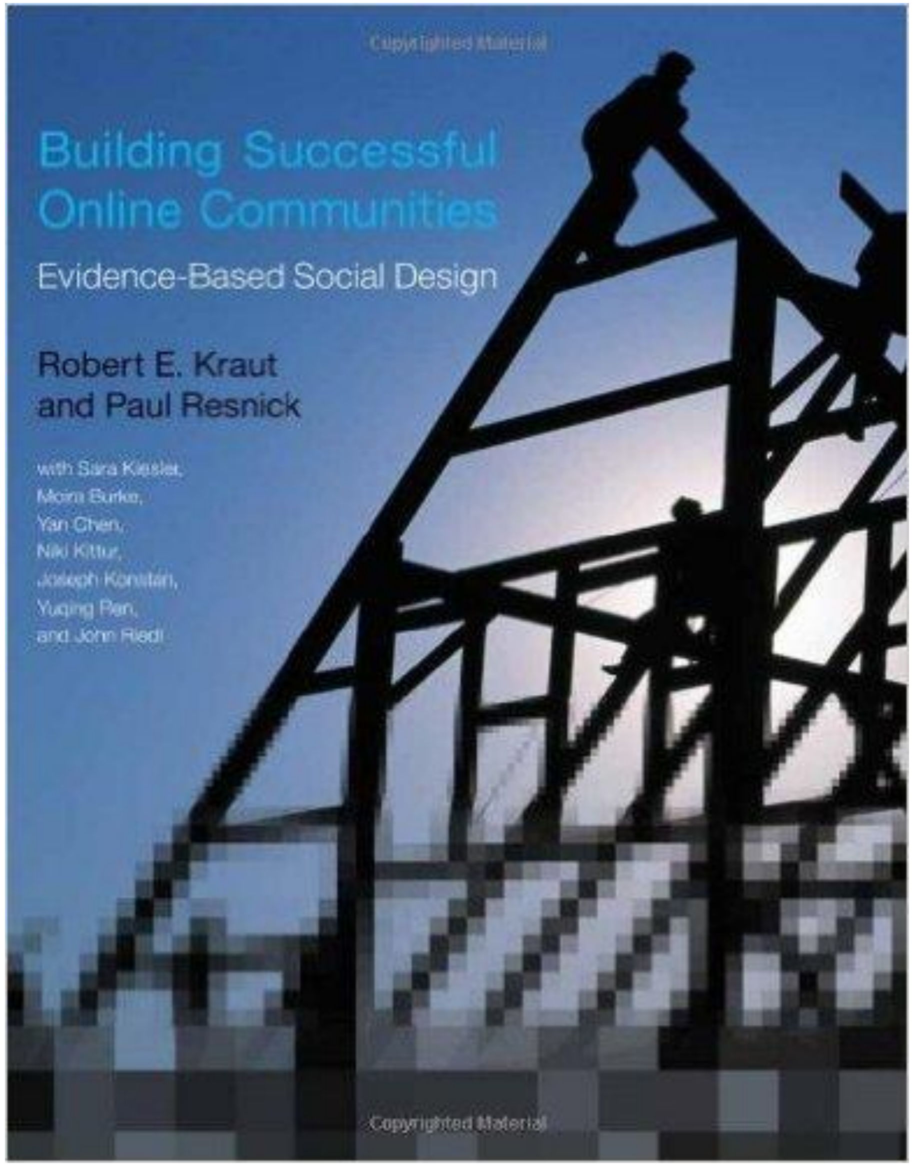
“Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Software Projects”

K.R. Lakhani and R.G. Wolf

Perspectives on Free and Open Source Software, MIT Press, 2005

Motivation	**% of respondents indicating up to 3 statements that best reflect their reasons to contribute (%)**	**% volunteer contributors**	**% paid contributor**	**Significant difference (t statistic/p value)**
Enjoyment based Intrinsic Motivation				
Code for project is intellectually stimulating to write	44.9	46.1	43.1	n.s.
Like working with this development team	20.3	21.5	18.5	n.s.
Economic/Extrinsic based Motivations				
Improve programming skills	41.3	45.8	33.2	3.56 (p=0.0004)
Code needed for user need (work and/or non-work)*	58.7	-	-	-
- Work need only	33.8	19.3	55.7	10.53 (p=0.0000)
- Non-work need	29.7	37.0	18.9	5.16 (p=0.0000)
Enhance professional status	17.5	13.9	22.8	3.01 (p=0.0000)
Obligation/Community based Intrinsic Motivations				
Believe that source code should be open	33.1	34.8	30.6	n.s.
Feel personal obligation to contribute because use F/OSS	28.6	29.6	26.9	n.s.
Dislike proprietary software and want to defeat them	11.3	11.5	11.1	n.s.
Enhance reputation in F/OSS community	11.0	12.0	9.5	n.s.

What Do We Know about What Motivates Participation in Online Communities?

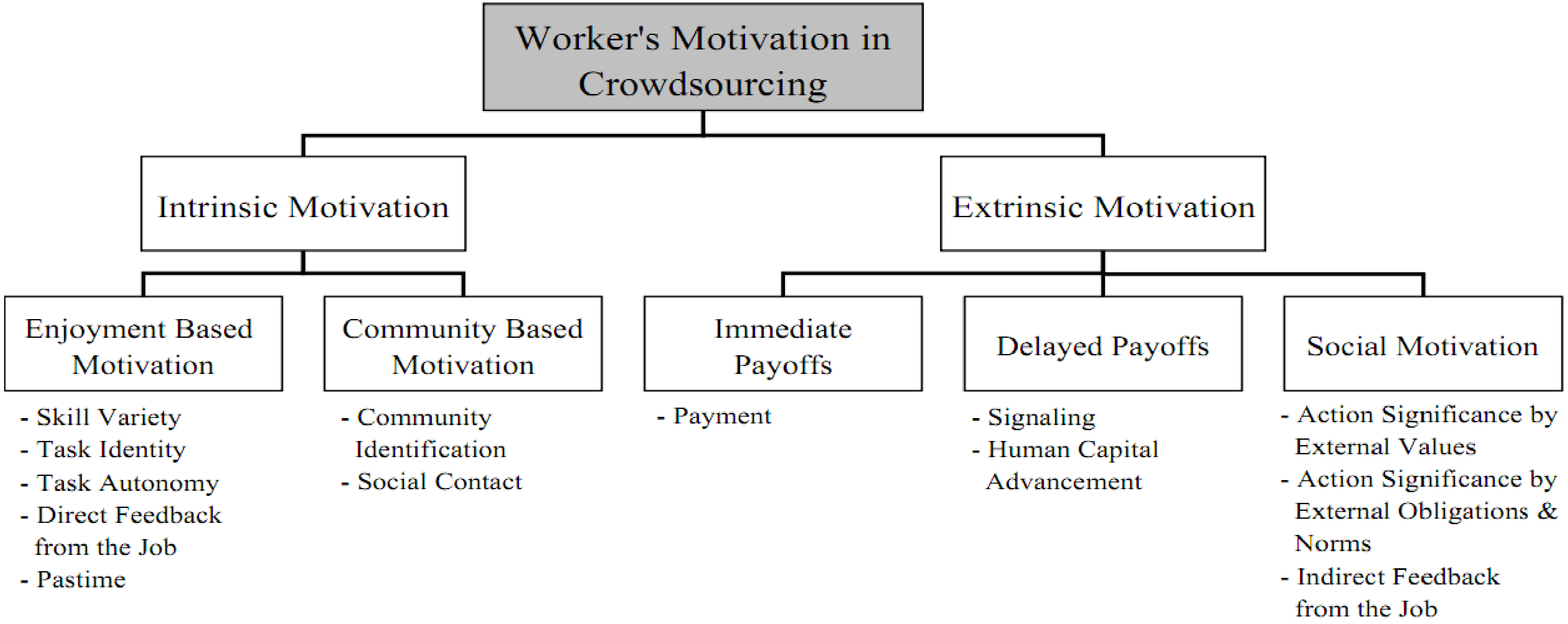

Copyrighted Material
Building Successful
Online Communities
Evidence-Based Social Design
Robert E. Kraut
and Paul Resnick
with Sara Kiesler,
Moira Burke,
Yan Chen,
Niki Kittur,
Joseph Konstan,
Yuqing Ren,
and John Riedl
Copyrighted Material

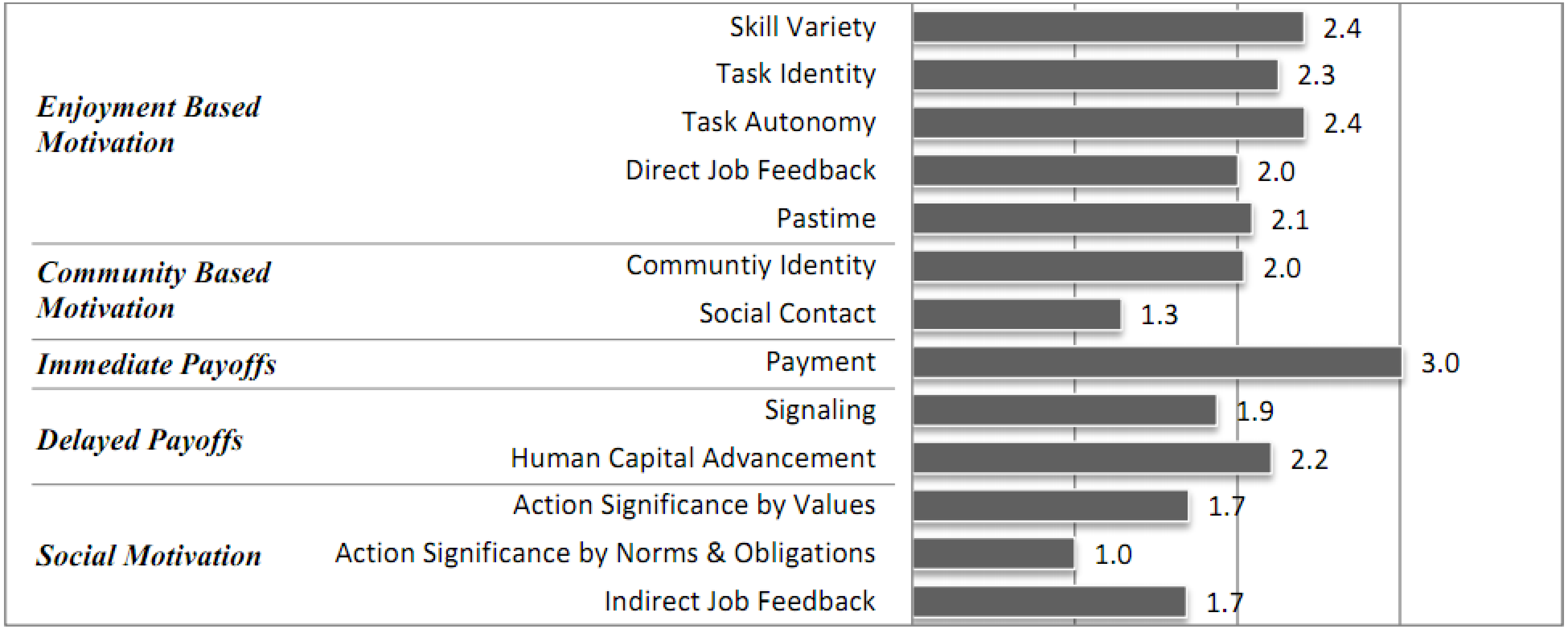
"More than fun and money.
Worker Motivation in Crowdsourcing – A Study on Mechanical Turk“

