Deterring Cheating in Online Environments

Henry Corrigan-Gibbs
Stanford University

Nakull Gupta
Microsoft Research India

Curtis Northcutt
MIT

Edward Cutrell
Microsoft Research India

William Thies
Microsoft Research India
But...
But...

Disposable identities
But...

Disposable identities

Few enforceable rules
But...

Disposable identities

Few enforceable rules

Different social norms
Craigslist fake check scam is going around again
Craigslist fake check scam is going around again

Local Chinese Restaurant Calls Out Dishonest Yelper With Surveillance Footage

by Andrew Dalton Mar 24, 2015, 2:00p @dolftown | 41 COMMENTS
Craigslist fake check scam is going around again

Local Chinese Restaurant Calls Out Dishonest Yelper With Surveillance Footage

by Andrew Dalton Mar 24, 2015, 2:00p @dolftown | 41 COMMENTS

Dozens of Plagiarism Incidents Are Reported in Coursera's Free Online Courses
Local Chinese Restaurant Calls Out Dishonest Yelper With Surveillance Footage

by Andrew Dalton Mar 24, 2015, 2:00p @dolftown | 41 COMMENTS

Dozens of Plagiarism Incidents Are Reported in Coursera's Free Online Courses

Common eBay scams and how to avoid them

Cheating on Amazon Mechanical Turk
Craigslist fake check scam is going around again

Local Chinese Restaurant Calls Out Dishonest Yelper With Surveillance Footage
by Andrew Dalton Mar 24, 2015, 2:00p @dolftown | 41 COMMENTS

Dozens of Plagiarism Incidents Are Reported in Coursera's Free Online Classes

Common eBay scams and how to avoid them

Cheating on Amazon's Mechanical Turk Examinations go hi-tech, so do methods to cheat

Swati Shinde | TNN | Aug 3, 2015, 05.47 AM IST
Research Questions

1. How can we deter dishonest behavior in online environments?

2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)

II. Amazon Mechanical Turk (US & India)
Research Questions

1. How can we deter dishonest behavior in online environments?
2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)
II. Amazon Mechanical Turk (US & India)
Research Questions

1. How can we deter dishonest behavior in online environments?

2. How will we know when we have succeeded? 
   (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)

II. Amazon Mechanical Turk (US & India)
Research Questions

1. How can we deter dishonest behavior in online environments?
2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)
II. Amazon Mechanical Turk (US & India)
A Theory of Dishonesty
A Theory of Dishonesty

Moral Consequences

- Will this make me feel guilty?
- What would my parents think?
A Theory of Dishonesty

Moral Consequences

• Will this make me feel guilty?
• What would my parents think?

Material Consequences

• Will I get caught?
• How much do I stand to gain / lose?
Intervention I: Honor Code

• Make cheating “morally costly”
• Require signed statement declaring intent to behave honestly
• Evidently effective in in-person interactions
  [Mazar et al. 2008], [Pruckner and Sausgruber 2013], [Rosenbaum et al. 2014]
Intervention I: Honor Code

- Make cheating “morally costly”
- Require signed statement declaring intent to behave honestly
- Evidently effective in in-person interactions

[Mazar et al. 2008], [Pruckner and Sausgruber 2013]

“I shall neither give nor receive help during this examination.”
Intervention I: Honor Code

• Make cheating “morally costly”
• Require signed statement declaring intent to behave honestly
• Evidently effective in in-person interactions

[Mazar et al. 2008], [Pruckner and Sausgruber 2013], [Rosenbaum et al. 2014]
Intervention II: Warning

- Inform the participant of the potential material costs of cheating
  [Fellner et al. 2013]
Intervention II: Warning

• Inform the participant of the potential material costs of cheating
  [Fellner et al. 2013]

Warning
If we discover that you violated the rules of this online exam, we may:
- Cancel your exam.
- Cancel your account.
- Notify your university.
Intervention II: Warning

- Inform the participant of the potential material costs of cheating

[Fellner et al. 2013]
General Method

1. Randomize participants into treatment groups:
   (a) Control
   (b) Honor Code
   (c) Warning

2. Measure rate of cheating in each group

3. Determine whether treatment had any measurable effect
General Method

1. Randomize participants into treatment groups:
   (a) Control
   (b) Honor Code
   (c) Warning

2. Measure rate of cheating in each group

3. Determine whether treatment had any measurable effect
Research Questions

1. How can we deter dishonest behavior in online environments?

2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)

II. Amazon Mechanical Turk (US & India)
Research Questions

1. How can we deter dishonest behavior in online environments?

2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)

