

# What I've Learned About Learning

Peter Norvig  
Director of Research, Google

**(1) Motivation,  
Not Information**





Love is a better teacher than duty  
- Albert Einstein

# Motivation, not Information

- Willpower (hyperbolic discounting)
- Due dates (self-commitment)
- Peer support (forums, outside groups)
- Faculty encouragement (email)
- Pride of accomplishment
- Authenticity of course/instructors/institution
- Early adopter (community)



## Timeline (successful Coursera course)

ramp up + exponential decay + due dates  
synchronous, evergreen,  
semi-synchronous, bus route

**(2) It's the Learner, Stupid**



• Learning results from what the student does and thinks and only from what the student does and thinks. The teacher can advance learning only by influencing what the student does to learn.

- Herb Simon

*For when you don't know what's going on*

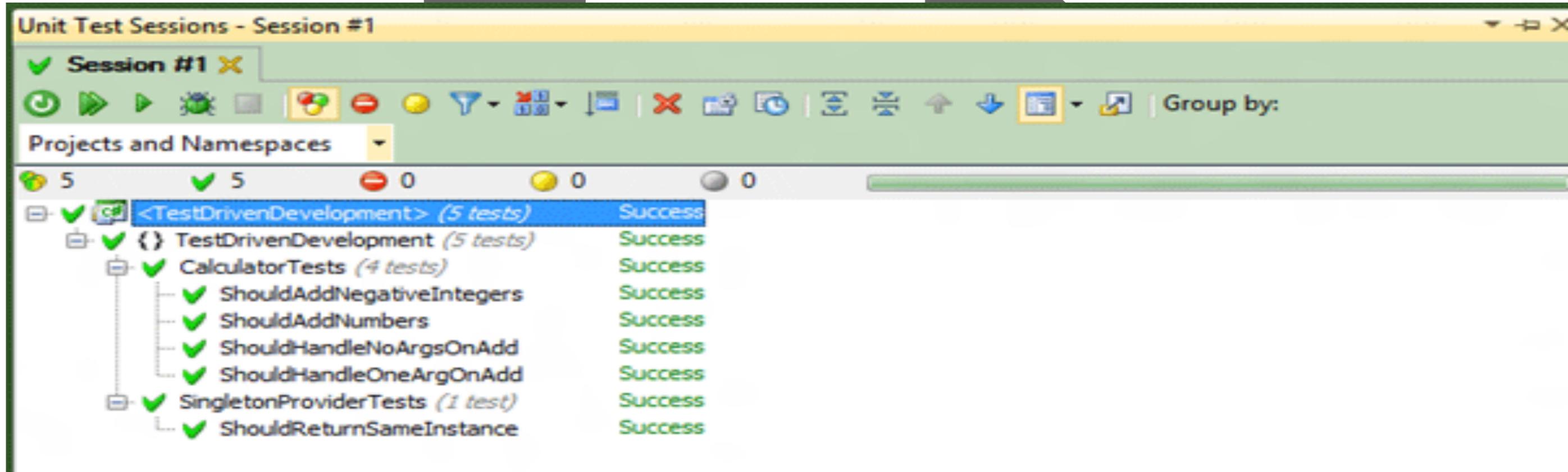
## 20.2.5 Learning Bayes net structures

So far, we have assumed that the structure of the Bayes net is given and we are just trying to learn the parameters. The structure of the network represents basic causal knowledge about the domain that is often easy for an expert, or even a naive user, to supply. In some cases, however, the causal model may be unavailable or subject to dispute—for example, certain corporations have long claimed that smoking does not cause cancer—so it is important to understand how the structure of a Bayes net can be learned from data. This section gives a brief sketch of the main ideas.

The most obvious approach is to *search* for a good model. We can start with a model containing no links and begin adding parents for each node, fitting the parameters with the methods we have just covered and measuring the accuracy of the resulting model. Alternatively, we can start with an initial guess at the structure and use hill-climbing or simulated annealing search to make modifications, retuning the parameters after each change in the structure. Modifications can include reversing, adding, or deleting links. We must not introduce cycles in the process, so many algorithms assume that an ordering is given for the variables, and that a node can have parents only among those nodes that come earlier in the ordering (just as in the construction process in Chapter 14). For full generality, we also need to search over possible orderings.