Nicolas Kaufmann, Thimo Schulze, and Daniel Veit

Proceedings of the 17th Americas Conference on Information Systems, 2011

		Intrinsic Motivation		Extrinsic Motivation		
Paper	**Focus**	**Enjoyment Based Motivation**	**Community Based Motivation**	**Immediate Payoffs**	**Delayed Payoffs**	**Social Motivation**
(Leimeister et al., 2009)	Idea Competitions	-	-	„Direct compensation“	„Learning“ „Self-Marketing“	„Social motives“
(Brabham, 2008)	Content Market	“Creative outlet”; “Fun”; “Produce [content] that I like”; “Passes the time when I am bored”	“Build a network of friends”	“Opportunity to make money”	“Improve skills” “Earn a reputation”	“Better way to make [content]” “Build a network with other creative people”
(Brabham, 2010)	Design Competition	-	“Love of community”; “‘Addiction’ to the community”	“Earn money”	“Improve creative skills” “Get employed as a freelancer”	-
(Ipeirotis, 2010)	Mechanical Turk	“Fruitful way to spend free time”; “To kill time”; “Tasks are fun”	-	“Primary source of income” “Secondary source of income”	-	-
(Organisciak, 2008)	Crowdsourcing	Fun; Boredom; achievement (by the action); Interest (curiosity)	Charity; Academia; Participation (Social Human Interaction)	Money	Self-Benefit (directly and indirectly from the action)	Forced

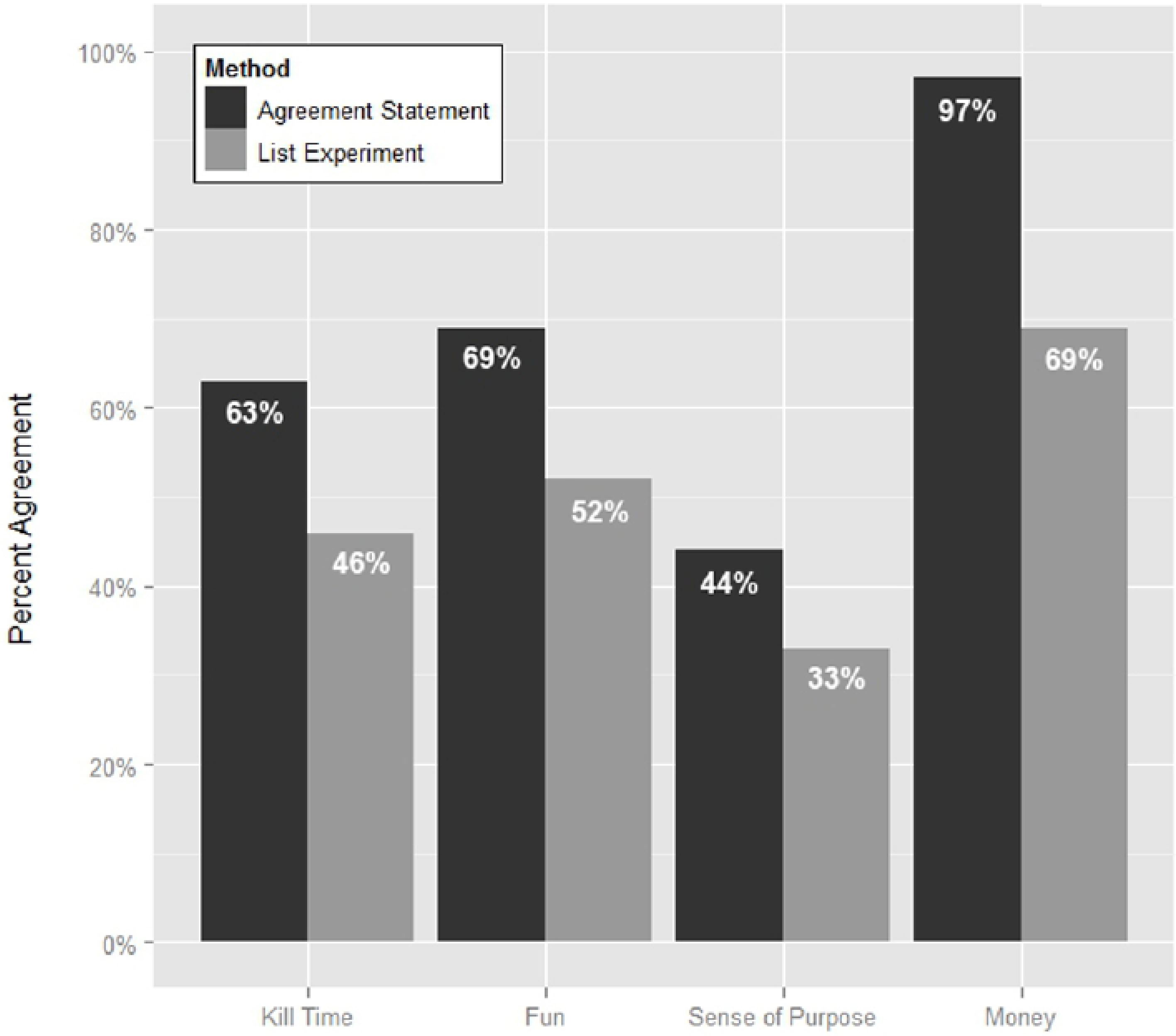

Worker's Motivation in Crowdsourcing
Intrinsic Motivation
Extrinsic Motivation
Enjoyment Based Motivation
Community Based Motivation
Immediate Payoffs
Delayed Payoffs
Social Motivation
- Skill Variety
- Task Identity
- Task Autonomy
- Direct Feedback from the Job
- Pastime
- Community Identification
- Social Contact
- Payment
- Signaling
- Human Capital Advancement
- Action Significance by External Values
- Action Significance by External Obligations & Norms
- Indirect Feedback from the Job


Enjoyment Based Motivation
Skill Variety 2.4
Task Identity 2.3
Task Autonomy 2.4
Direct Job Feedback 2.0
Pastime 2.1
Community Based Motivation
Communtiy Identity 2.0
Social Contact 1.3
Immediate Payoffs
Payment 3.0
Delayed Payoffs
Signaling 1.9
Human Capital Advancement 2.2
Social Motivation
Action Significance by Values 1.7
Action Significance by Norms & Obligations 1.0
Indirect Job Feedback 1.7

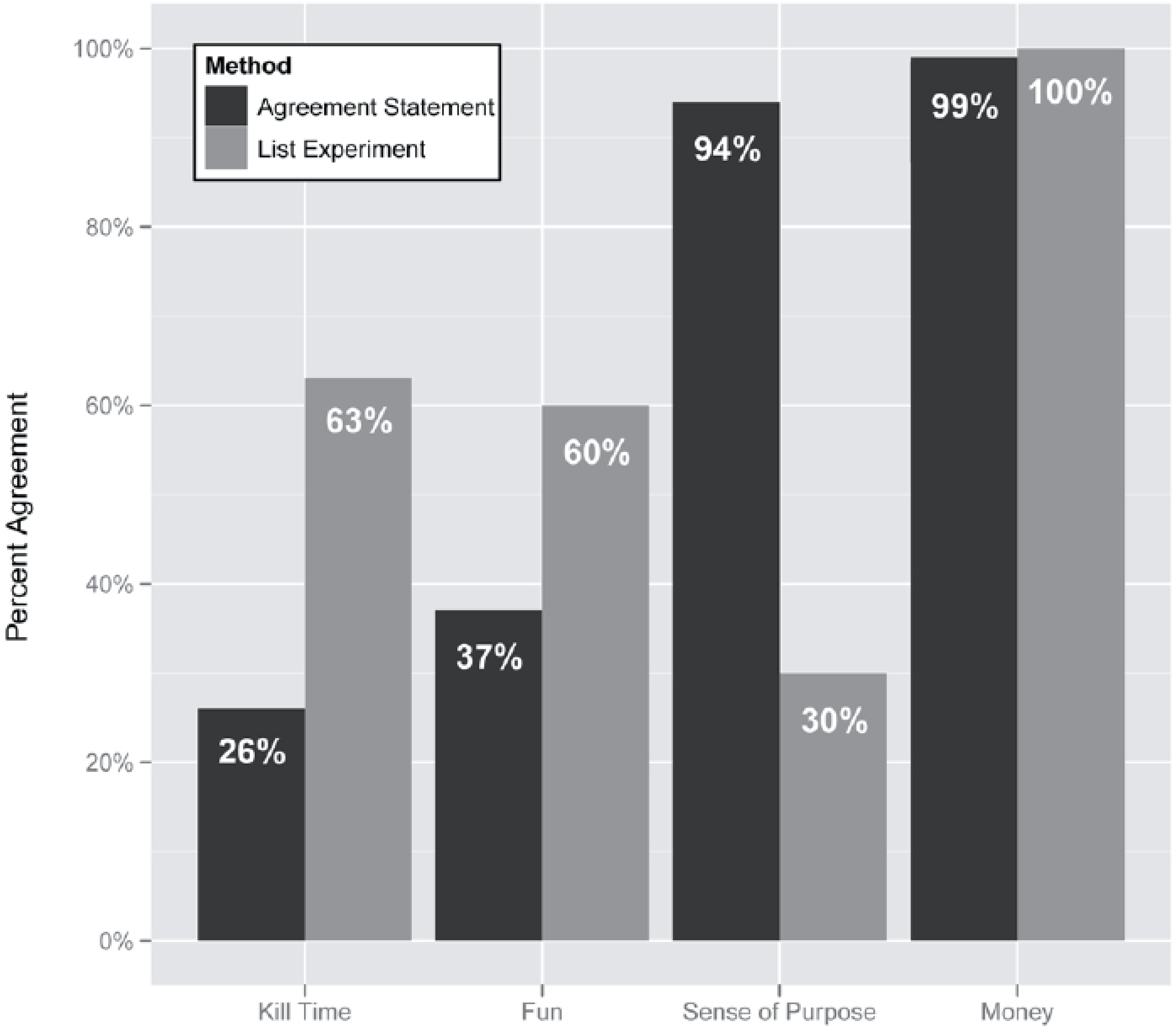
“Social Desirability Bias in Reports of Motivation for US and India Workers on Mechanical Turk“

