Teaching Innovations based on Social Science Research

Gary King
Institute for Quantitative Social Science

Talk at the Harvard Board of Overseers Meeting, 2/5/2012

The Context

What is Harvard's Biggest Threat?

Social Science Principles for Teaching Innovation

Social Science Principles for Teaching Innovation

1. Social connections motivate

Social Science Principles for Teaching Innovation

1. Social connections motivate
    - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!

Social Science Principles for Teaching Innovation

1. Social connections motivate
   - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
   - E.g., Voter turnout drives fail, unless you hear about your neighbors

Social Science Principles for Teaching Innovation

1. Social connections motivate
   - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
   - E.g., Voter turnout drives fail, unless you hear about your neighbors
   - People mind wander: 50% of the time

1. Social connections motivate
    - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
    - E.g., Voter turnout drives fail, unless you hear about your neighbors
    - People mind wander: 50% of the time; During conversation: only 25%

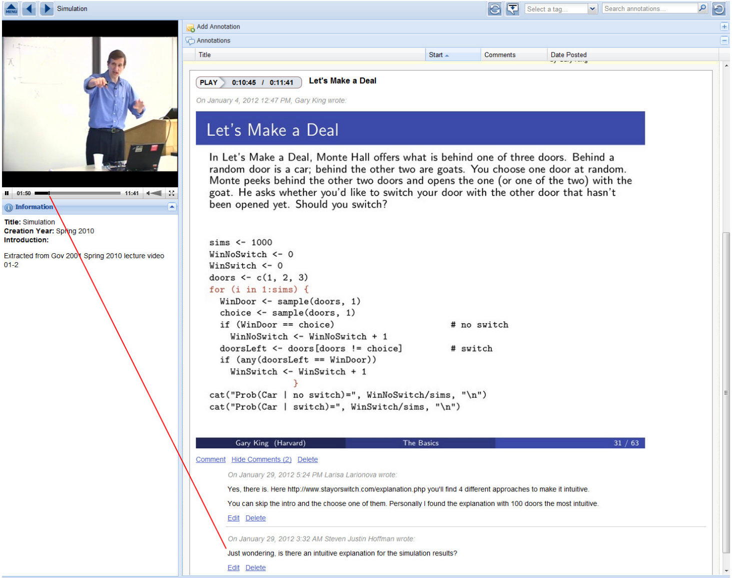
Social Science Principles for Teaching Innovation

1. Social connections motivate
    - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
    - E.g., Voter turnout drives fail, unless you hear about your neighbors
    - People mind wander: 50% of the time; During conversation: only 25%
2. Teaching teaches the teacher! (Help students teach each other)

Social Science Principles for Teaching Innovation

1. Social connections motivate
    - E.g., easier to get you to recycle (for the community) than to exercise (so you can live longer)!
    - E.g., Voter turnout drives fail, unless you hear about your neighbors
    - People mind wander: 50% of the time; During conversation: only 25%
2. Teaching teaches the teacher! (Help students teach each other)
3. Instant feedback improves learning

Collaborative Video Annotation



Collaborative Text Annotation

22 **2 Conceptualizing uncertainty and inference**

distinguish between these two cases, I refer to the hypothetical parameter value as $\tilde{\theta}$ and the single unobserved true value as θ. In the next chapter, I will introduce $\hat{\theta}$ as a point estimator for θ, based on the maximum of the likelihood with respect to $\tilde{\theta}$. $\hat{\theta}$ is a number in a single experiment, but a random variable across hypothetical experiments.

The *likelihood* that a hypothetical model (summarized by the hypothetical parameter value $\tilde{\theta}$) produced the data we observe, given $\mathcal{M}^*$, is denoted $L(\tilde{\theta}|y,\mathcal{M}^*)$, where $\mathcal{M}^*$ may again be suppressed since it appears in all subsequent expressions. The *likelihood axiom* then defines this concept as follows:

$$\begin{aligned} L(\tilde{\theta}|y,\mathcal{M}^*) &\equiv L(\tilde{\theta}|y) \\ &= k(y)\Pr(y|\tilde{\theta}) \\ &\propto \Pr(y|\tilde{\theta}). \end{aligned} \tag{2.5}$$

