

# How to Teach a Class to Grade Itself

Nicolaas Kaashoek and William Wu

Mentors: Matt Weinberg and  
Christos Tzamos

# The Motivation

coursera



edx

Previously, classes  
only had 20-30  
Students in them,  
But with online courses,

**How are you going to grade all the assignments?**  
That number had grown rapidly and ceased

# What should be done?



- Idea 0: Do tons of work →
- Idea 0.5: Hire tons of TAs →
  
- Idea 1: Just use multiple choice
- Limited, some classes can't just use multiple choice
- Idea 2: Automated Grading:

*The New York Times*

April 4, 2013

## **Essay-Grading Software Offers Professors a Break**

But all systems can be duped, teachers “do a much better job of providing feedback than a machine ever could.”

# Here's a better idea!

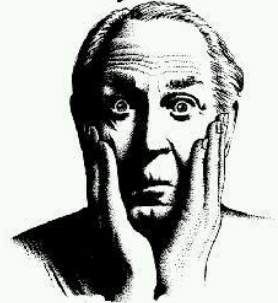
Make the students grade it themselves!



# Why is this hard?

- Students think work is bad, might not care about the grades of their peers
- Might want to help friends or hurt enemies
- So how can they be encouraged to grade correctly
  - By being incentivized!
- Our Approach: Use ideas from *Game Theory* and *Mechanism Design*

Oh, no!



# Why Game Theory?

- Allows us to understand how people behave
- Mechanism design allows us to create a set of conditions to force people to behave just how we want them to



Final goal? Design a mechanism that will encourage students to grade correctly because it is in their own best interests

