

October 27, 2015

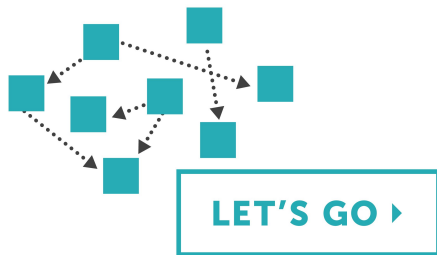


Kevin H. Wilson
Principal Data Scientist

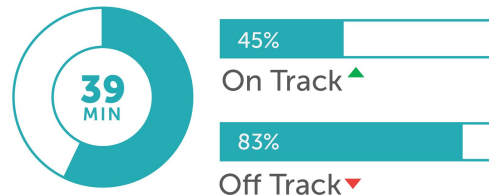
Determining Skill Levels

What does Knewton do?

Recommendations for students



Learning analytics for students and teachers



Content insights for publishers



What does Knewton do?

☆ Your Assignments

+

 Assignments

MATH

Multiplication And Divis...
Category: Numbers & Comput...

$4 \times 3 = 12$
 $12 \div 3 = 4$

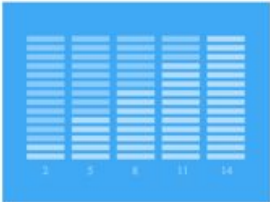
Sample

Keep Going

🔴 Past Due: Apr 11th

MATH

Arithmetic Sequence
Category: Numbers & Comput...



Sample

Start

🔴 Past Due: Apr 17th

MATH

Solving Logarithmic Eq...
Category: Algebra & Algebra II

\log_b

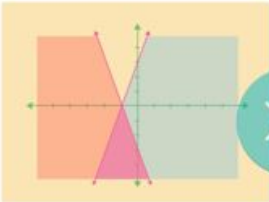
Sample

Start

🔴 Past Due: Apr 29th

MATH

Systems Of Linear Equ...
Category: Algebra & Algebra II



Sample


Start

🔴 Past Due: May 18th

💡 Suggested Assignments

MATH


Circle Graph
Category: Data Analysis, Prob...



Why? Sample

+ Add Start

Line Graph
Category: Data Analysis, Prob...



Why? Sample

+ Add Start


Stem And Leaf Plot
Category: Data Analysis, Prob...

stem	leaf
9	0 4
8	3 4 5 7 7 9
7	0 2 2 9
6	5 8 9
5	3

Why? Sample

+ Add Start

Venn Diagram
Category: Data Analysis, Prob...



Why? Sample

+ Add Start

Some Data Problems Knewton Faces:

- What do students know?
- When do students learn it?
- Are students engaged?
- How effective are particular pieces of content?
- How do we measure success?

Some Data Problems Knewton Faces:

- **What do students know?**
- **When do students learn it?**
- Are students engaged?
- How effective are particular pieces of content?
- How do we measure success?

What do students know?

- How to assess?
- About a particular subject?
- At a particular time?
- In a particular setting?

How to assess?



There's no way to open a student's mind and, with a single measurement, determine how much she knows.

About a particular subject?

Summarize and describe distributions.

CCSS.MATH.CONTENT.6.SP.B.4

Display numerical data in plots on a number line, including dot plots, histograms, and box plots.

CCSS.MATH.CONTENT.6.SP.B.5

Summarize numerical data sets in relation to their context, such as by:

CCSS.MATH.CONTENT.6.SP.B.5.A

Reporting the number of observations.

CCSS.MATH.CONTENT.6.SP.B.5.B

Describing the nature of the attribute under investigation, including how it was measured and its units of measurement.

CCSS.MATH.CONTENT.6.SP.B.5.C

Giving quantitative measures of center (median and/or mean) and variability (interquartile range and/or mean absolute deviation), as well as describing any overall pattern and any striking deviations from the overall pattern with reference to the context in which the data were gathered.

CCSS.MATH.CONTENT.6.SP.B.5.D

Relating the choice of measures of center and variability to the shape of the data distribution and the context in which the data were gathered.

<http://www.corestandards.org/Math/Content/SP/>

At a particular time?



In a particular setting?