Judd Antin and Aaron Shaw

CSCW 2011



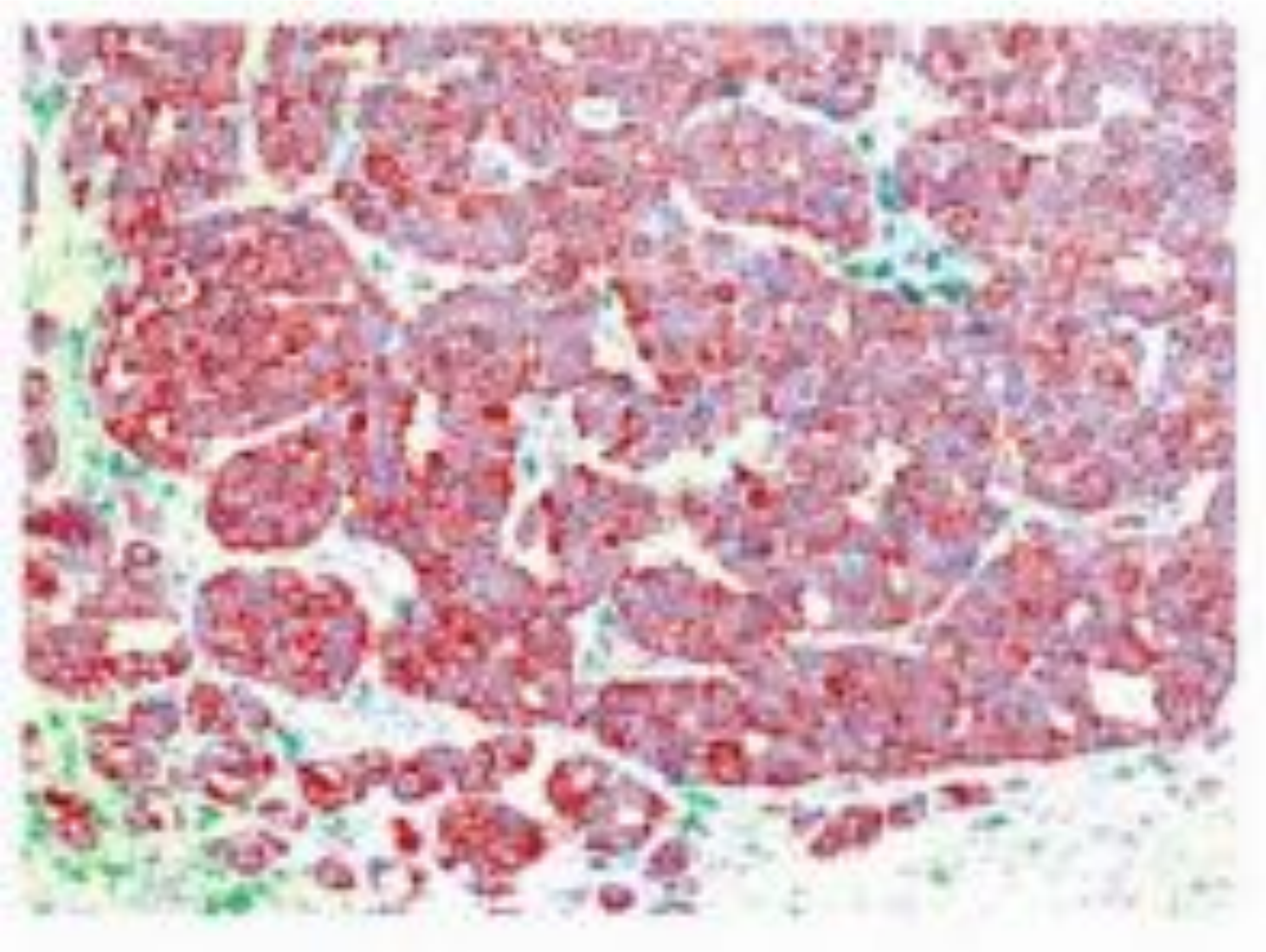
Method
Agreement Statement
List Experiment
Percent Agreement
100%
80%
60%
40%
20%
0%
63%
46%
69%
52%
44%
33%
97%
69%
Kill Time
Fun
Sense of Purpose
Money

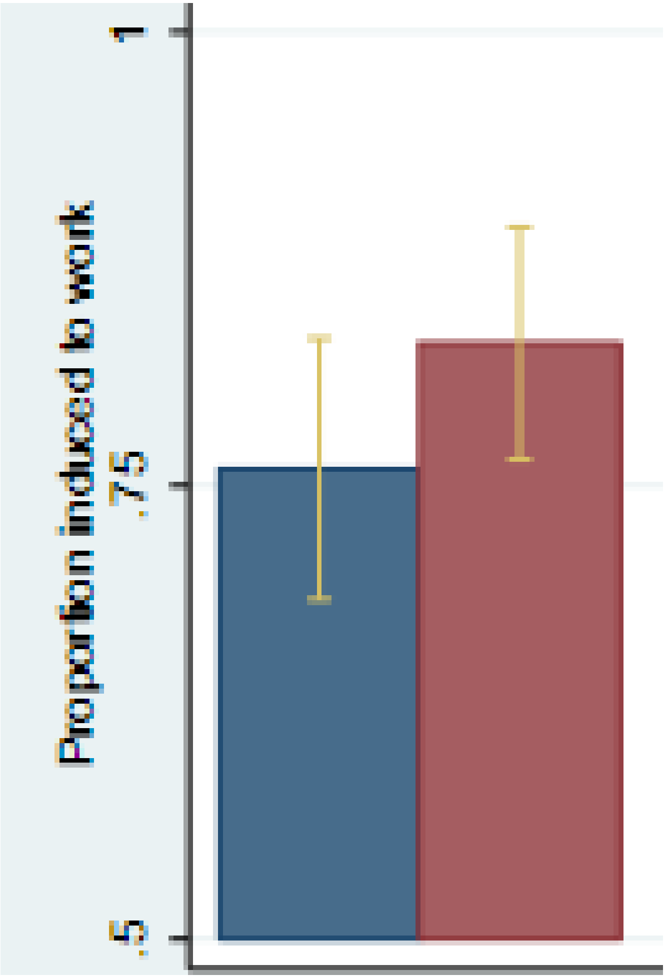

Method
Agreement Statement
List Experiment
Percent Agreement
100%
80%
60%
40%
20%
0%
26%
63%
37%
60%
94%
30%
99%
100%
Kill Time
Fun
Sense of Purpose
Money

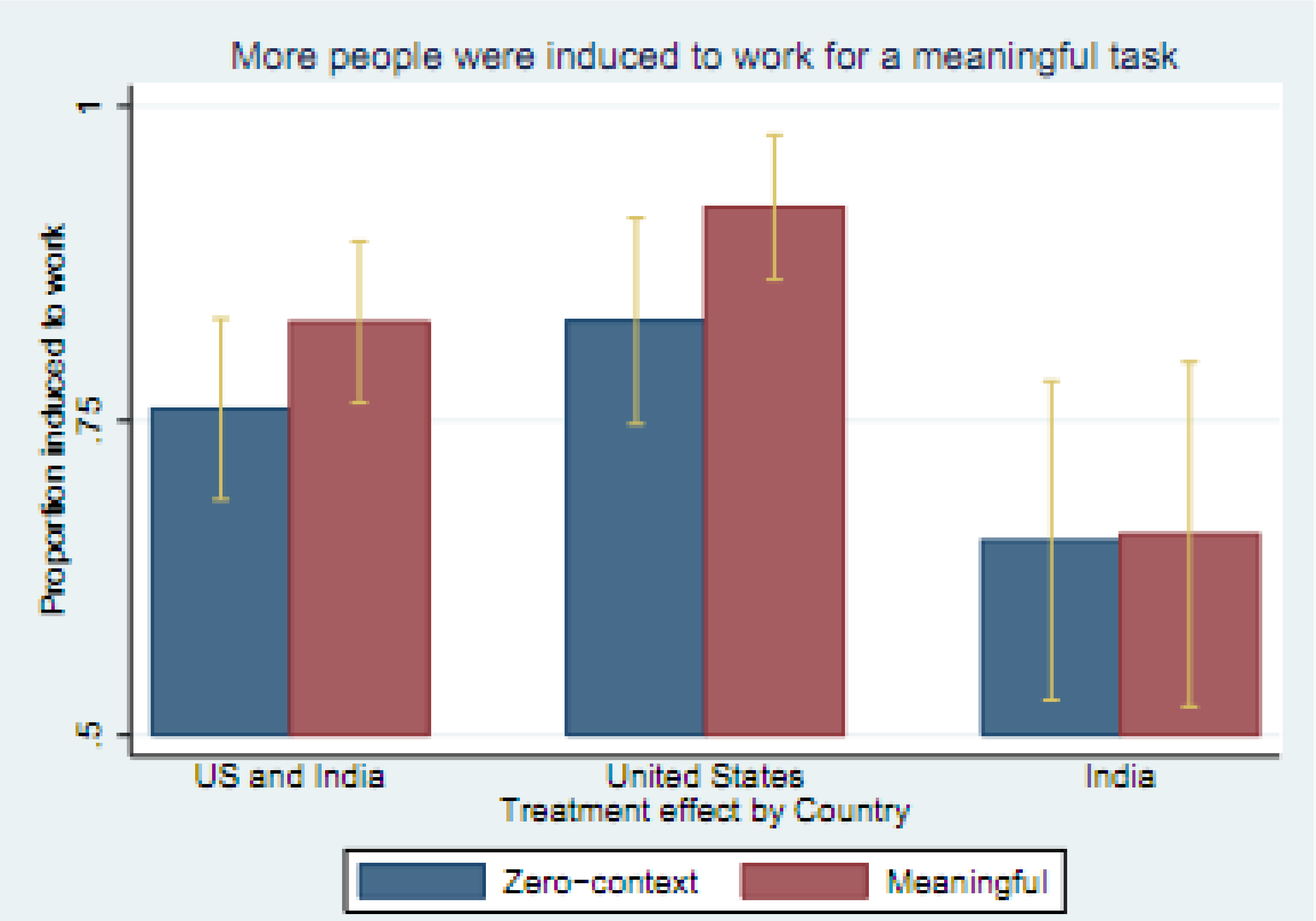
"Breaking Monotony with Meaning"

D. Chandler and A. Kapelner

Journal of Economic Behavior & Organization 90, 123-133 (2013)




Proportion induced to work
1
.75
.5


More people were induced to work for a meaningful task
Proportion induced to work
1
.75
.5
US and India
United States
India
Treatment effect by Country
Zero−context
Meaningful

“Financial Incentives and the `Performance of Crowds‘”

W. Mason and D. J. Watts

Proceedings of the First Workshop on Human Computation, 2009

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. 
2. 
3. 