# The First Step: Understand Student Behavior (a very simple model)

1) Students want to be as happy as possible

In math terms: Students have function  $H$ , want to maximize

2) Students want good grades

In math terms:  $H$  increases as grade increases

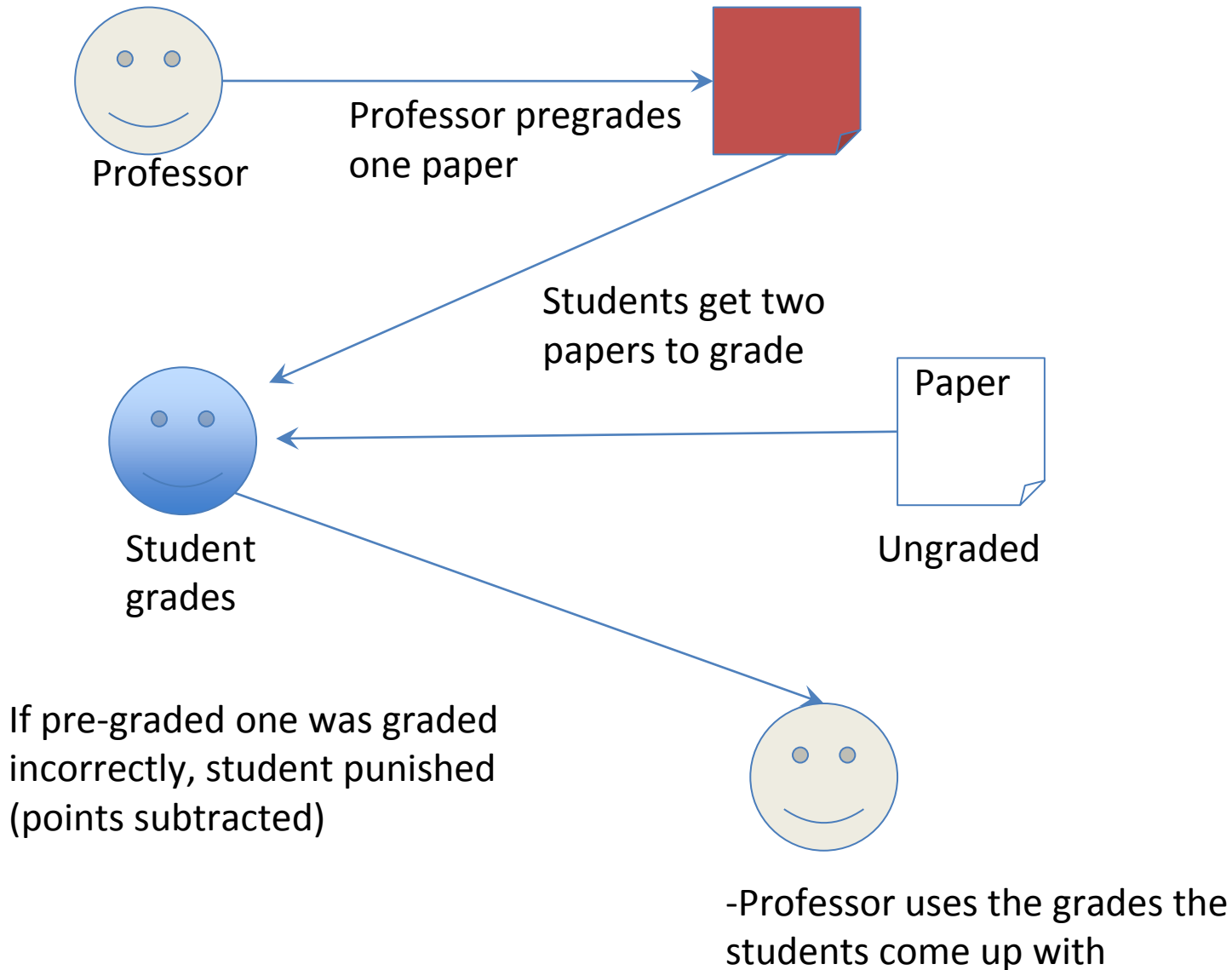
3) Students don't like to do work

In math terms: When work is done,  $H$  goes down, grading a paper costs one unit of happiness

4) Students only care about themselves (not fairness, etc)

In math terms:  $H$  depends only on the grade they receive and the amount of work they do

# The First Idea

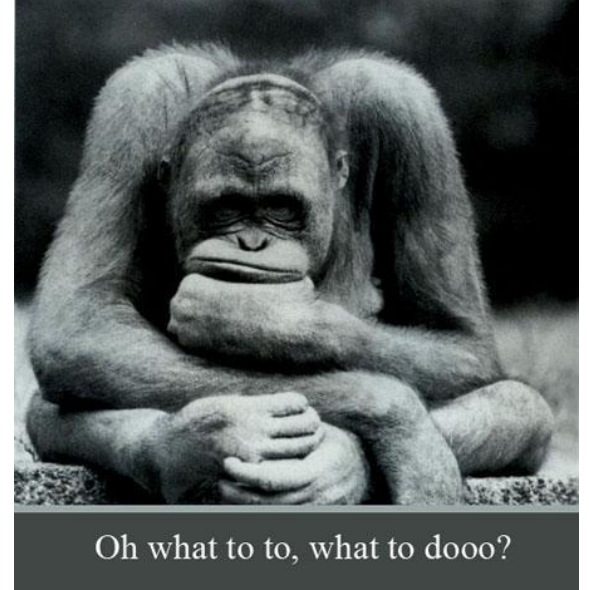




# Why it works

What can a student do?

1. Grade both papers
2. Ignore both papers
3. Grade one and not the other



Students will do #1 because #2 and #3 involve a risk of punishment.

# Why it Works: The Math

G.Assigned = grade assigned by grader

G.Minimum = some minimum grade

G.Actual =  $\text{Max}\{\text{G.Minimum}, \text{G.Assigned}\}$

**Happiness for grading just one paper:  $H(\text{G.Actual})/2 + H(0)/2 - 1$**

**For Grading both papers:  $H(\text{G.Actual}) - 2$**

**For Grading neither paper:  $H(0)$**

If  $H(\text{G.Actual}) - 2 > H(0)$ , then student will grade both papers.

**Choose G.Minimum such that  $H(\text{G.Minimum}) > H(0) + 2$ .**

i.e. every student who grades correctly receives some minimum grade equivalent to 2 units of work (e.g. maybe a 30%)

# So, Are We Done?

- Strong assumption: people can't communicate
- With communication, students can discover which one the professor graded
- Why grade the student's paper?