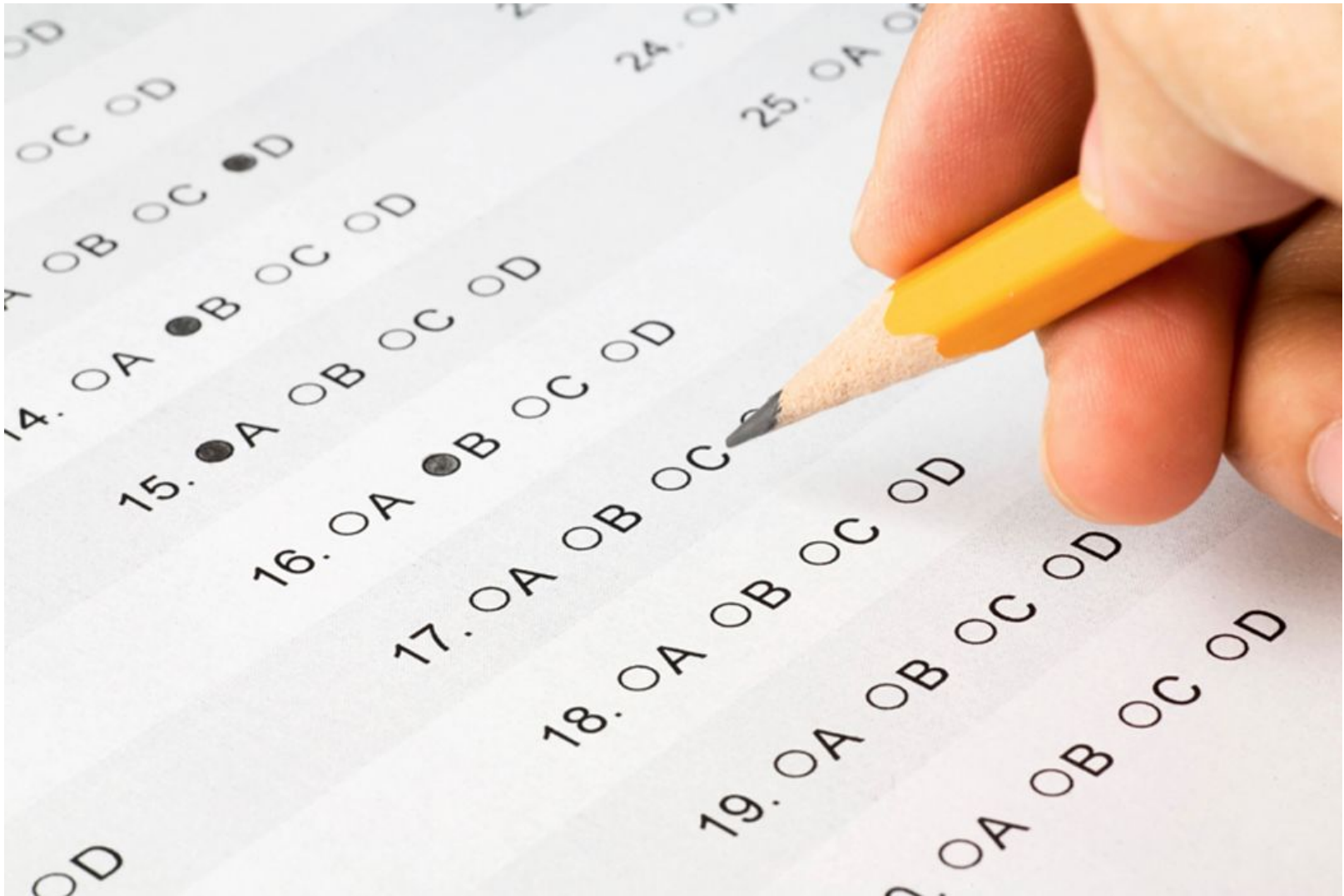


What do students know?

- About a particular subject?
- At a particular time?
- In a particular setting?

Take as fixed for this talk.

Thus is born....



Types of assessments:

- Fill-in-the-blank
- Free response
- Multiple choice
- Matching

Types of assessments:

- **Fill-in-the-blank**
- Free response
- **Multiple choice**
- Matching

Types of assessments:

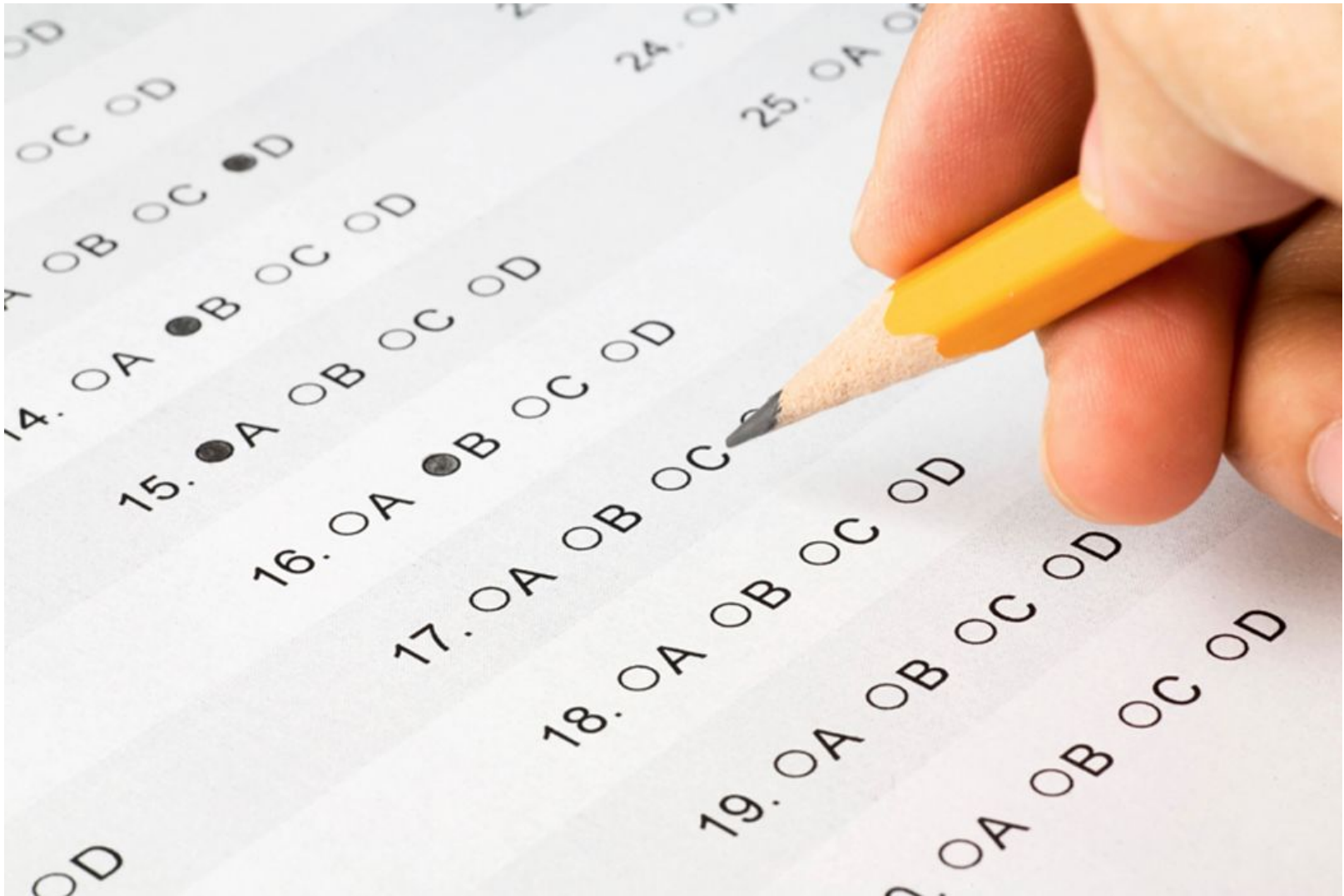
- **Fill-in-the-blank**
- Free response
- **Multiple choice**
- Matching

Correct/Incorrect is easy

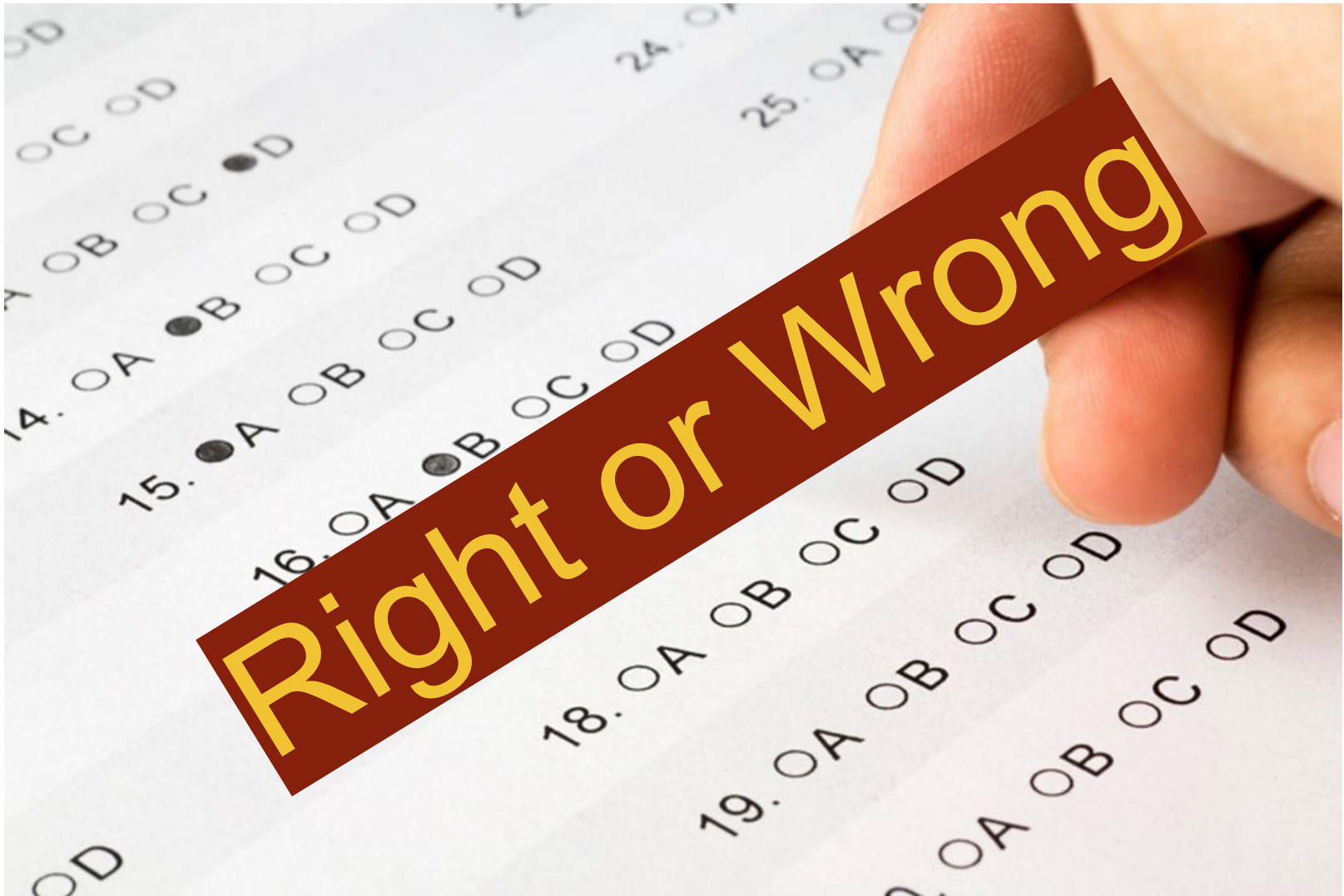
Boiling it down:

- Measuring latent “proficiency” parameter through binary tasks
- Some tasks may be more difficult
- Some tasks may be more discriminating

Exams



Exams



Is this the number 3?



Is this the number 3?



Football Games



Football Games



Triplet Question



A

Which (B or C)
is most similar
to the left (A)?



B



C

Triplet Question



A

Which (B or C)
is most
to



C

Greater or Lesser

A Quick Tour of Objective Functions

The Situations....

A few dimensions:

- Are responses **deterministic** or **probabilistic**?
- Are comparisons **head-to-head** or **via an intermediary**?
- Do you **know the answer** or is the answer **unknown**?

The Situations....

Known		Deterministic	Probabilistic
	Head-to-head	Sorting	Ranking
	Intermediary	20 Questions	Proficiency

Unknown		Deterministic	Probabilistic
	Head-to-head	SMEs	Crowdsourcing
	Intermediary	SMEs	Crowdsourcing

Head-to-Head, Non-deterministic, Known



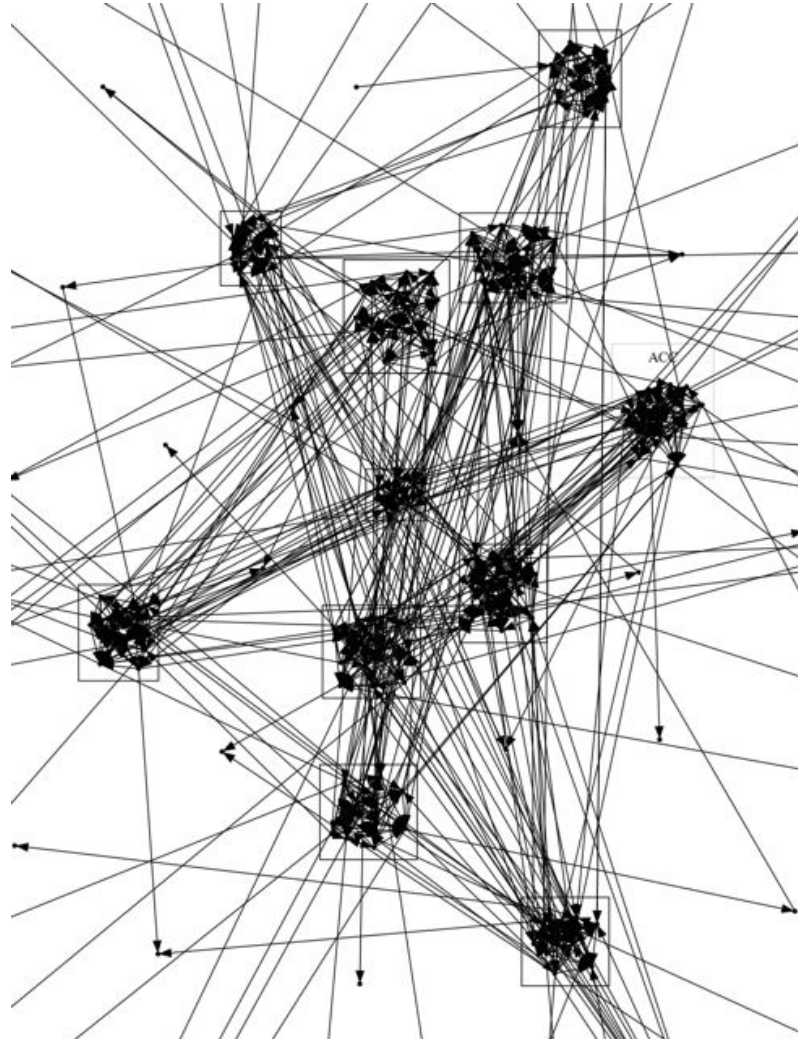
Who wins?

Head-to-Head, Non-deterministic, Known



Obvious!