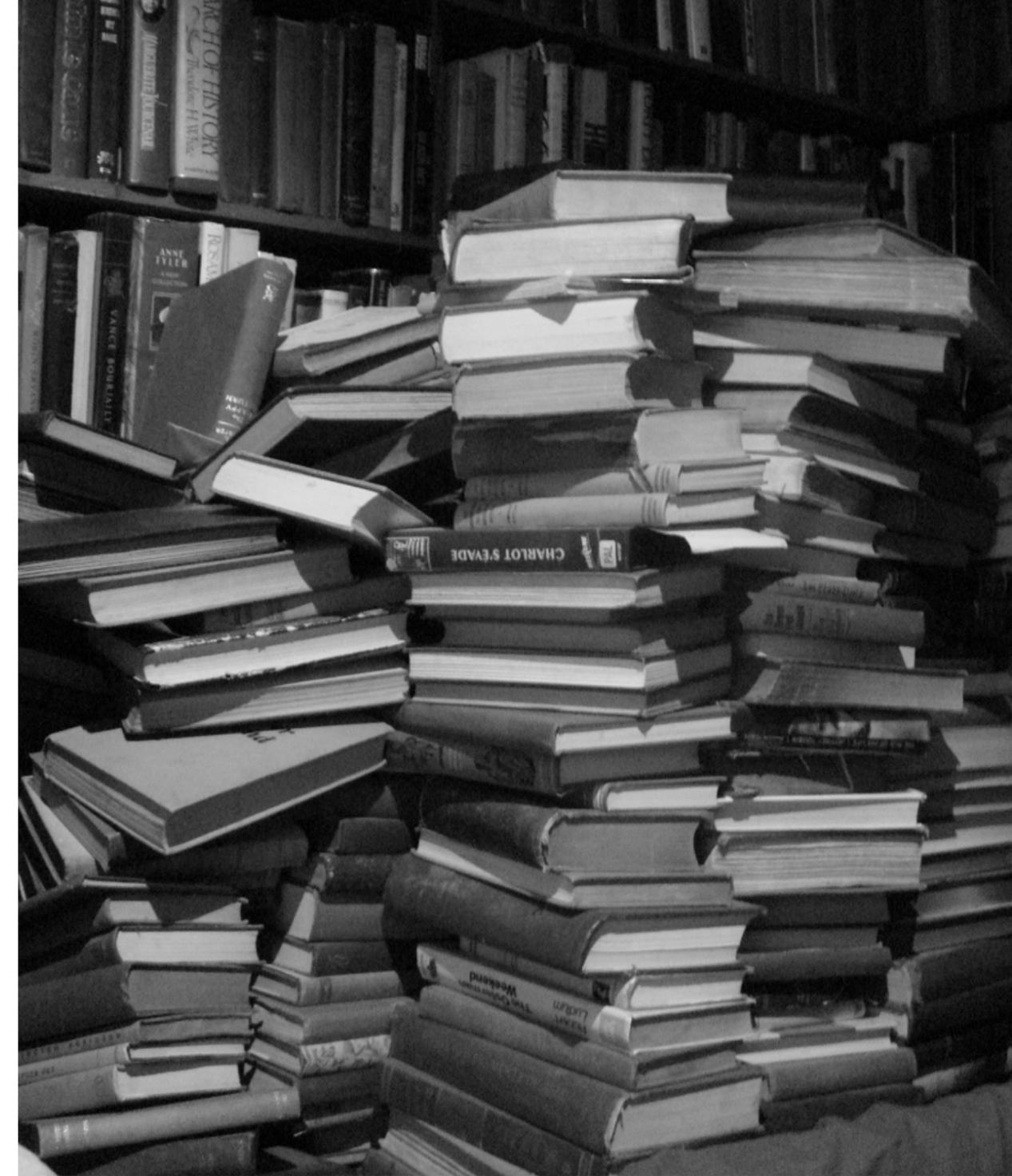
## **“what the student does”**

- Problem first, then explanation (Peer learning)
- Make students predict (see it is wrong)
- (They didn't like that.)
- Experiments with more open-ended problems



- “Richard Hamming told me his secret: First get together the problem sets and exams that you want the students to be able to solve. Then write a book that will teach them how to solve them.” -- Hal Varian (1993)

# (3) Mastery Learning



Hal Abelson

photo by austinevan

# The 2 Sigma Problem: The Search for Methods of Group Instruction as Effective as One-to-One Tutoring

BENJAMIN S. BLOOM

*University of Chicago and Northwestern University*

Two University of Chicago doctoral students in education, Anania (1982, 1983) and Burke (1984), completed dissertations in which they compared student learning under the following three conditions of instruction:

1. *Conventional.* Students learn the subject matter in a class with about 30 students per teacher. Tests are given periodically for marking the students.