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

1. 
2. 
3. 

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.

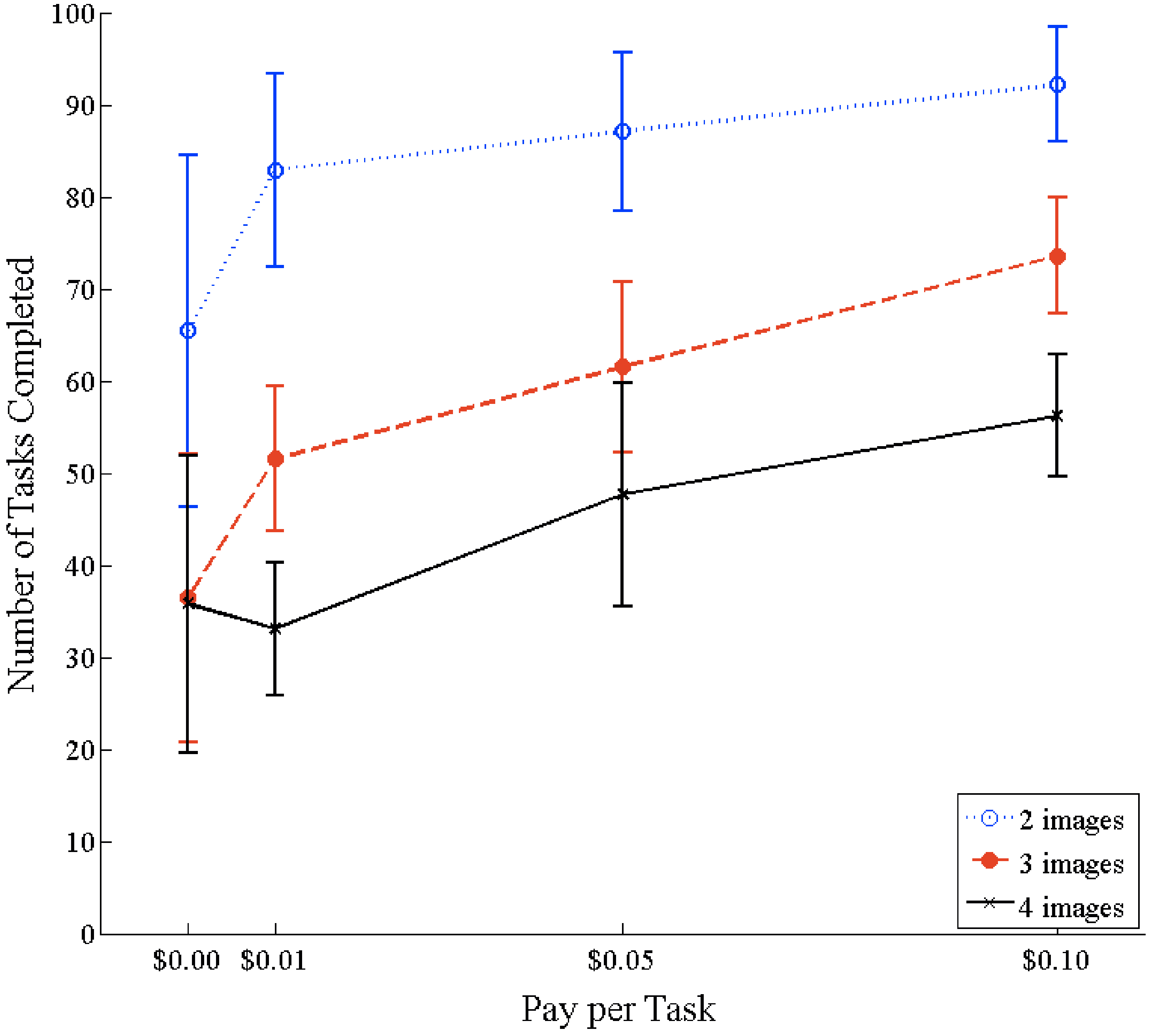
To reorder a list, click and drag a photo to the position it belongs. The other photos will move accordingly. Once you believe the list is in the correct order, click on the "Submit" button at the top of the page.

If you do not want to complete any more tasks, click on the "Finished" button at the bottom of the page.

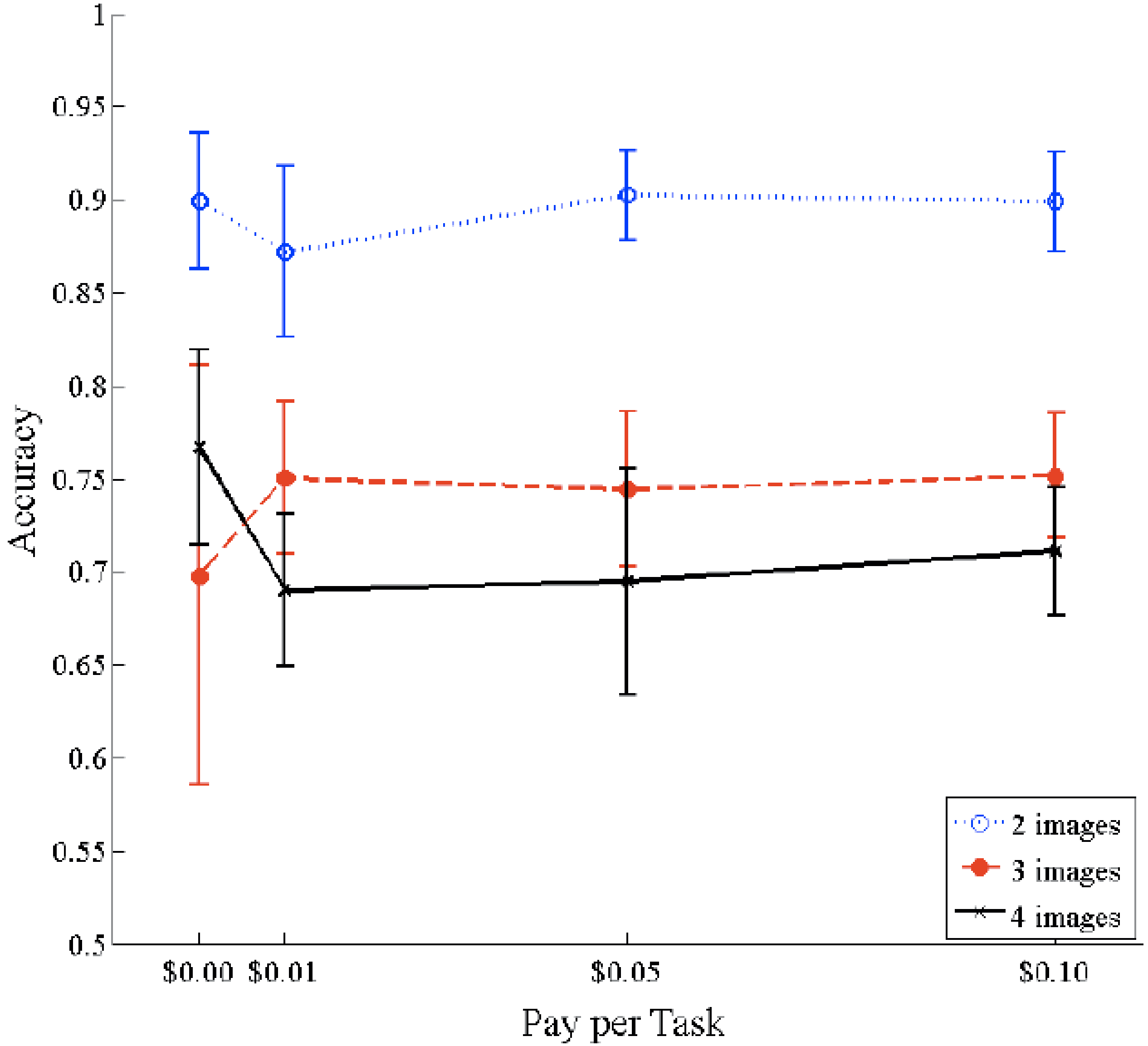
(This button will not be available in the next 3 practice examples)

Click here to practice: Practice

Images provided by Trafficland.com



Number of Tasks Completed
0
10
20
30
40
50
60
70
80
90
100
$0.00
$0.01
$0.05
$0.10
Pay per Task
2 images
3 images
4 images


1
0.95
0.9
0.85
0.8
0.75
0.7
0.65
0.6
0.55
0.5
Accuracy
$0.00
$0.01
$0.05
$0.10
Pay per Task
2 images
3 images
4 images

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.