Head-to-Head, Non-deterministic, Known



Head-to-Head, Non-deterministic, Known

Ranking

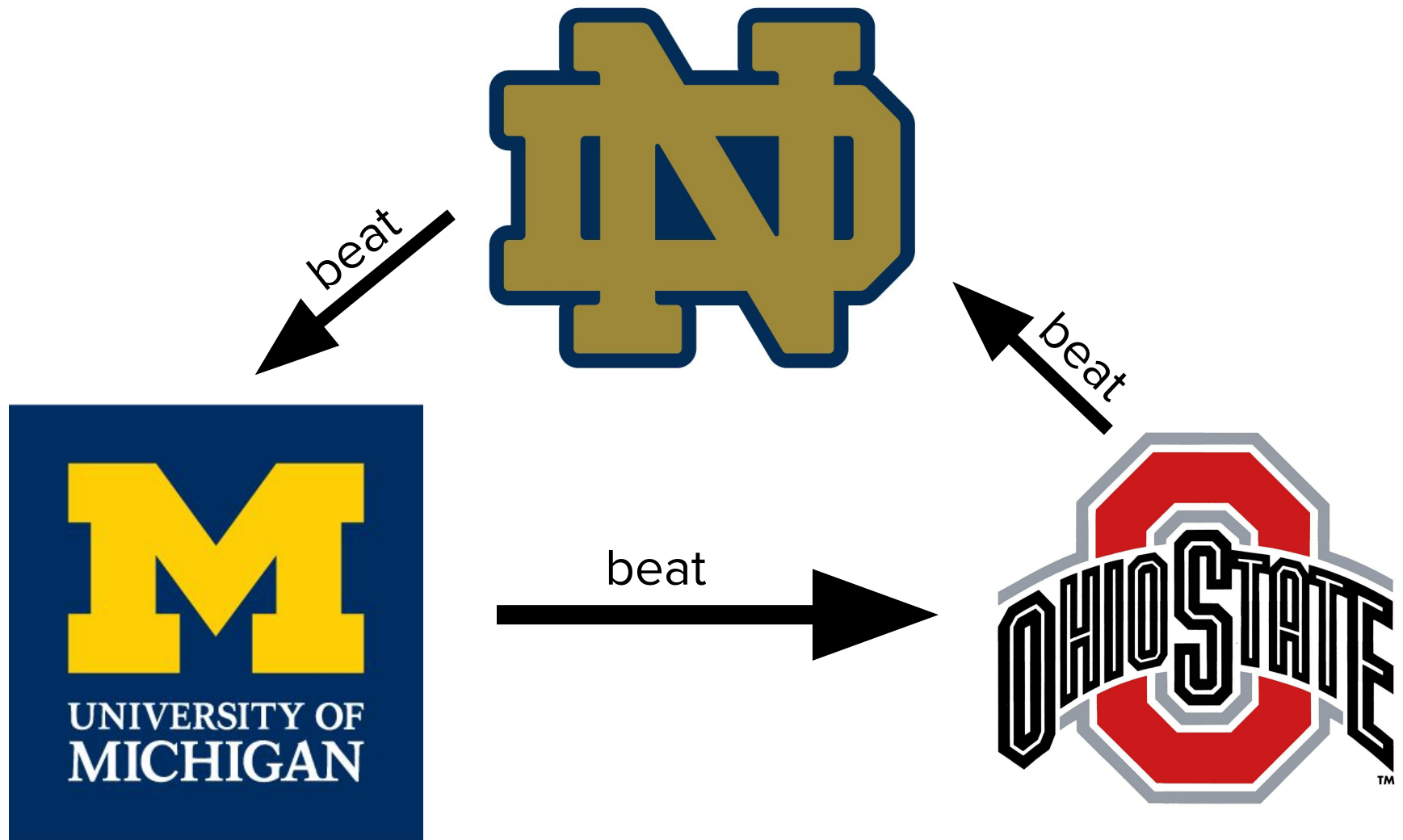
Objective function:

Head-to-Head, Non-deterministic, Known

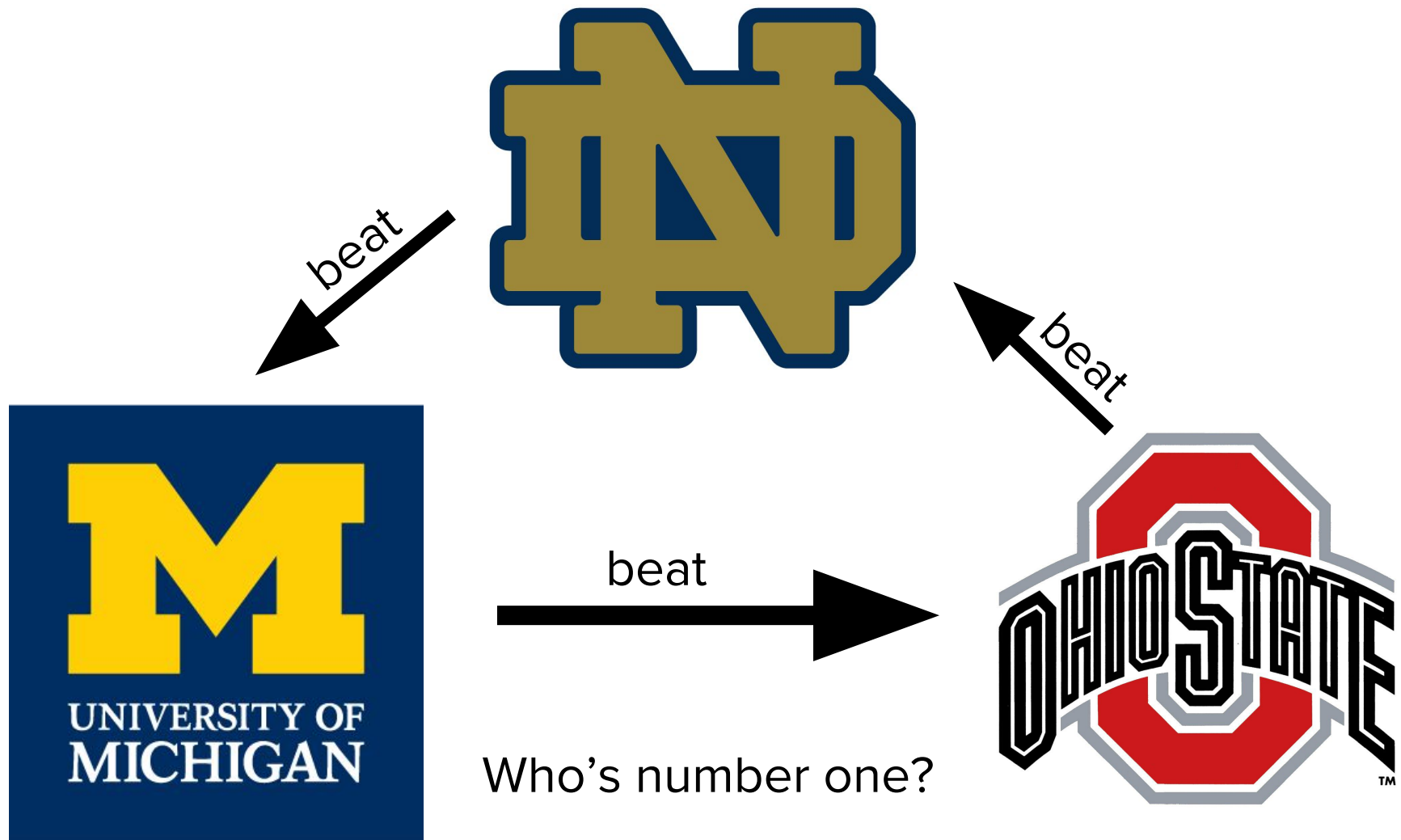
Ranking

Objective function: Minimize
number of out of order teams

Head-to-Head, Non-deterministic, Known



Head-to-Head, Non-deterministic, Known



Head-to-Head, Non-deterministic, Known

Ranking

Objective function: Minimize out of order teams

Problems: Ambiguity, no room for priors!

Head-to-Head, Non-deterministic, Known

Ranking

Objective function: Minimum

Feedback Arc Set Problem

Head-to-Head, Non-deterministic, Known

Ranking

Objective function: Minimum
Feedback Arc Set Problem

APX-Hard :- (

Head-to-Head, Non-deterministic, Known

Some approximations:

- igraph package provides Eades algorithm and exact solution
- Naeff and Nichols' recent work from the education perspective (NIPS Workshops 2014)

Head-to-Head, Non-deterministic, Known Ranking

Objective function: Define latent “awesomeness” parameter θ_t that determines probability of winning in head-to-head matchup. Maximize likelihood.

Head-to-Head, Non-deterministic, Known Ranking

Objective function: Elo's Method

$$Pr(A \text{ beats } B) = \frac{1}{1 + \exp(-(\theta_A - \theta_B))}$$

$$\operatorname{argmax}_{\theta} = \sum_{A \text{ beat } B} -\log(1 + \exp(-(\theta_A - \theta_B)))$$

Head-to-Head, Non-deterministic, Known

[(34.347870666232524, 'Georgia Southern'),
(29.838761263481281, 'Florida State'),
(25.432582458372824, 'North Dakota St'),
(23.821632753205929, 'Northern Iowa'),
(18.679517268505776, 'McNeese State'),
(18.625173744416838, 'Towson'),
(14.917821539733199, 'Fordham'),
(10.959894887882886, 'Alabama'),
(10.919577390031504, 'Auburn'),
(10.296523121897861, 'Missouri'),
(9.9650384068581026, 'Louisiana State'),
(9.9401026629153737, 'Clemson'),
(9.8242445700308938, 'South Carolina'),
(9.6859376591590625, 'Georgia'),
(9.6360200598307966, 'Ohio State'),

...