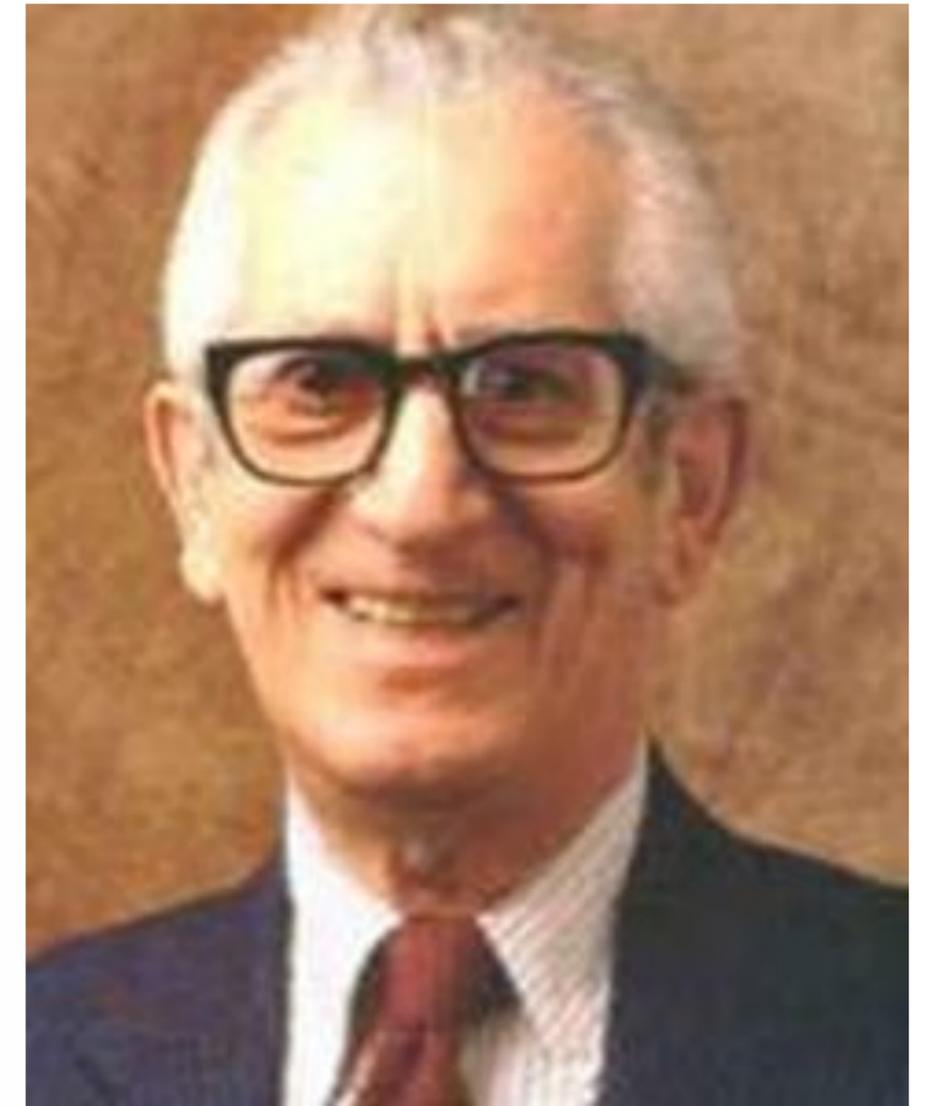
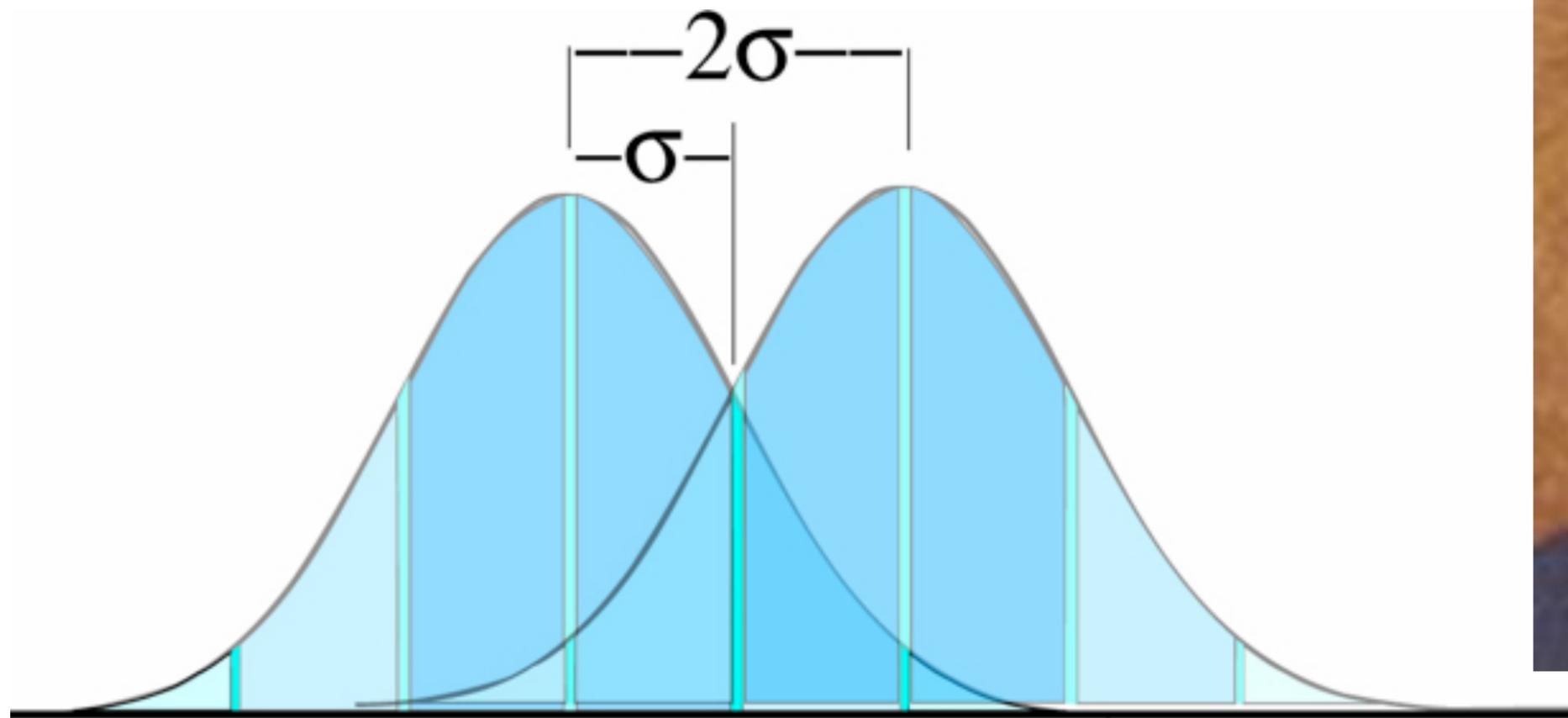
2. *Mastery Learning.* Students learn the subject matter in a class with about 30 students per teacher. The instruction is the same as in the conventional class (usually with the same teacher). Formative tests (the