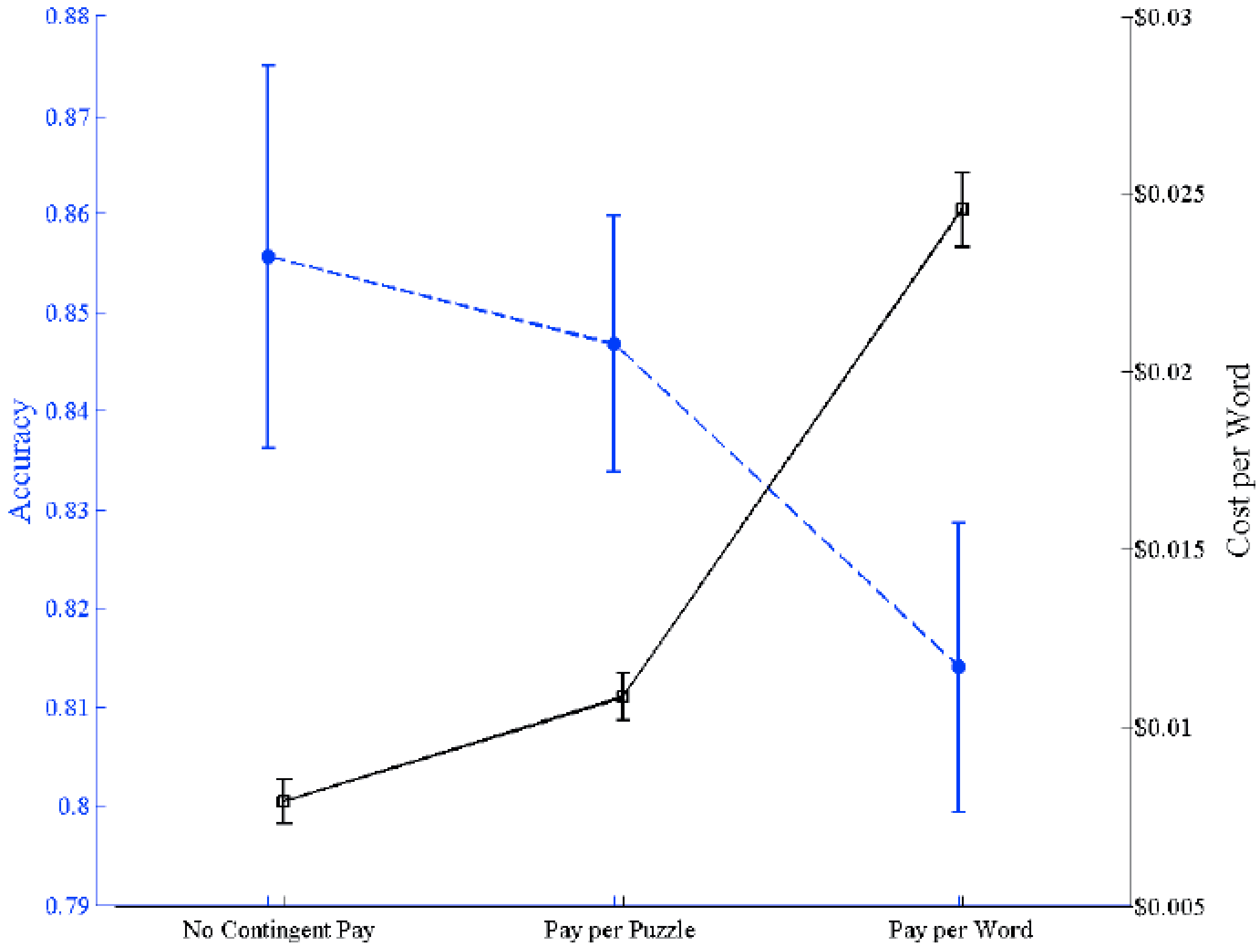
For each puzzle you will see a set of *possible* words and their category. **Not all of the words listed are in the puzzle!** In addition, the number of words in each puzzle changes. The list of *possible* words follows:
ACHIEVE, ATTAIN, BUILDING, CHAIR, COMPETE, GREEN, LAMP, MASTER, MUSIC, PLANT, STAPLE,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

RANDOM WORDS

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

SUCCEED
BUILDING
ACHIEVE

Submit Puzzle

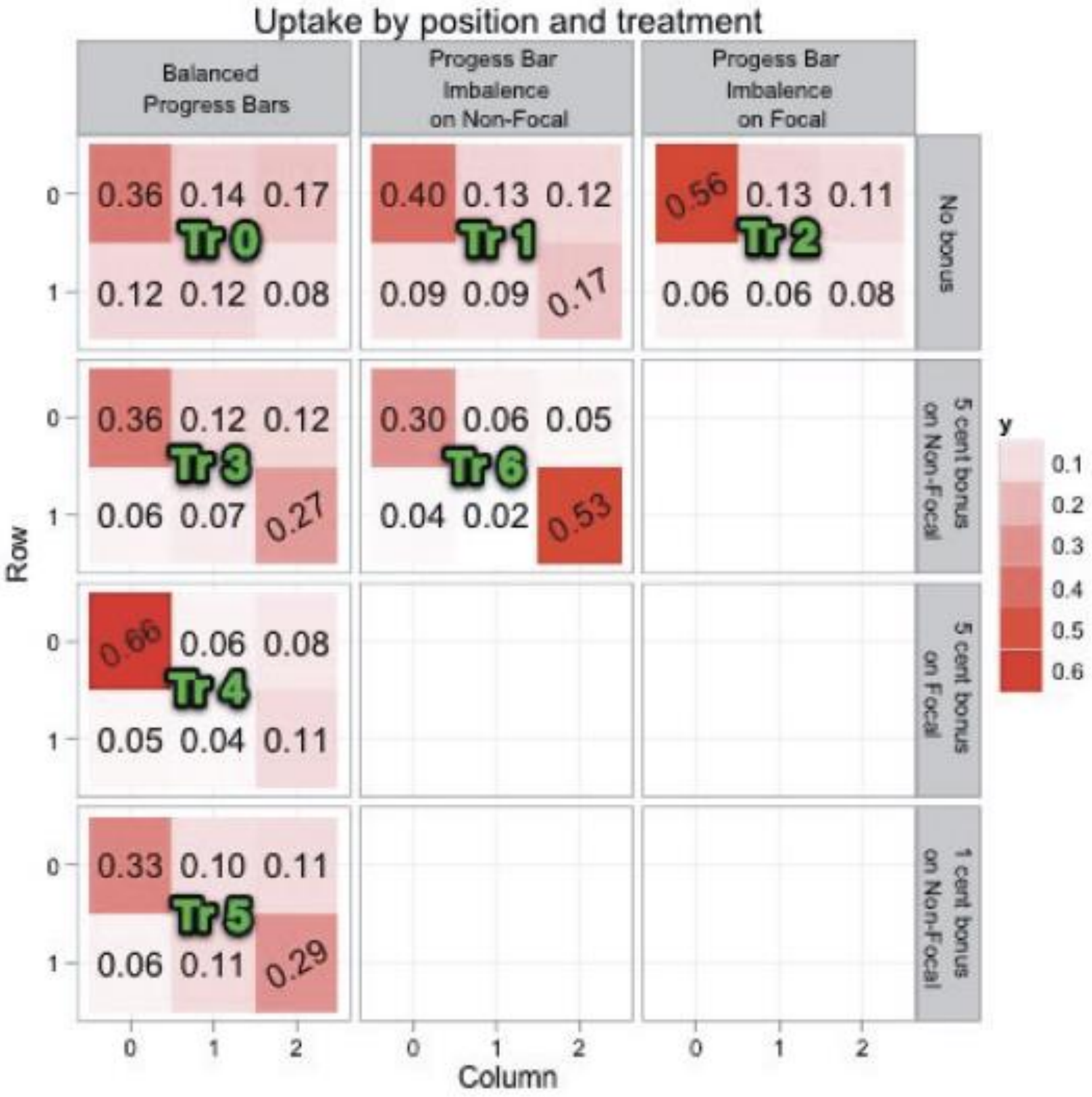


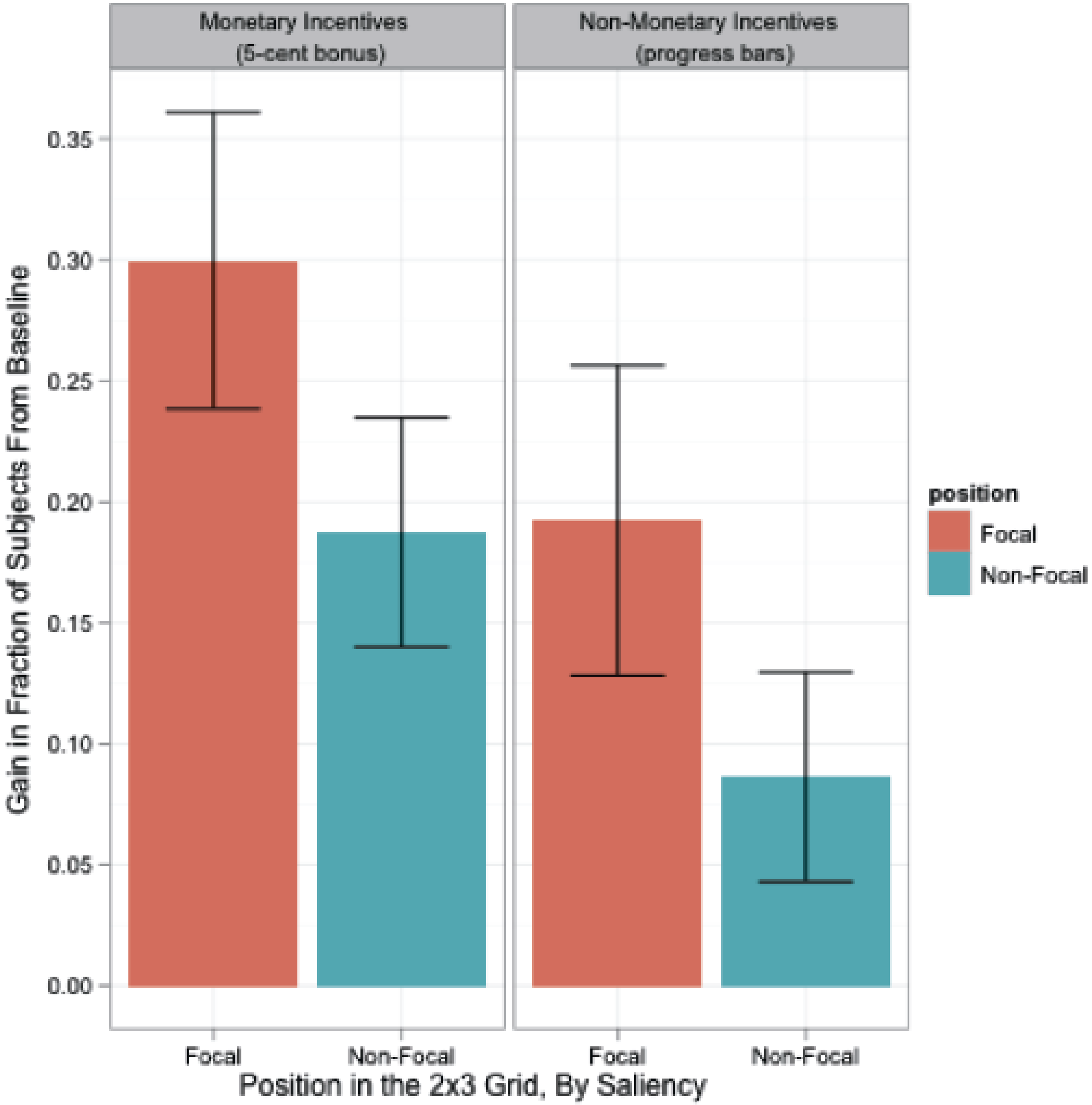
0.88
0.87
0.86
0.85
0.84
0.83
0.82
0.81
0.8
0.79
Accuracy
$0.03
$0.025
$0.02
$0.015
$0.01
$0.005
Cost per Word
No Contingent Pay
Pay per Puzzle
Pay per Word