Head-to-Head, Non-deterministic, Known

[(34.347870666232524, 'Georgia Southern'),
(29.838761263481281, 'Florida State'),
(25.432582458372824, 'North Dakota St'),
(23.821632753205929, 'Northern Iowa'),
(18.679517268505776, 'McNeese State'),
(18.625173744416838, 'Towson'),
(14.917821539733199, 'Fordham'),
(10.959894887882886, 'Alabama'),
(10.919577390031504, 'Auburn'),
(10.296523121897861, 'Missouri'),
(9.9650384068581026, 'Louisiana State'),
(9.9401026629153737, 'Clemson'),
(9.8242445700308938, 'South Carolina'),
(9.6859376591590625, 'Georgia'),
(9.6360200598307966, 'Ohio State'),

...



Head-to-Head, Non-deterministic, Known

Extensions: Priors / Regularization

- Florida should have a high starting rating
- Towson should not
- Michigan should be better than them all (well, not this year....)

Makes objective convex!

Allows for domain knowledge to make up for sparsity.

Head-to-Head, Non-deterministic, Known

Extensions: Stochastic Gradient
Ascent

$$\theta'_A = \theta_A + K \cdot (1(A \text{ won}) - \text{Pr}(A \text{ beats } B))$$

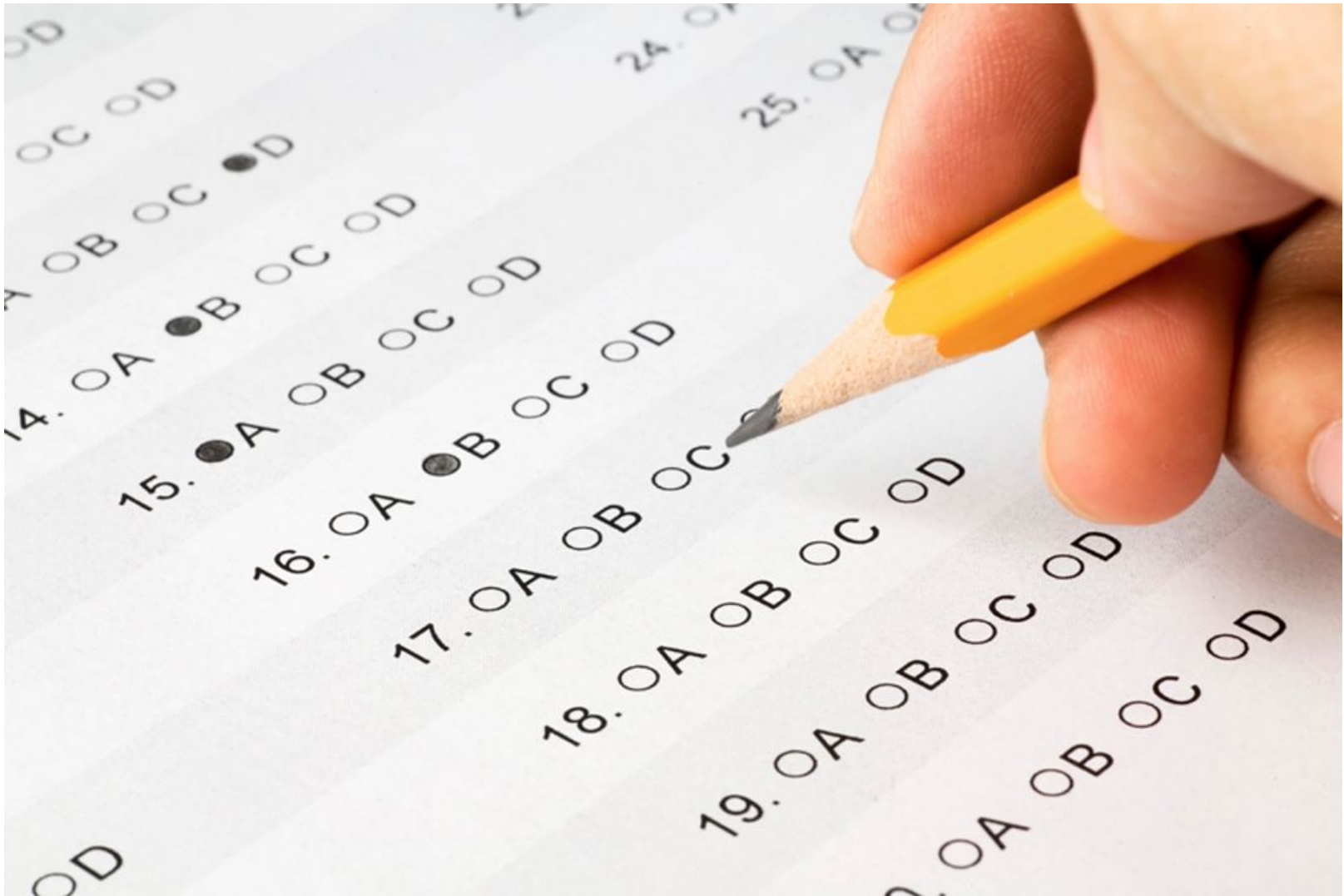
Allows for online updates with
many users (XBox, Yahoo!, etc.)