The students were randomly assigned the three learning conditions, and their initial aptitude tests scores, previous achievement in the subject, and initial attitudes and interests in the subject were similar. The amount of time for instruction was the same in all three groups except for the corrective work in the mastery learning and tutoring groups. Burke (1984) and Anania (1982, 1983) replicated the study with four different samples of students at grades four, five, and eight and with two different subject matters, Probability and Cartography. In each sub-study, the instructional treatment was limited to

students under conventional instructional conditions. (See Figure 1.)

There were corresponding changes in students' time on task in the classroom (65% under conventional instruction, 75% under Mastery Learning, and 90+% under tutoring) and students' attitudes and interests (least positive under conventional instruction and most positive under tutoring). There were great reductions in the relations between prior measures (aptitude or achievement) and the summative achievement measures. Typically, the aptitude-achievement correlations changed from +.60 under conventional to +.35 under

- The average under *tutoring* [with *mastery learning*] was about two standard deviations above the average of the control. (Benjamin Bloom, 1984)



Was  
Bloom  
Right?

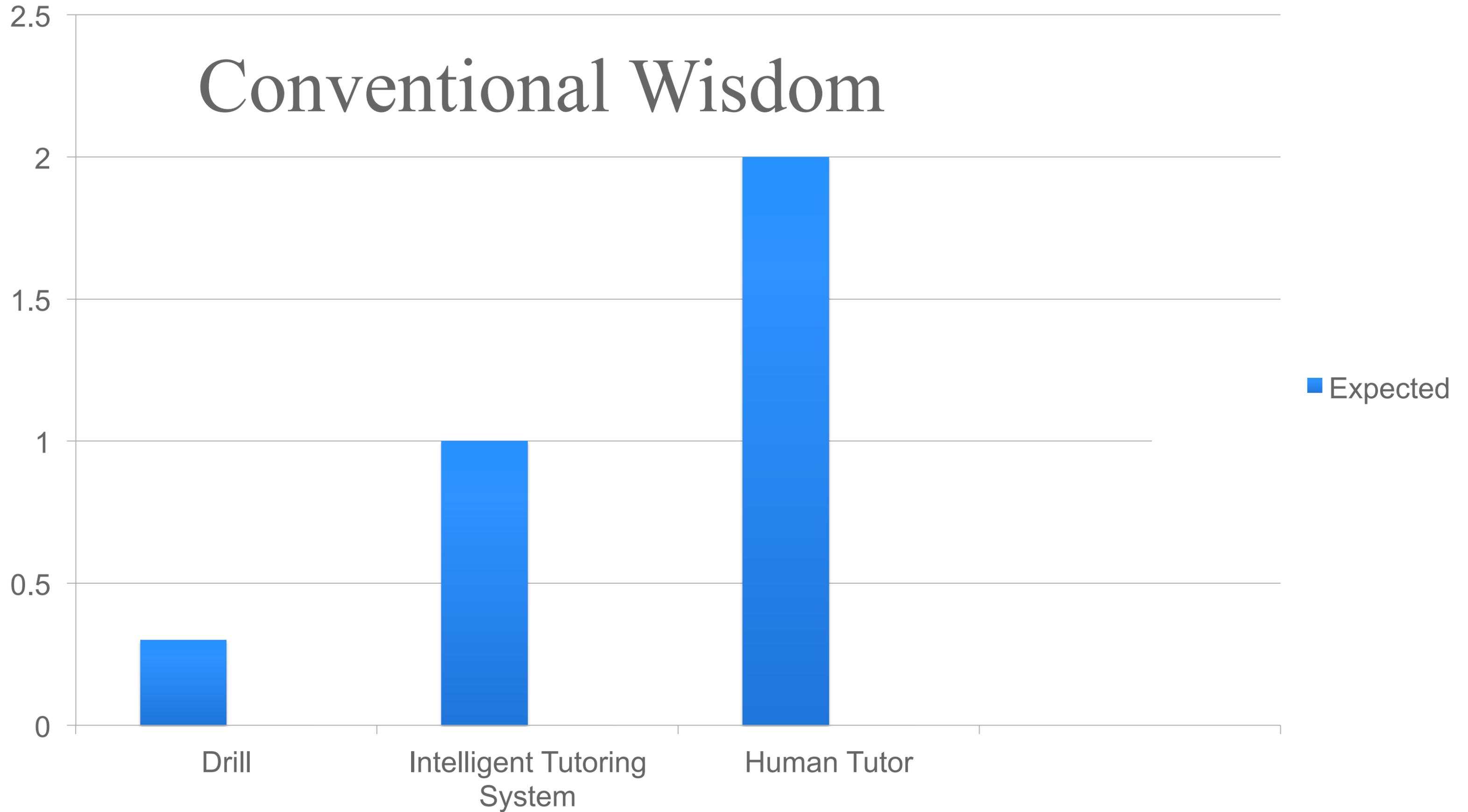
# The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems

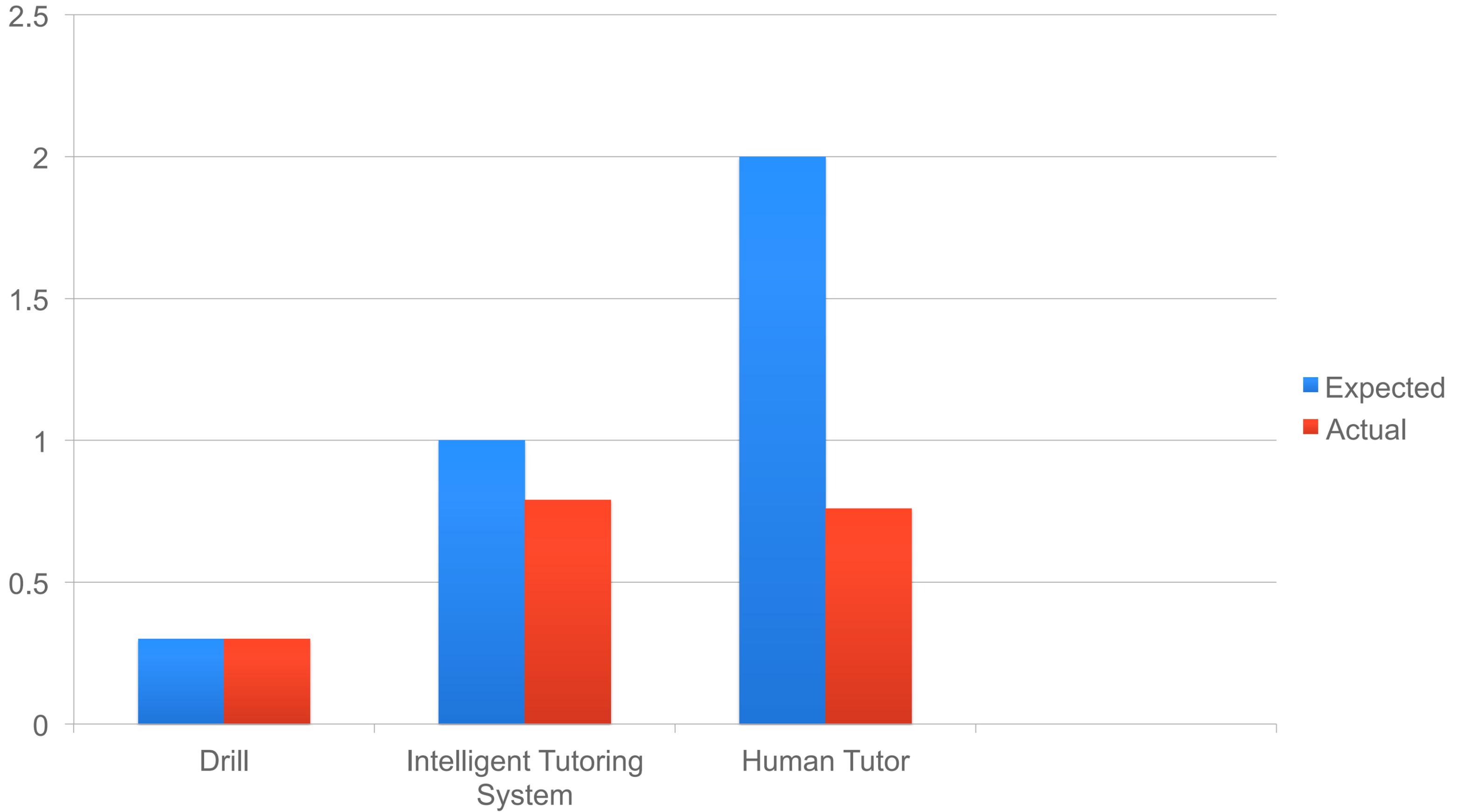
Kurt VanLehn

*Computing, Informatics and Decision Systems Engineering  
Arizona State University*