"Labor Allocation in Paid Crowdsourcing:
Experimental Evidence on Positioning, Nudges and Prices“

Dana Chandler and John Horton

Proceedings of the Third Human Computation Workshop, 2011

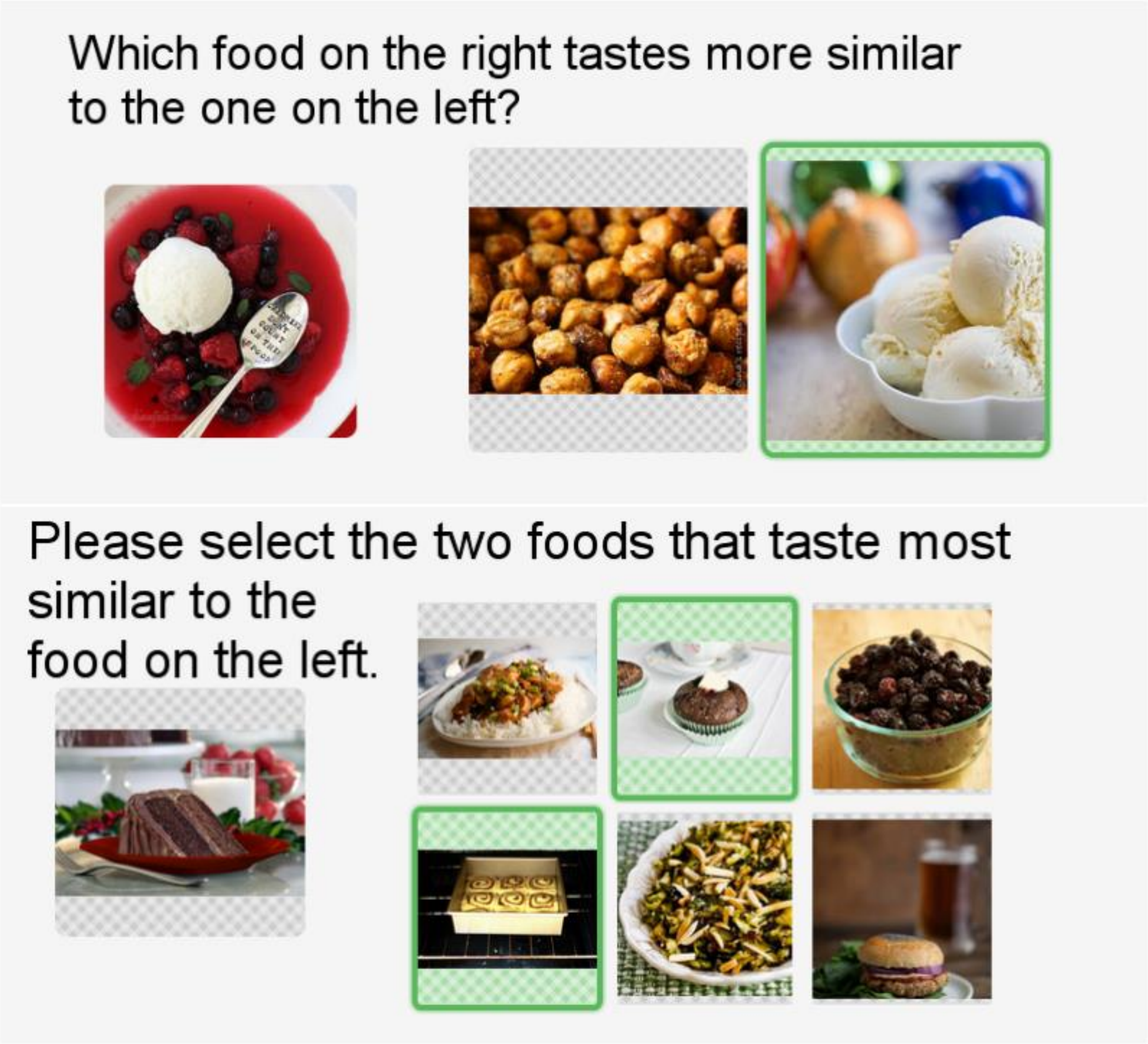

Uptake by position and treatment
Balanced Progress Bars
Progess Bar Imbalence on Non-Focal
Progess Bar Imbalence on Focal
No bonus
5 cent bonus on Non-Focal
5 cent bonus on Focal
1 cent bonus on Non-Focal
Tr 0
0.36 0.14 0.17
0.12 0.12 0.08
Tr 1
0.40 0.13 0.12
0.09 0.09 0.17
Tr 2
0.56 0.13 0.11
0.06 0.06 0.08
Tr 3
0.36 0.12 0.12
0.06 0.07 0.27
Tr 6
0.30 0.06 0.05
0.04 0.02 0.53
Tr 4
0.66 0.06 0.08
0.05 0.04 0.11
Tr 5
0.33 0.10 0.11
0.06 0.11 0.29
Row
Column
y
0.1
0.2
0.3
0.4
0.5
0.6

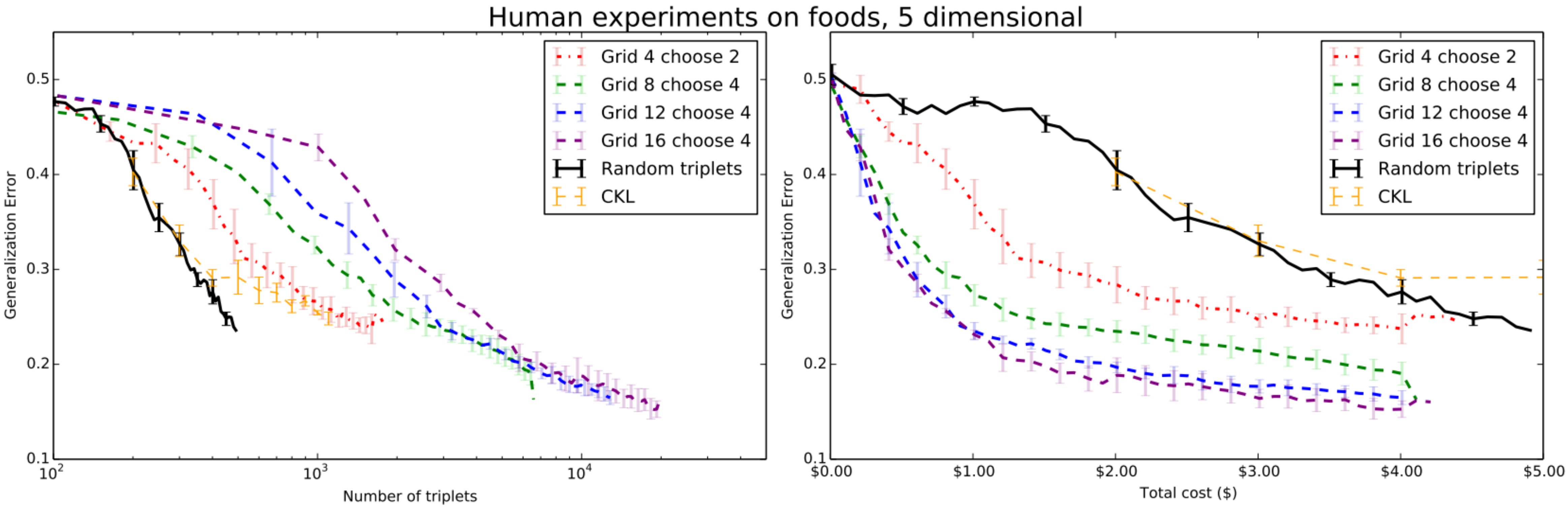

Monetary Incentives
(5-cent bonus)
Non-Monetary Incentives
(progress bars)
Gain in Fraction of Subjects From Baseline
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
Focal
Non-Focal
Focal
Non-Focal
Position in the 2x3 Grid, By Saliency
position
Focal
Non-Focal

"Cost-Effective HITs for Relative Similarity Comparisons "

M. Wilber, I. Kwak, and S. Belongie

Proceedings of the 2014 Conference on Human Computation


Which food on the right tastes more similar to the one on the left?
Please select the two foods that taste most similar to the food on the left.



Human experiments on foods, 5 dimensional
Generalization Error
Number of triplets
Total cost ($)
Grid 4 choose 2
Grid 8 choose 4
Grid 12 choose 4
Grid 16 choose 4
Random triplets
CKL
0.5
0.4
0.3
0.2
0.1
10^2
10^3
10^4
$0.00
$1.00
$2.00
$3.00
$4.00
$5.00

"Incentives to Counter Bias in Human Computation“

B. Faltings, P. Pu, B.D. Tran, and R. Jurca

Proceedings of the 2014 Conference on Human Computation

Proposition 2 *Whenever the agents' prior belief $Pr(x)$ is equal to the publicly available distribution $R(x)$, the Peer Truth Serum makes truthful reporting a Nash Equilibrium.*

Proof: Note that the expected reward for an agent who solves the task, obtains answer x and reports y is:

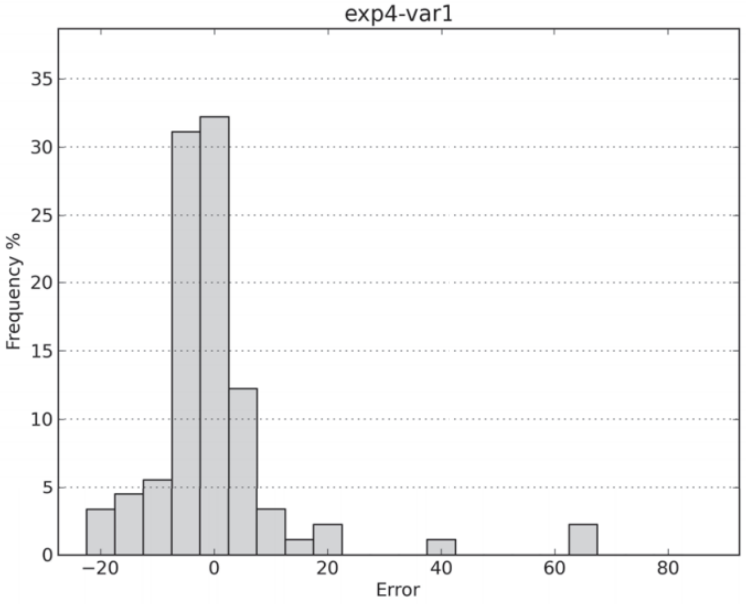
$$pay(x,y) = Pr_x(y) \cdot f(y,y,R)$$

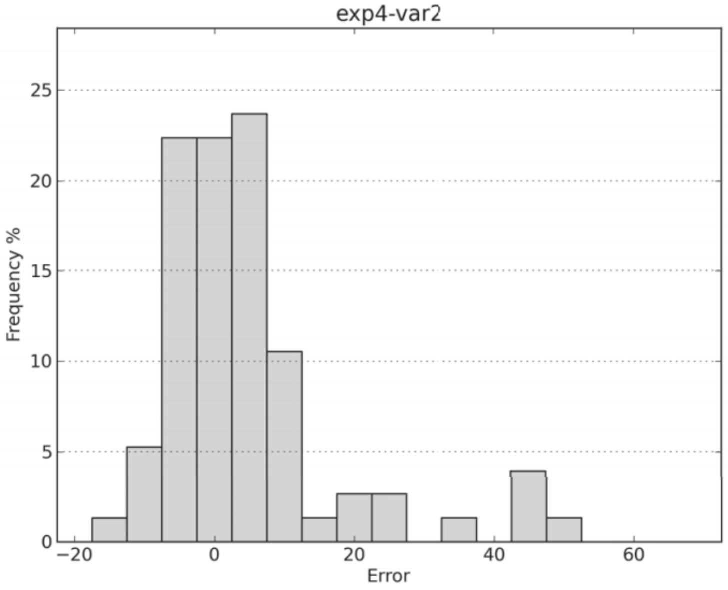
The condition for solving the task and truthful reporting is being the best response by a margin greater than γ is:

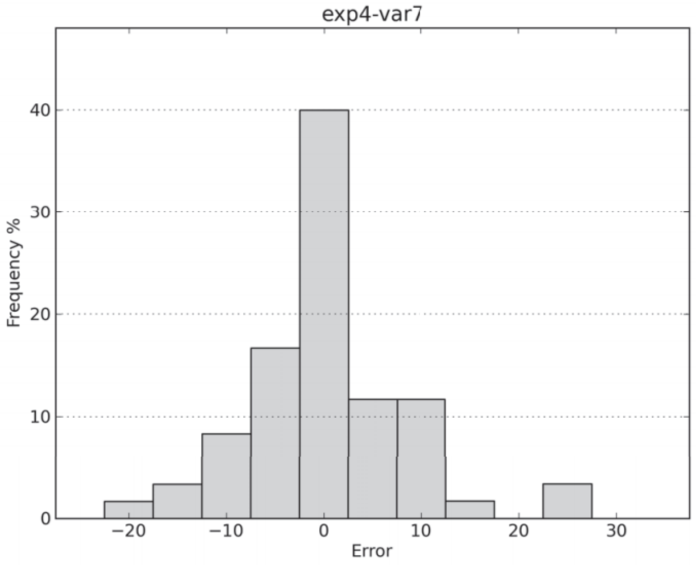
$$\begin{aligned}
\forall x,y,x \neq y : pay(x,x) - \gamma &> pay(x,y)\\
Pr_x(x)f(x,x,R) - \gamma &> Pr_x(y)f(y,y,R)\\
Pr_x(x)f(x,x,Pr) - \gamma &> Pr_x(y)f(y,y,Pr)
\end{aligned}$$

where γ is the cost of effort for solving the task and obtaining answer x. If $f(x,x,R) = c/R(x)$ and $\gamma = c\epsilon$, the truthfulness condition is just the self-predicting condition 2. The scaling constant c has to be chosen in function of the margin ϵ that can be assumed in condition 2.

Note that this reward scheme has a very intuitive nature: it rewards answers that go against the biases expressed by $R(x)$, but on the other hand still requires matching another