II. Amazon Mechanical Turk (US & India)
Measuring Cheating
Measuring Cheating

onlineexam.edu
Measuring Cheating

onlineexam.edu
Measuring Cheating

onlineexam.edu
Measuring Cheating

onlineexam.edu
Measuring Cheating

onlineexam.edu

wikipedia.org
Measuring Cheating

onlineexam.edu

wikipedia.org
Measuring Cheating
Measuring Cheating

• For free-response questions, can use plagiarism detection tools
Measuring Cheating

- For free-response questions, can use plagiarism detection tools
  
  “Did Alice copy verbatim from Wikipedia?”
Measuring Cheating

- For free-response questions, can use plagiarism detection tools
  
  "Did Alice copy verbatim from Wikipedia?"
  
  "Did Alice and Bob submit exactly the same response?"
Measuring Cheating

- For free-response questions, can use plagiarism detection tools
  
  "Did Alice copy verbatim from Wikipedia?"

  "Did Alice and Bob submit exactly the same response?"

- In use today (e.g., Turnitin), and we use these techniques too
Measuring Cheating

• For free-response questions, can use plagiarism detection tools
  
  “Did Alice copy verbatim from Wikipedia?”
  
  “Did Alice and Bob submit exactly the same response?”

• In use today (e.g., Turnitin), and we use these techniques too

... but for multiple-choice questions?
“Honey pot”

onlineexam.edu

[Provos 2004]
“Honey pot”

onlineexam.edu

exam-answers.org

[Provos 2004]
“Honey pot”

onlineexam.edu

exam-answers.org

[Provos 2004]
“Honey pot”

onlineexam.edu

exam-answers.org

[Provos 2004]
We use the cookie to identify dishonest behavior

onlineexam.edu

exam-answers.org
Caveat: False Negatives

onlineexam.edu

exam-answers.org
Caveat: False Negatives

onlineexam.edu

exam-answers.org
Caveat: False Negatives
Caveat: False Negatives

onlineexam.edu

exam-answers.org
Caveat: False Negatives

onlineexam.edu

exam-answers.org
Caveat: False Negatives

onlineexam.edu

exam-answers.org
Research Questions

1. How can we deter dishonest behavior in online environments?

2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)

II. Amazon Mechanical Turk (US & India)
Research Questions

1. How can we deter dishonest behavior in online environments?
2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)
II. Amazon Mechanical Turk (US & India)
Research Questions

1. How can we deter dishonest behavior in online environments?
2. How will we know when we have succeeded? (i.e., How do we measure the rate of cheating?)

Experiments

I. Microsoft’s online course platform (India)
II. Amazon Mechanical Turk (US & India)
Experiment I: Online Course

- MSR India free online algorithms course (2014-15)
- Online final examination
  - 14 multiple-choice questions + 1 free response
  - 409 students in spring 2014
  - 223 students in winter 2015
- “Closed book, closed Internet” exam

All students had an opportunity to opt out and we did not penalize students we suspected of cheating.
Practice Questions

Question 1:

Let $A(n)$ and $B(n)$ denote the worst-case running time for algorithms A and B, respectively, as a function of the input size $n$. Consider the following statements:

I. If $A(n)$ is $\Theta(B(n))$, then $B(n)$ is $\Theta(A(n))$

II. If $A(n)$ is $O(B(n))$, then $B(n)$ is $O(A(n))$

A. Statement I is true and Statement II is false
B. Statement I is false and Statement II is true
C. Both Statement I and Statement II are true
D. Both Statement I and Statement II are false

Click to see the answer
Practice Questions

Question 1:

Let $A(n)$ and $B(n)$ denote the worst-case running time for algorithms A and B, respectively, as a function of the input size $n$. Consider the following statements:

I. If $A(n)$ is $\Theta(B(n))$, then $B(n)$ is $\Theta(A(n))$
II. If $A(n)$ is $O(B(n))$, then $B(n)$ is $O(A(n))$

A. Statement I is true and Statement II is false
B. Statement I is false and Statement II is true
C. Both Statement I and Statement II are true
D. Both Statement I and Statement II are false

Click to see the answer
Practice Questions

Question 1:

Let $A(n)$ and $B(n)$ denote the worst-case running time for algorithms $A$ and $B$, respectively, as a function of the input size $n$. Consider the following statements:

I. If $A(n)$ is $\Theta(B(n))$, then $B(n)$ is $\Theta(A(n))$  
II. If $A(n)$ is $O(B(n))$, then $B(n)$ is $O(A(n))$

A. Statement I is true and Statement II is false  
B. Statement I is false and Statement II is true  
C. Both Statement I and Statement II are true  
D. Both Statement I and Statement II are false
Let $A(n)$ and $B(n)$ denote the worst-case running time for algorithms $A$ and $B$, respectively, as a function of the input size $n$. Consider the following statements:

I. If $A(n)$ is $\Theta(B(n))$, then $B(n)$ is $\Theta(A(n))$
II. If $A(n)$ is $O(B(n))$, then $B(n)$ is $O(A(n))$

A. Statement I is true and Statement II is false
B. Statement I is false and Statement II is true
C. Both Statement I and Statement II are true
D. Both Statement I and Statement II are false

Click to see the answer
let \( a(n) \) \( b(n) \) denote the worst case running time
let \( a(n) \) \( b(n) \) denote the worst case running time

Practice questions - People.csail.mit.edu
people.csail.mit.edu/thies/6.046-fall04/practice-questions.htm
Let \( A(n) \) and \( B(n) \) denote the worst-case running time for algorithms \( A \) and \( B \), respectively, as a function of the input size \( n \). Consider the following statements:

[PDF] 1 Analysis of Merge Sort - Usf
www-bcf.usc.edu/~dkepme/.../11-07.p...
Let's use \( T(n) \) to denote the worst-case running time of Merge Sort on an array of \( n \)...
general form is \( T(n) = a \cdot T(n/b) + f(n) \), for some constant numbers \( a \) and \( b \), ...

CS240: Data Structures & Algorithms I
https://www.csp.edu/~ftang/courses/CS240/lectures/analysis.htm
The running time of an algorithm or a data structure method typically grows with the ....
... Time taken by the slowest primitive operation. Let \( T(n) \) be the worst-case array of...
Let $a(n)$ and $b(n)$ denote the worst case running time for algorithms A and B, respectively, as a function of the input size $n$. Consider the following statements:
Study population
603 students
Study population
603 students

Plagiarized on free-response
102 students
Study population
603 students

Plagiarized on free-response
102 students

Visited honey pot
50 students
Study population: 603 students

- Plagiarized on free-response: 102 students
- Visited honey pot: 50 students
- Both: 5 students
Results: Cheating vs. Score

- Scores (Honest) vs. Frequency
- Scores (Cheating) vs. Frequency
Results: Cheating vs. Score

Cheating doesn’t pay! (p < 0.0001)
Results: HC and Warning

Percentage cheating:
- 0%
- 10%
- 20%
- 30%
- 40%
- 50%

Online Exam February 2015:
- Control: 31.7%
- Honor Code: 25.9%
- Warn: 14.6%
Results: HC and Warning

Honor code group saw a drop in cheating rate (not stat. signif.)

Percentage cheating

Online Exam

Control: 31.7%
Honor Code: 25.9%
Warn: 14.6%
Results: HC and Warning

- Control: 31.7%
- Honor Code: 25.9%
- Warn: 14.6%

Online Exam
Results: HC and Warning

Warning has an effect ($N=398, \ p < 0.001$)
Results: HC and Warning

Online Exam

- Control: 31.7%
- Honor Code: 25.9%
- Warn: 14.6%
Results: HC and Warning

Online Exam

- Control: 31.7%
- Honor Code: 25.9%
- Warn: 14.6%

MTurk (India)

- Control: 41.6%
- Honor Code: 35.8%
- Warn: 18.3%
Where we go from here...
Where we go from here...

- Does the effect of a warning diminish over time?
Where we go from here…

- Does the effect of a warning diminish over time?
- Do warnings scare off honest and dishonest users at different rates?
Where we go from here…

- Does the effect of a warning diminish over time?
- Do warnings scare off honest and dishonest users at different rates?
- Can we use honey pots to automate the detection of unethical behavior online? Abuse, etc.
Where we go from here…

• Does the effect of a warning diminish over time?

• Do warnings scare off honest and dishonest users at different rates?

• Can we use honey pots to automate the detection of unethical behavior online? Abuse, etc.

• Can warnings help deter other bad behavior online? Nasty comments, etc.
Summary

Online Exam February 2015

- Control: 31.7%
- Honor Code: 25.9%
- Warn: 14.6%
Summary

• We studied the effect of an honor code and a warning at deterring cheating in an online course and Mechanical Turk.
We studied the effect of an honor code and a warning at deterring cheating in an online course and Mechanical Turk. We find evidence that the warning was effective in online settings… less so the honor code.
Summary

• We studied the effect of an **honor code** and a **warning** at deterring cheating in an **online course** and **Mechanical Turk**.

• We find evidence that the warning was effective in online settings… less so the honor code.

• “Honey pots” may be a useful tool for measuring rates of cheating in future work.