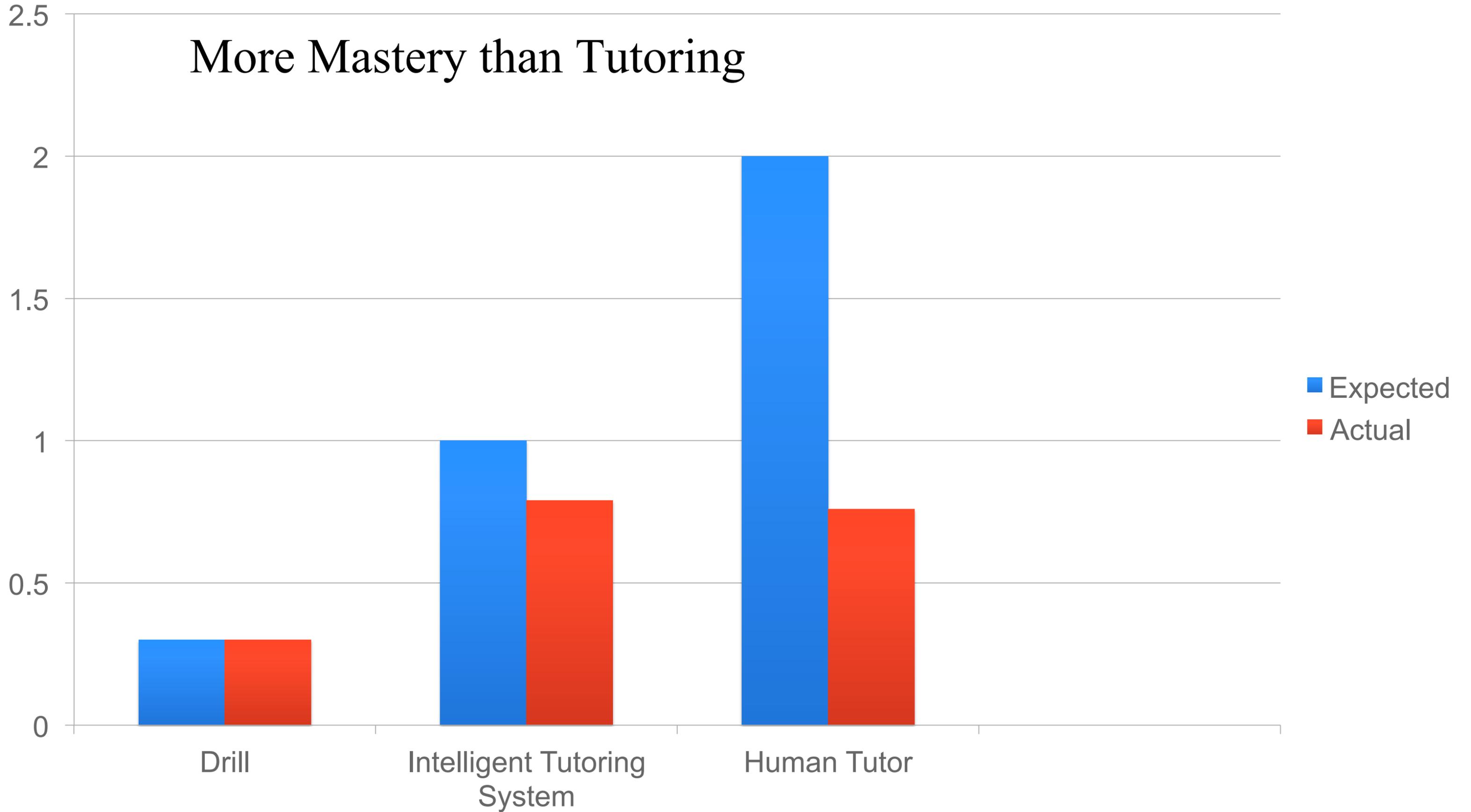
This article is a review of experiments comparing the effectiveness of human tutoring, computer tutoring, and no tutoring. “No tutoring” refers to instruction that teaches the same content without tutoring. The computer tutoring systems were divided by their granularity of the user interface interaction into answer-based, step-based, and substep-based tutoring systems. Most intelligent tutoring systems have step-based or substep-based granularities of interaction, whereas most other tutoring systems (often called CAI, CBT, or CAL systems) have answer-based user interfaces. It is widely believed as the granularity of tutoring decreases, the effectiveness increases. In particular, when compared to No tutoring, the effect sizes of answer-based tutoring systems, intelligent tutoring systems, and adult human tutors are believed to be  $d = 0.3$ , 1.0, and 2.0 respectively. This review did not confirm these beliefs. Instead, it found that the effect size of human tutoring was much lower:  $d = 0.79$ . Moreover, the effect size of intelligent tutoring systems was 0.76, so they are nearly as effective as human tutoring.