exp4-var1
Frequency %
Error
0
5
10
15
20
25
30
35
−20
0
20
40
60
80

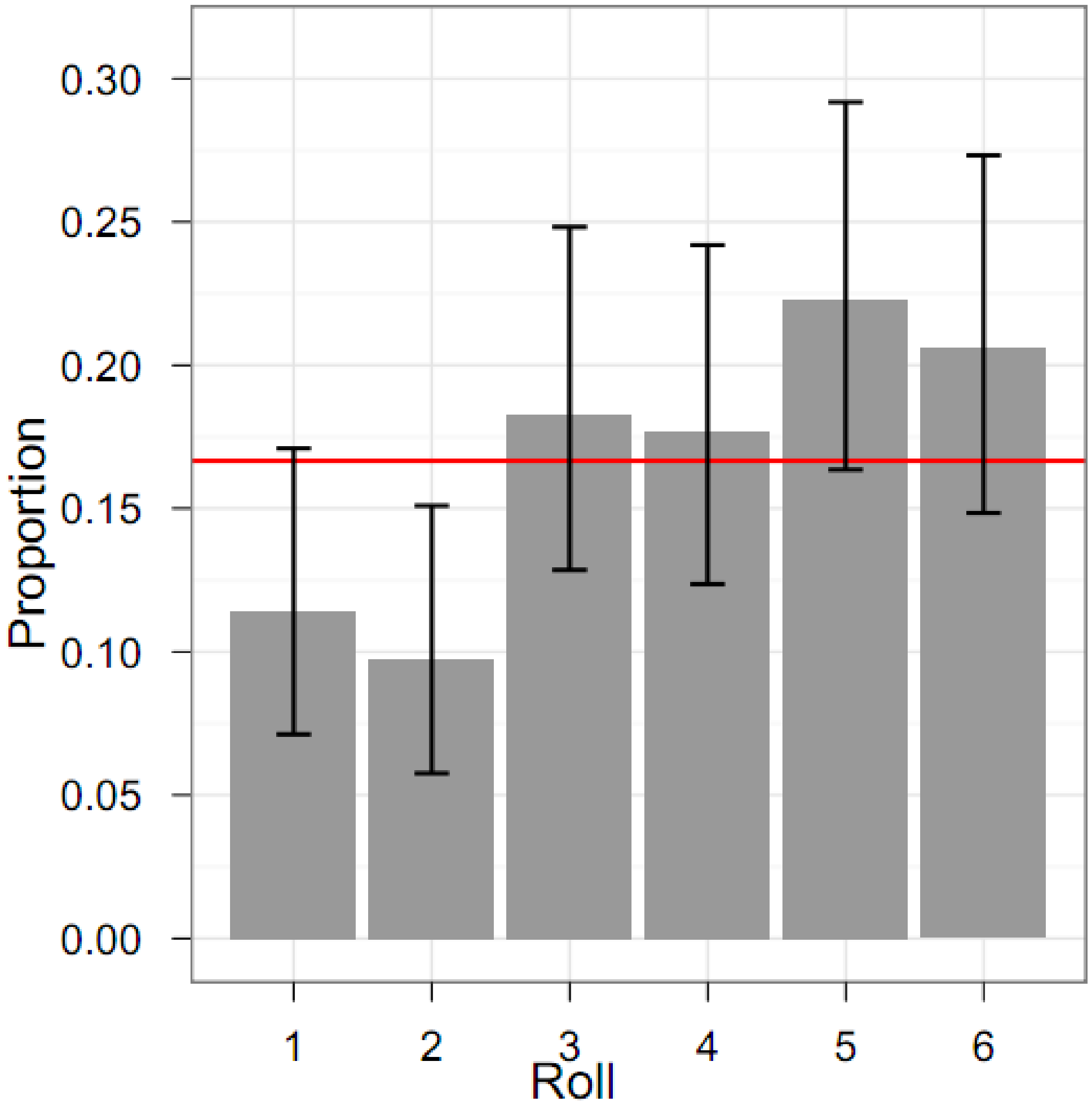

exp4-var2
Frequency %
Error
0
5
10
15
20
25
−20
0
20
40
60


exp4-var7
Frequency %
Error
0
10
20
30
40
−20
−10
0
10
20
30

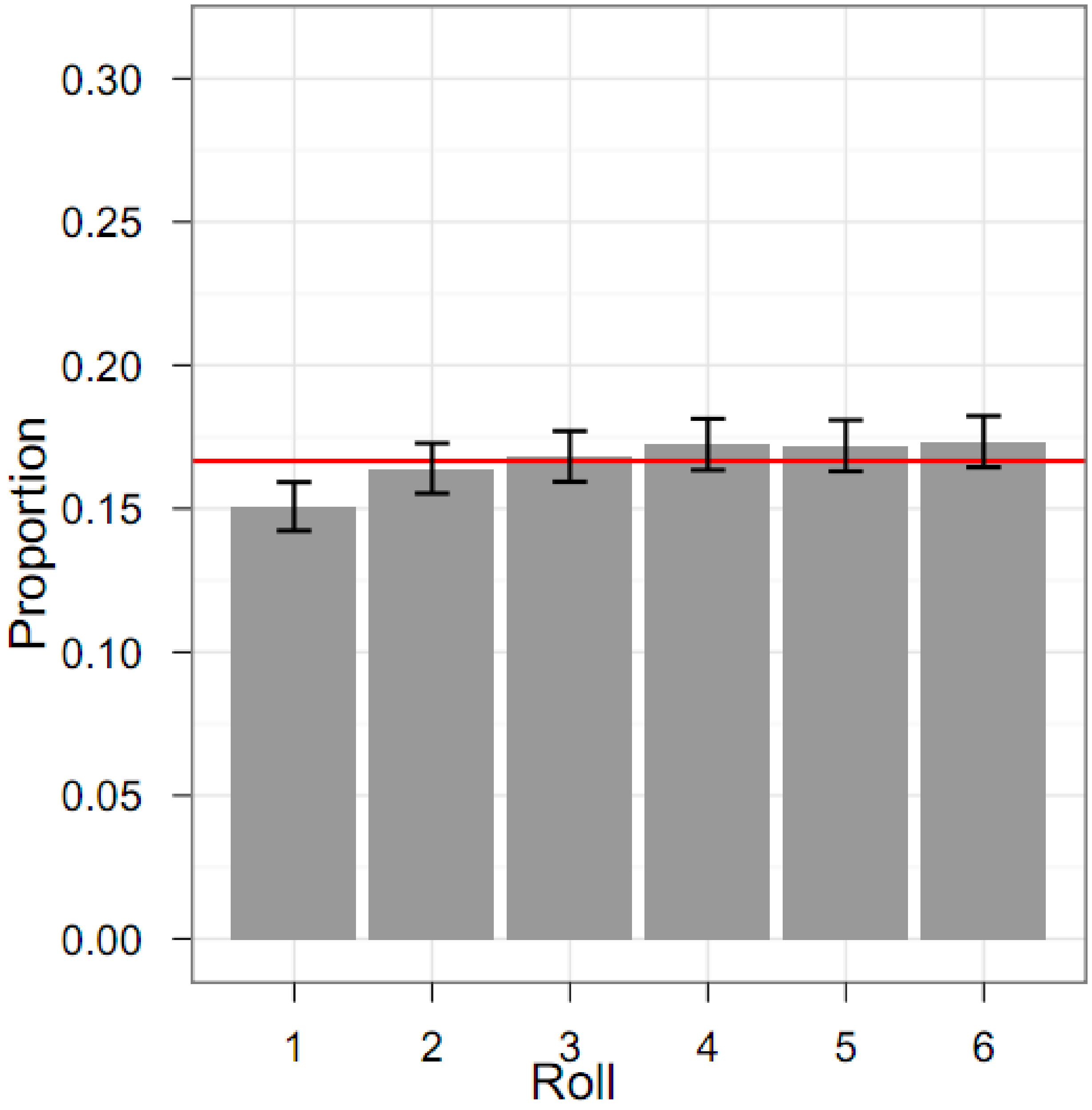
"Honesty in an Online Labor Market“

Siddharth Suri, Daniel G. Goldstein, and Winter A. Mason

In *Proceedings of the Third Human Computation Workshop*, 2011



0.30
0.25
0.20
0.15
0.10
0.05
0.00
Proportion
1
2
3
4
5
6
Roll

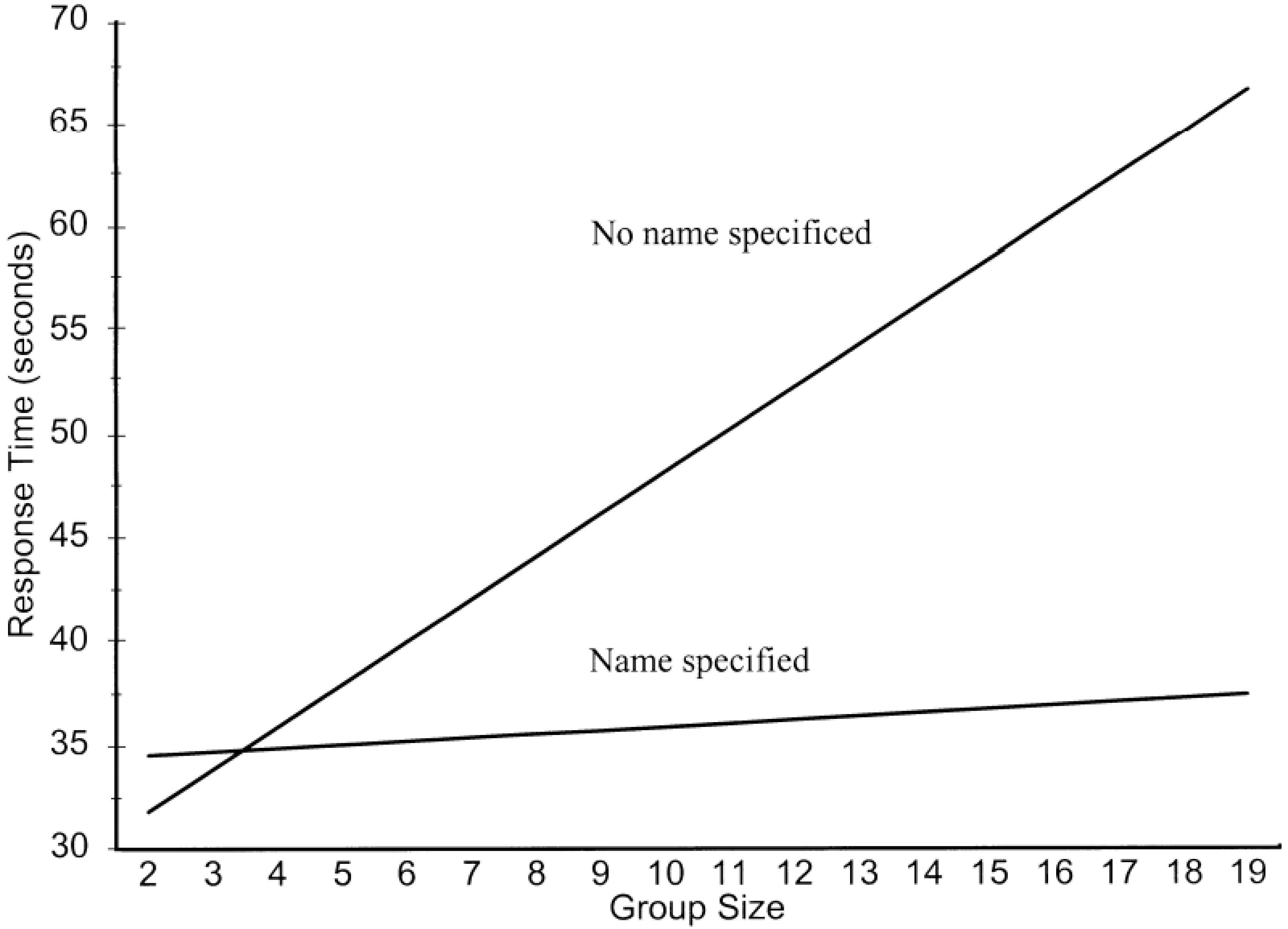


0.30
0.25
0.20
0.15
0.10
0.05
0.00
Proportion
1
2
3
4
5
6
Roll

"Bystander Intervention in Computer-Mediated Communication"

P.M. Markey

Computers in Human Behavior (2000)

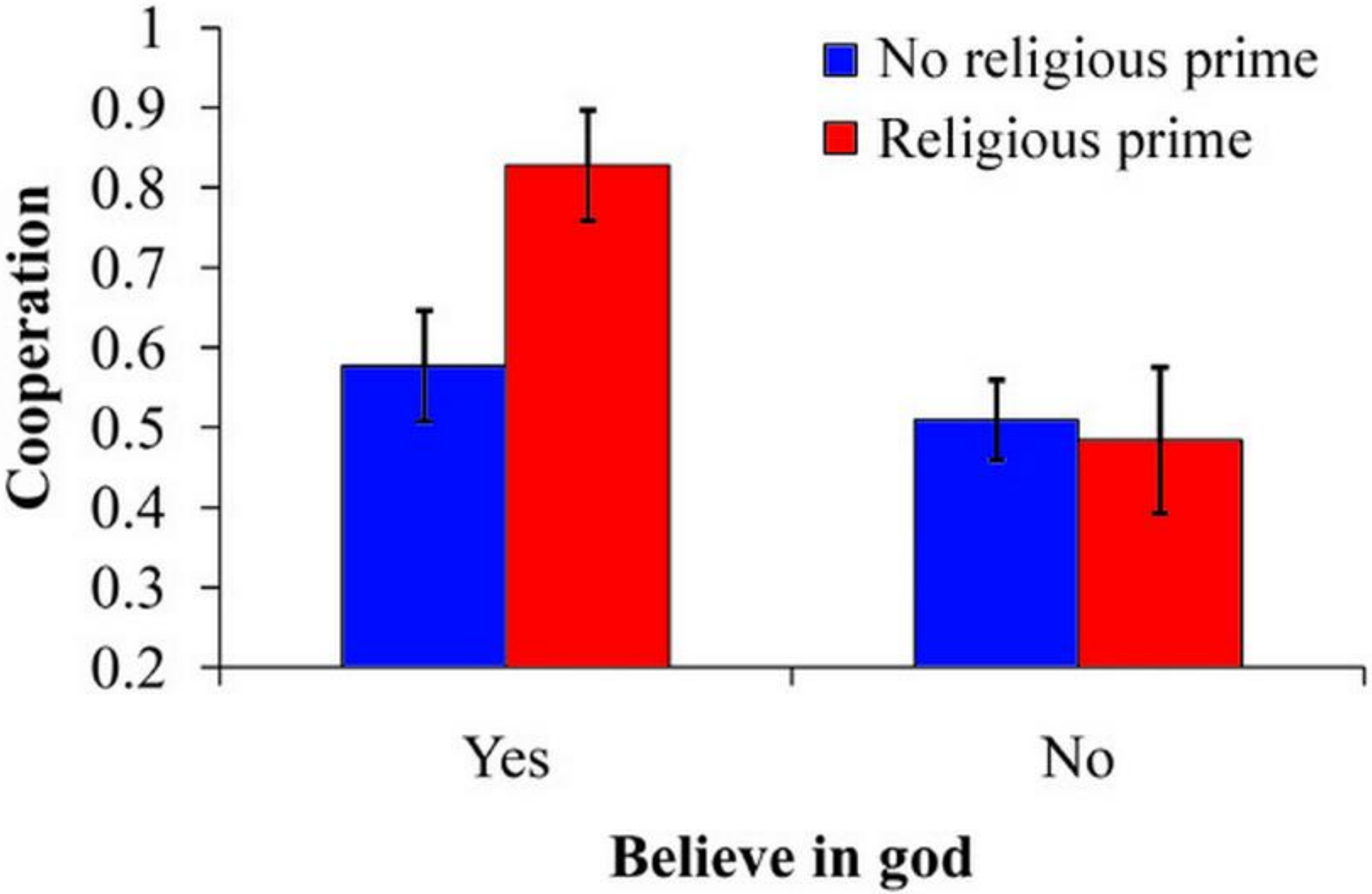


70
65
60
55
50
45
40
35
30
Response Time (seconds)
No name specificed
Name specified
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
Group Size

“God Is Watching You: Priming God Concepts Increases Prosocial Behavior in and Anonymous Economic Game”

A.Shariff and A. Norenzayan

Psychological Science, 18:9 803-809 (2007)



1
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
Cooperation
No religious prime
Religious prime
Yes
No
Believe in god

Readings for Next Time

- Tuesday, March 1:
  Infotopia, Chapter 2