Problem: Everyone shares a paper



# A Simple Fix

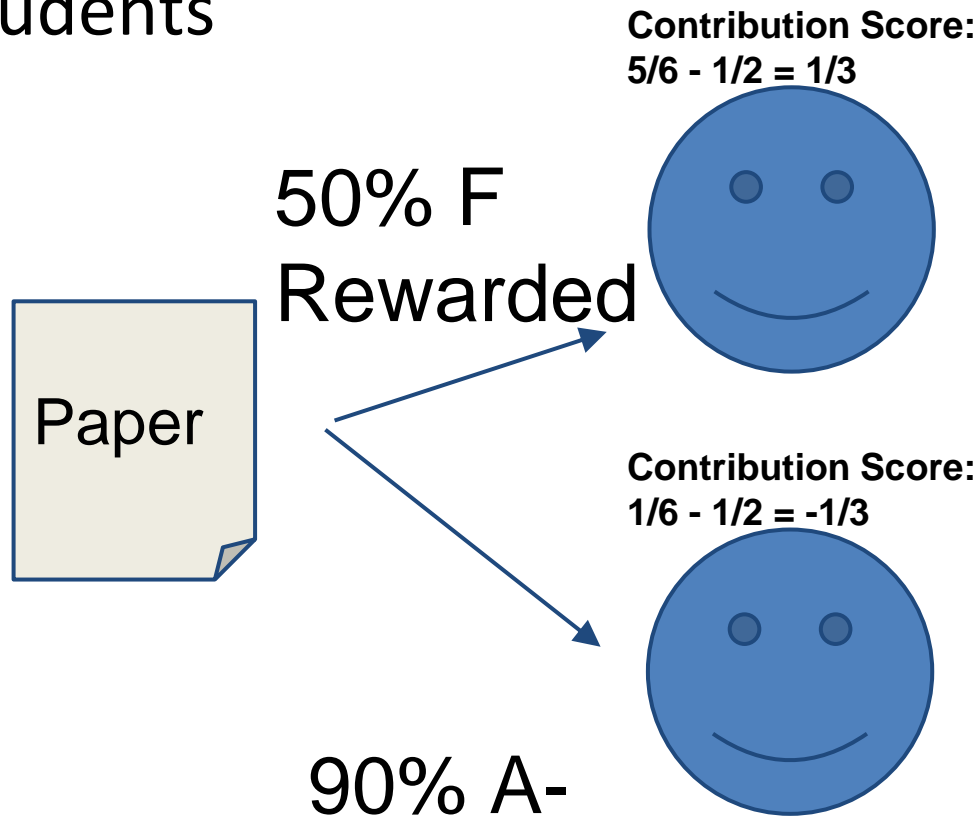
- More calibrated papers
- Distribute each paper multiple times, papers appear same number of times, regardless of calibration
- Can't tell what is calibrated
- This creates a lot of work for both teachers and students, bad
- Need a more powerful idea

# The Next Step

- Need a different way to incentivize people
- Calibrating is like the professor just checking intelligently, need a new idea
  - Idea: Have the students do the checking!
- The incentive: A competition, 2 graders compete to most effectively grade the paper

# The New Mechanism

Every paper goes to 2 students



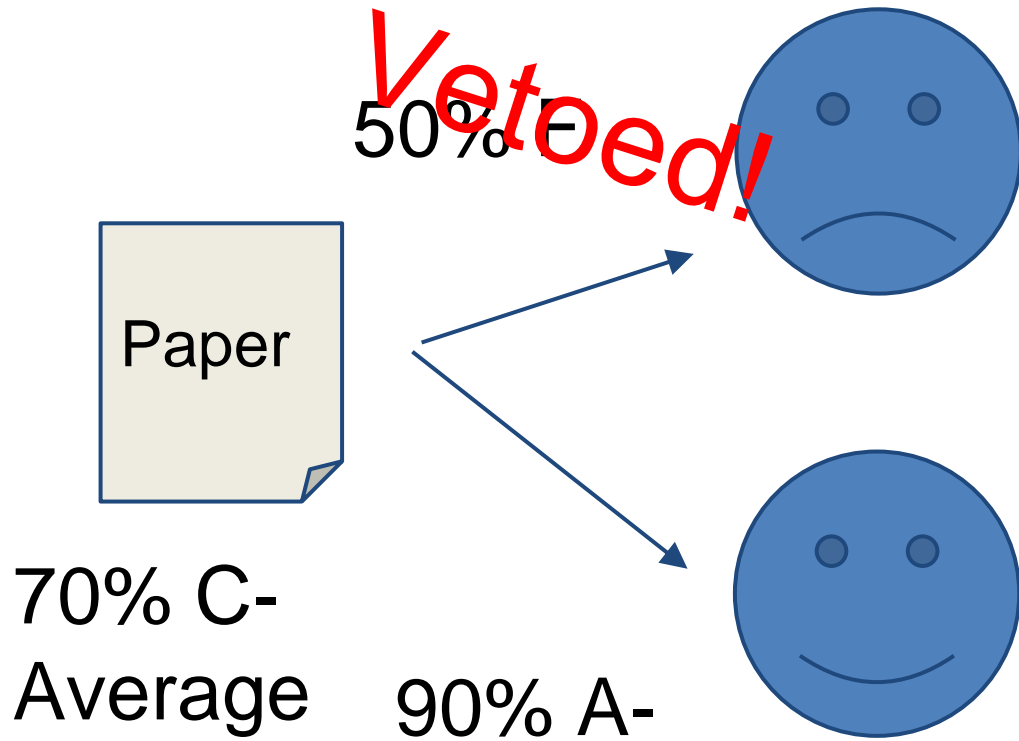
Each student takes off points with a justification

Students are then given a **contribution score** from -1 to 1, From each of the 2 assignments they grade, they get their points deducted/total points deducted-0.5