In the second line of this equation, $k(y)$ is an unknown function of the data; since it is not a function of $\tilde{\theta}$, it is treated as an unknown positive constant. In the third line, "$\propto$" means "is proportional to." The third line is only a more convenient way of writing the second without the constant. For a given set of observed data, $k(y)$ remains the same over all possible hypothetical values of $\tilde{\theta}$. However, $k(y)$ is a function of y and therefore may change as y changes. The likelihood $L(\tilde{\theta}|y)$ is similar to the concept of inverse probability in that it permits one to measure and compare the uncertainty one has about alternative hypothetical values of $\tilde{\theta}$. However, the unknown value $k(y)$ ensures that likelihood is a relative rather than an absolute measure of uncertainty. This likelihood axiom is but one way to make the measure explicitly relative. Indeed, one could use any monotonic function of $\Pr(y|\tilde{\theta})$.[5] The choice represented in Equation (2.5) is arbitrary, just as is the choice of making the scale of probability range between 0 and 1. The advantage of likelihood is that it can be calculated from a traditional probability, whereas inverse probability cannot be calculated in any way.

If the data are continuous rather than discrete, the likelihood is calculated in the same way, except that the underlying probability distribution is now a density. Hence, a more general way to write the formula is as follows:

1 thread on page 2

[2] I'm trying to understand what "under essentially the same conditions" means. When we say

1 thread on page 3

[2] If the meaning of "model" is the same here as it was in the first chapter, this appears to

3 threads on page 4

[3] According to an earlier statement in this book, that exact probability should be known to

[2] Likelihood is a measure of relative uncertainty, but relative to what? If it is the relati

[5] I'm a bit uncomfortable about this apparent interchangeability of "M" as a model and "M" a

2 threads on page 6

[1] Why is this always our null hypothesis then? Does it make sense to have more sophisticated

[2] how was this equation derived?

1 thread on page 7

[1] Yes! I am always answering this question while teaching stat to others.

1 thread on page 8

[1] I don't see why it's necessarily impossible. For example, you could know that in the past

4 threads on page 9

[6] Shouldn't it be "is NOT a number"?

[3] I'm not really clear on what k(y) *is*, besides some unknowable constant. Is it Pr(~theta

[5] I have a conflict between my understanding (or misunderstanding...) of likelihood inferenc

[2] What if it's not monotonic?

Shouldn't it be "is NOT a number"? *Larisa Larionova* - 28 Jan, 08:27PM

I think so. *Tracey Shollenberger* - 29 Jan, 01:32AM

It's a number, because once you've run the single experiment, it just *is* whatever value came from that experiment. It's no longer a variable, it's a constant thing that you know - a number. *Jesse Heitner* - 29 Jan, 03:08AM

Maybe my English deceives me, but it seems like, it should be either:

"is NOT a number, but a variable" or
"is a number, but NOT a variable".
So, do you mean, Jesse, that it's the latter? *Larisa Larionova* - 29 Jan, 11:24AM

Thanks for your post, Jesse. The structure of the sentence confused me, too, Larisa. I now think it means that there are two cases - In the case of a single experiment, it is a number. In the case of multiple experiments (not sure why they're "hypothetical"), it's not a single number, but a random variable. *Tracey Shollenberger* - 29 Jan, 02:11PM

I think they're hypothetical because you don't necessarily run them. You just think of theta_hat as something that could vary if you ran your experiment several times. Eg, in a linear regression, you get only one estimate of beta, your effect, but you also use statistical theory to calculate its standard error, which measures the hypothetical variation it would show if you ran the regression on several samples from the same population. *Ekedi Mpondo-Dika* - 29 Jan, 06:56PM

Gov2001 Email Search: March 2003 Archives

Messages sorted by: [subject] [author] [date]

- [gov2001-l] HW 2 Code *Andrew Reeves*
  - [gov2001-l] HW 2 Code *Kosuke Imai*
- [gov2001-l] hw3 Question 2 *Kosuke Imai*
- [gov2001-l] question 2b *Mee-Jung Jang*
  - [gov2001-l] question 2b *Kosuke Imai*
    - [gov2001-l] loading state files in R *Stanislav Markus*
      - [gov2001-l] loading state files in R *Kosuke Imai*
      - [gov2001-l] loading state files in R *Olivia Lau*
      - [gov2001-l] loading state files in R *Kosuke Imai*
    - [gov2001-l] question 2b *Daniel Hopkins*
      - [gov2001-l] question 2b *Kosuke Imai*
- [gov2001-l] back to basics: Question 1(b) *Ryan Thomas Moore*
  - [gov2001-l] back to basics: Question 1(b) *Kosuke Imai*
    - [gov2001-l] back to basics: Question 1(b) *Phillip Y. Lipscy*
      - [gov2001-l] back to basics: Question 1(b) *Kosuke Imai*
- [gov2001-l] 2(a) *Traci Burch*
  - [gov2001-l] 2(a) *Jennifer_Fitzgerald*