Head-to-Head, Non-deterministic, Known

But warning!

- No room for improvement
- Top of the heap has little incentive to play
- Sparsity means that priors have strong effect (which is why everybody hated the BCS)

Intermediary, Non-deterministic, Known



Intermediary, Non-deterministic, Known

Proficiency

Objective function:

Intermediary, Non-deterministic, Known Proficiency

Objective function: Define probability of answering correctly given latent proficiency parameter θ_s and latent difficulty parameter β_q .

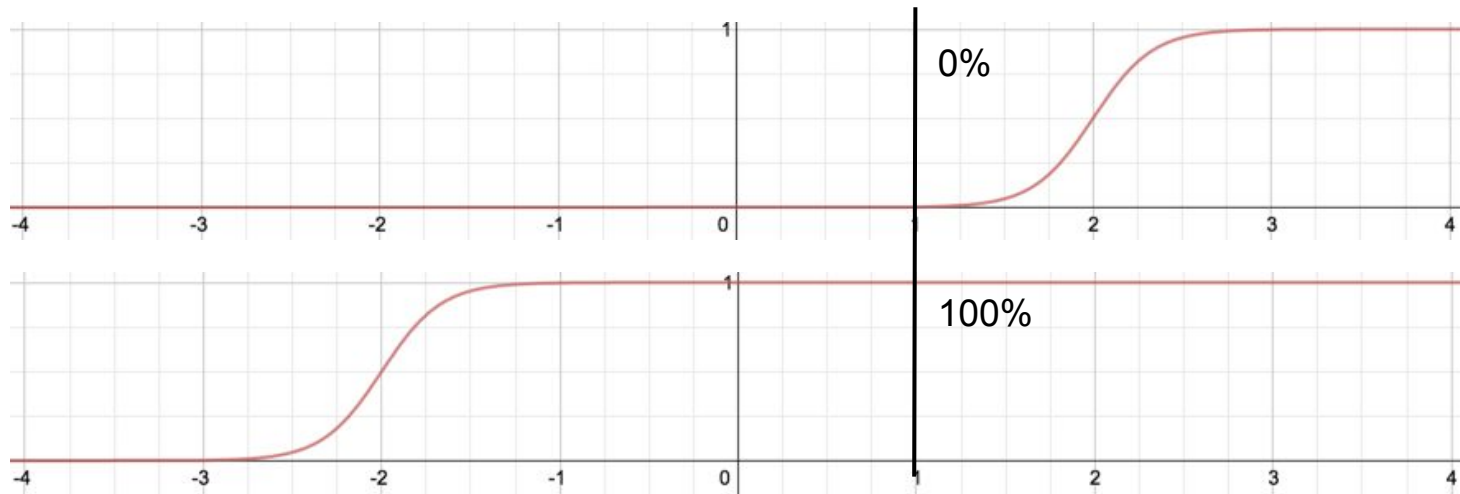
Intermediary, Non-deterministic, Known Proficiency

Objective function: Item Response
Theory (IRT)

$$Pr(s \text{ gets } q \text{ right}) = \frac{1}{1 + \exp(-(\theta - \beta))}$$

$$\operatorname{argmax}_{\theta, \beta} - \sum_{(s, q) \text{ right}} \log(1 + \exp(-(\theta - \beta))) - \sum_{(s, q) \text{ wrong}} \log(1 + \exp(-(\beta - \theta)))$$

Intermediary, Non-deterministic, Known



Intermediary, Non-deterministic, Known

Features:

- Students and questions are ranked on the same scale
- Items can be more or less difficult
- Can tailor an exam!

Intermediary, Non-deterministic, Known

Features:

- Students and questions are ranked on the same scale
- Items can be more or less difficult
- Can tailor an exam!

Problems:

- Distinguishing between features of questions
- Just right/wrong?
- Learning over time

Intermediary, Non-deterministic, Known Performance?

- Typically measured in *prediction accuracy*
 - Given a student's history, how well can you predict their future responses?
- AUC a common metric
- Typically see 10-15 point improvement in AUC for accounting for differences in questions

Intermediary, Non-deterministic, Unknown



A

Which (B or C)
is most similar
to the left (A)?



B



C

Intermediary, Non-deterministic, Unknown

Crowdsourcing

Objective function components:

- Raters have a skill level
- Trust skilled raters more (likelihood of a skilled rater getting a correct answer is higher, even on unknown data)
- Choose classification that maximizes likelihood of responses

Intermediary, Non-deterministic, Unknown

Skill level?

- Fixed set of questions with known answers
- Then can use IRT to assess skill level!

Intermediary, Non-deterministic, Unknown

What does trust mean?

- Really hard question
- HCOMP is a conference dedicated to issues like this.

Thank You.

Kevin H. Wilson

Principal Data Scientist

Knewton, Inc.

kevin@knewton.com

khwilson@gmail.com

Code: github.com/khwilson/pydata2014