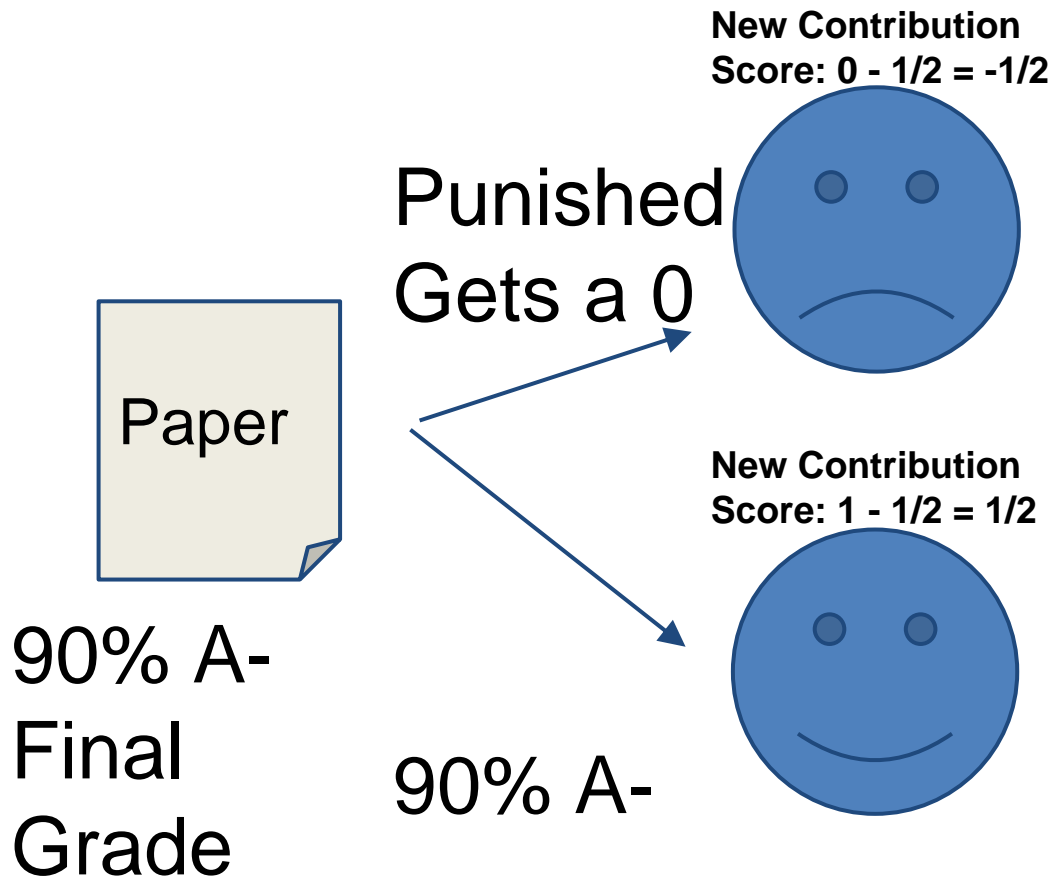
# The New Mechanism

Grade for the writer is the average of the two

If the writer doesn't like their grade, they “veto” the offending graders



# The New Mechanism



Resolved by professor, wrongdoer is punished

Final assignment grades:  $H^{-1}(\text{Contribution score} * 4 + H(\text{Average}))$



# Why Does This Work

What can students do?

1. Honestly grade: get rewarded
2. Be lazy, take off points without justification: be vetoed and punished
3. Be more lazy and take off no points: no reward

Game Theory: Students grade honestly, so vetoes won't happen

Not much work for students or professor

# Why it Works: The Math (part I)

Students can: Grade or not grade

If don't grade: Can take off points with no justification or give 100

- no justification --> vetoed and punished
- should give 100

If grade: Can either

- Not take off all points: Throwing away free points
- Take off extra points with no justification:

Punished

- Grade Correctly: Best Option

# Why it Works: The Math (part II)

By part 1, all graders will give 100 or the correct grade.

What happens when grading partner gives 100:

If don't grade: Contribution score = 0, Effort = 0

Happiness =  $H(\text{your score} - \text{your assignment})$

If you do grade: Contribution score = 0, Effort = 1

Happiness =  $H(\text{your score} - \text{your assignment}) + 3$ .

In both cases it is better for you to always grade because your happiness will be higher, so everyone will grade fairly

# Are We Done Yet?

Mechanism's theory is pretty good. However...

- Quite mean
- Encourages really harsh grading



# The Future

- Make it nice to students, positive competition
- Throughout this talk, some strong assumptions were made: everyone is a competent grader - remove assumption
- Make an experiment

# Acknowledgements

Our Mentors, Matt Weinberg and Christos  
Tzamos

Professor Costis Daskalakis

The Primes Program

Slava Gerovitch

Pavel Etingof

Tanya Khovanova