[gov2001-l] likelihood vs loglikelihood

Phillip Y. Lipscy lipscy at fas.harvard.edu
Wed Mar 5 00:37:25 EST 2003

```
For 2b, if we use the product term of the negative binomial distribution to
estimate the likelihood rather than the loglikelihood, am I right to assume that
we should get the same result for the maximum point?  i.e. we use loglikelihood
instead of likelihood to get rid of the product term, but our results should not
change?

Thanks,
Phillip.
```

Gov2001 Email Search: March 2003 Archives

Messages sorted by: [subject] [author] [date]

- [gov2001-l] HW 2 Code *Andrew Reeves* — **Boston University**
  - [gov2001-l] HW 2 Code *Kosuke Imai*
- [gov2001-l] hw3 Question 2 *Kosuke Imai*
- [gov2001-l] question 2b *Mee-Jung Jang* — **Cleary Gottlieb Steen & Hamilton LLP**
  - [gov2001-l] question 2b *Kosuke Imai*
    - [gov2001-l] loading state files in R *Stanislav Markus* — **University of Chicago**
      - [gov2001-l] loading state files in R *Kosuke Imai* — **Princeton University**
      - [gov2001-l] loading state files in R *Olivia Lau* — **Google, Inc**
      - [gov2001-l] loading state files in R *Kosuke Imai*
    - [gov2001-l] question 2b *Daniel Hopkins* — **Georgetown University**
      - [gov2001-l] question 2b *Kosuke Imai*
- [gov2001-l] back to basics: Question 1(b) *Ryan Thomas Moore* — **University of Washington, St Louis**
  - [gov2001-l] back to basics: Question 1(b) *Kosuke Imai*
    - [gov2001-l] back to basics: Question 1(b) *Phillip Y. Lipscy* — **Yale University**
      - [gov2001-l] back to basics: Question 1(b) *Kosuke Imai*
- [gov2001-l] 2(a) *Traci Burch* — **Northwestern University**
  - [gov2001-l] 2(a) *Jennifer_Fitzgerald* — **University of Colorado**

[gov2001-l] likelihood vs loglikelihood

Phillip Y. Lipscy lipscy at fas.harvard.edu
Wed Mar 5 00:37:25 EST 2003

```
For 2b, if we use the product term of the negative binomial distribution to
estimate the likelihood rather than the loglikelihood, am I right to assume that
we should get the same result for the maximum point?  i.e. we use loglikelihood
instead of likelihood to get rid of the product term, but our results should not
change?

Thanks,
Phillip.
```

What's left to do in class?

What's left to do in class? Focus on what they don't know!

- (Intensely participatory) discussions on topics they think they're confused about

What's left to do in class? Focus on what they don't know!

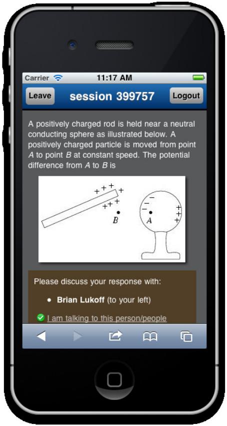
- (Intensely participatory) discussions on topics they think they're confused about
- Lecture on topics they are confused about

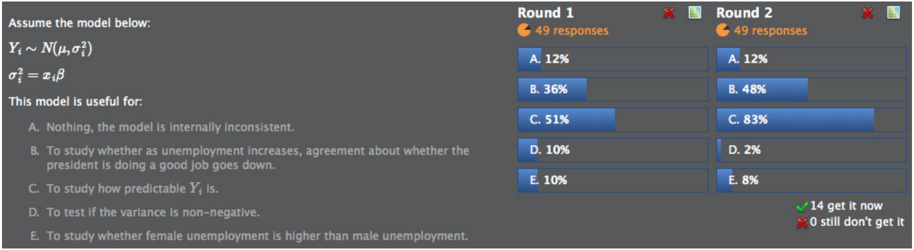
What's left to do in class? Focus on what they don't know!

- (Intensely participatory) discussions on topics they think they're confused about
- Lecture on topics they are confused about
- A version of "peer instruction"

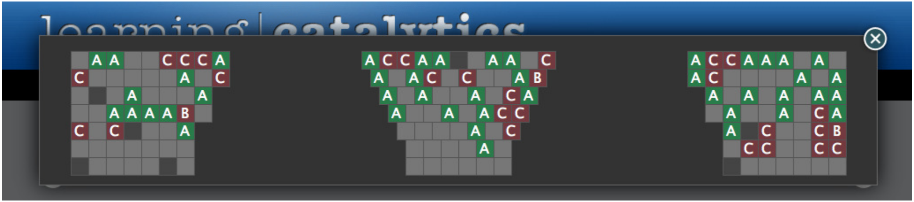
What's left to do in class? Focus on what they don't know!

- (Intensely participatory) discussions on topics they think they're confused about
- Lecture on topics they are confused about
- A version of "peer instruction"

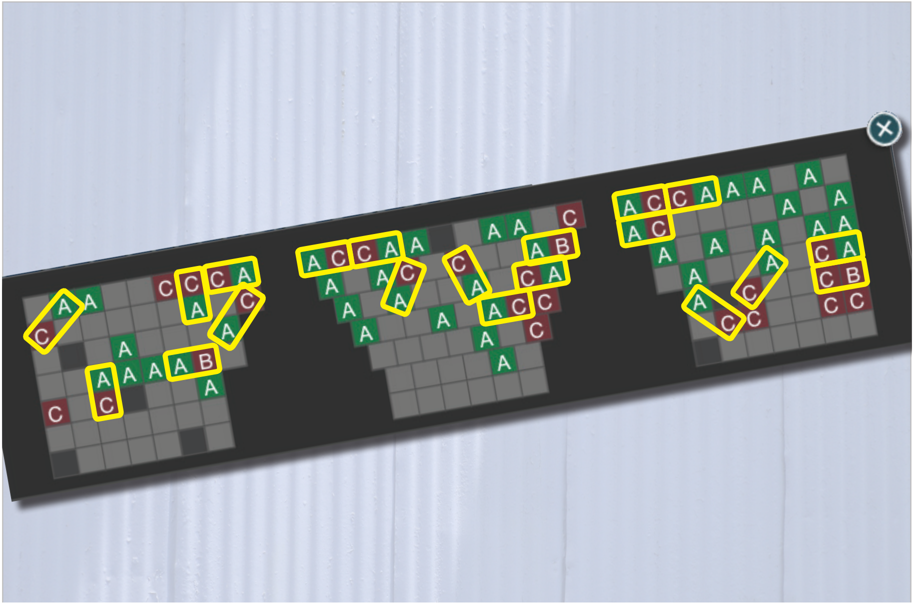


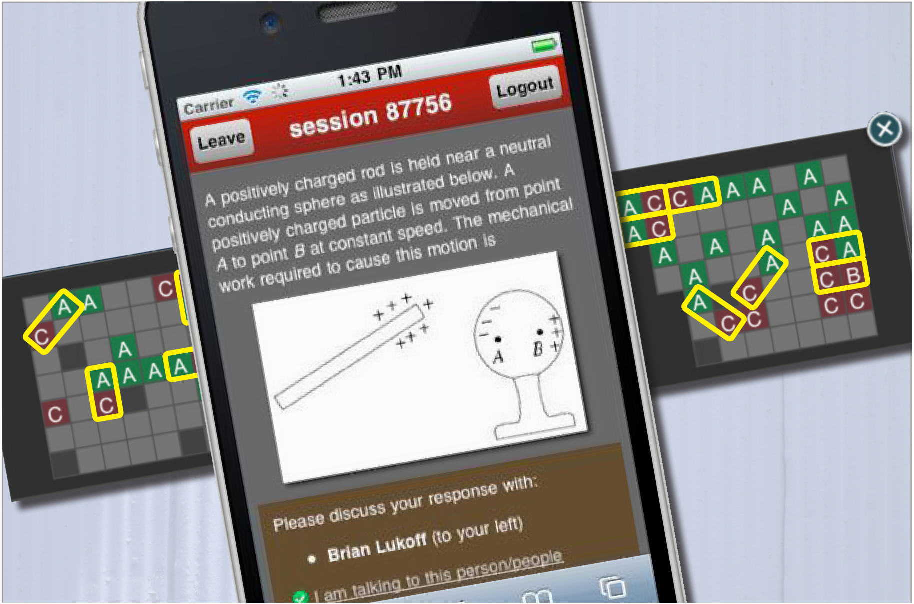


Assume the model below:
$Y_i \sim N(\mu, \sigma_i^2)$
$\sigma_i^2 = x_i\beta$
This model is useful for:
A. Nothing, the model is internally inconsistent.
B. To study whether as unemployment increases, agreement about whether the president is doing a good job goes down.
C. To study how predictable Y_i is.
D. To test if the variance is non-negative.
E. To study whether female unemployment is higher than male unemployment.
Round 1
49 responses
A. 12%
B. 36%
C. 51%
D. 10%
E. 10%
Round 2
49 responses
A. 12%
B. 48%
C. 83%
D. 2%
E. 8%
14 get it now
0 still don't get it

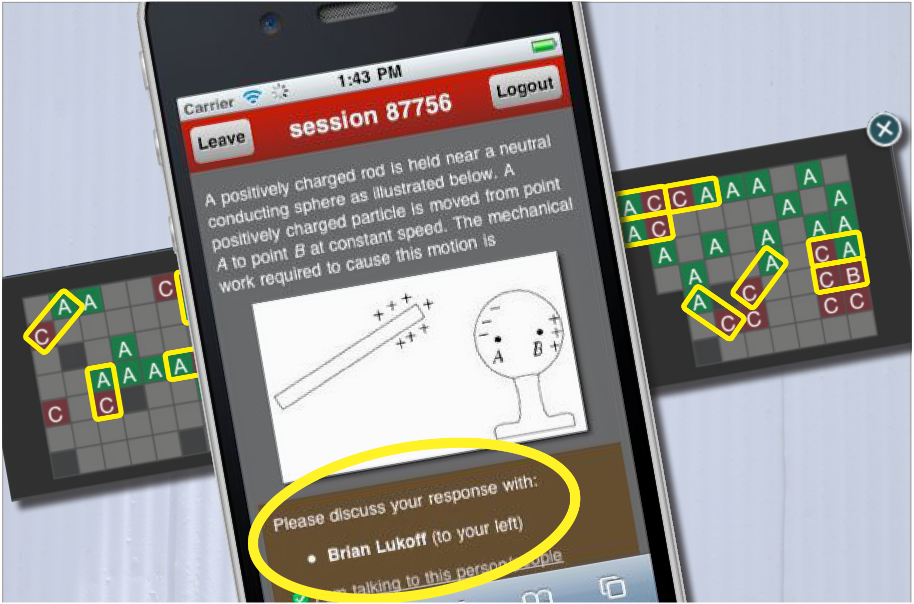


A A C C C A
C A C
A A
A A A A B
C C A
A C C A A A A C
A A C C A B
A A A C A
A A A C C
A C
A
A C C A A A A
A C A A
A A A A A
A A C A
A C C B
C C C C





Carrier
1:43 PM
Leave
session 87756
Logout
A positively charged rod is held near a neutral conducting sphere as illustrated below. A positively charged particle is moved from point A to point B at constant speed. The mechanical work required to cause this motion is
Please discuss your response with:
Brian Lukoff (to your left)
I am talking to this person/people



Carrier
1:43 PM
Leave
session 87756
Logout
A positively charged rod is held near a neutral conducting sphere as illustrated below. A positively charged particle is moved from point A to point B at constant speed. The mechanical work required to cause this motion is
A
B
Please discuss your response with:
Brian Lukoff (to your left)

Social Science Principles

Social Science Principles

1. Social connections motivate

Social Science Principles

1. Social connections motivate
2. Teaching teaches the teacher

Social Science Principles

1. Social connections motivate
2. Teaching teaches the teacher
3. Instant feedback improves learning

Social Science Principles

1. Social connections motivate
2. Teaching teaches the teacher
3. Instant feedback improves learning
4. The advantages of large scale measurement and analysis