# Conventional Wisdom

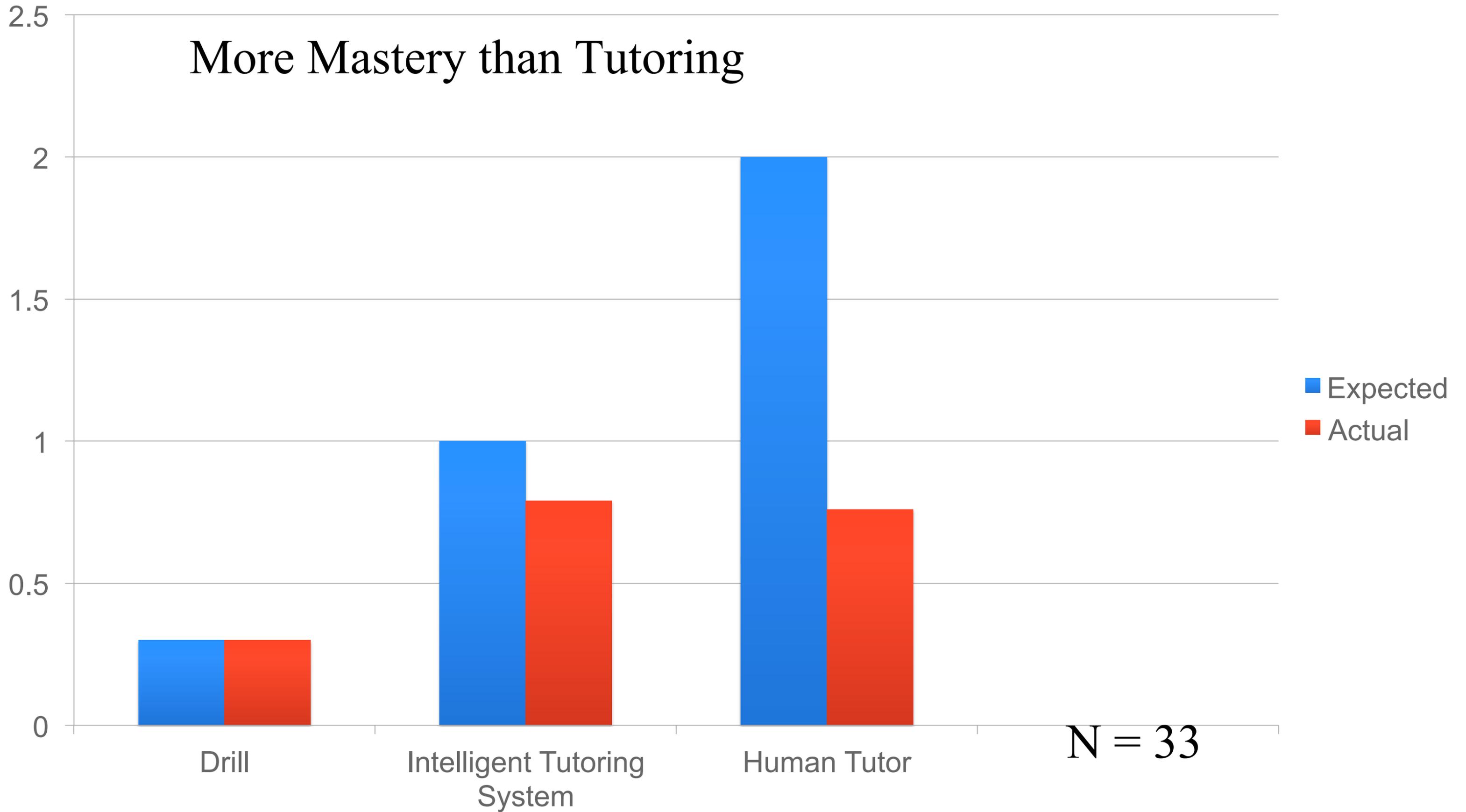




# More Mastery than Tutoring



# More Mastery than Tutoring



# (4) Disruptive Technologies Disaggregate Industries

# Disaggregation

- Research vs. Teaching
- Learning vs. Evaluation
- Admission vs. Learning vs. Graduation

(5) We (or at Least I)  
Don't Know  
What Media to Use

# Statistical Machine Translation

SEHR GEEHRTER GAST!  
KUNST, KULTUR UND  
KOMFORT IM HERZEN  
BERLIN.

DEAR GUESTS,  
ART, CULTURE AND  
LUXURY IN THE HEART  
OF BERLIN.

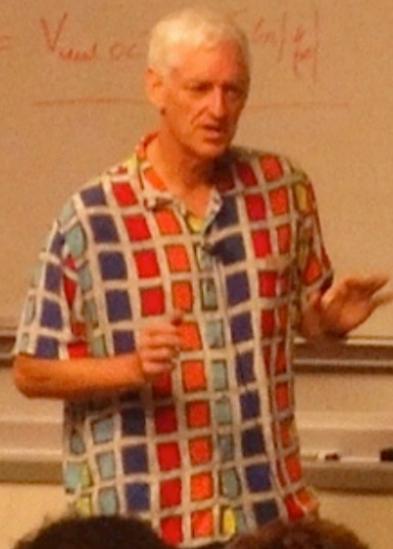
DIE ÖRTLICHE  
NETZSPANNUNG BETRÄGT  
220/240 VOLT BEI 50  
HERTZ.

THE LOCAL VOLTAGE  
IS 220/240 VOLTS 50 H

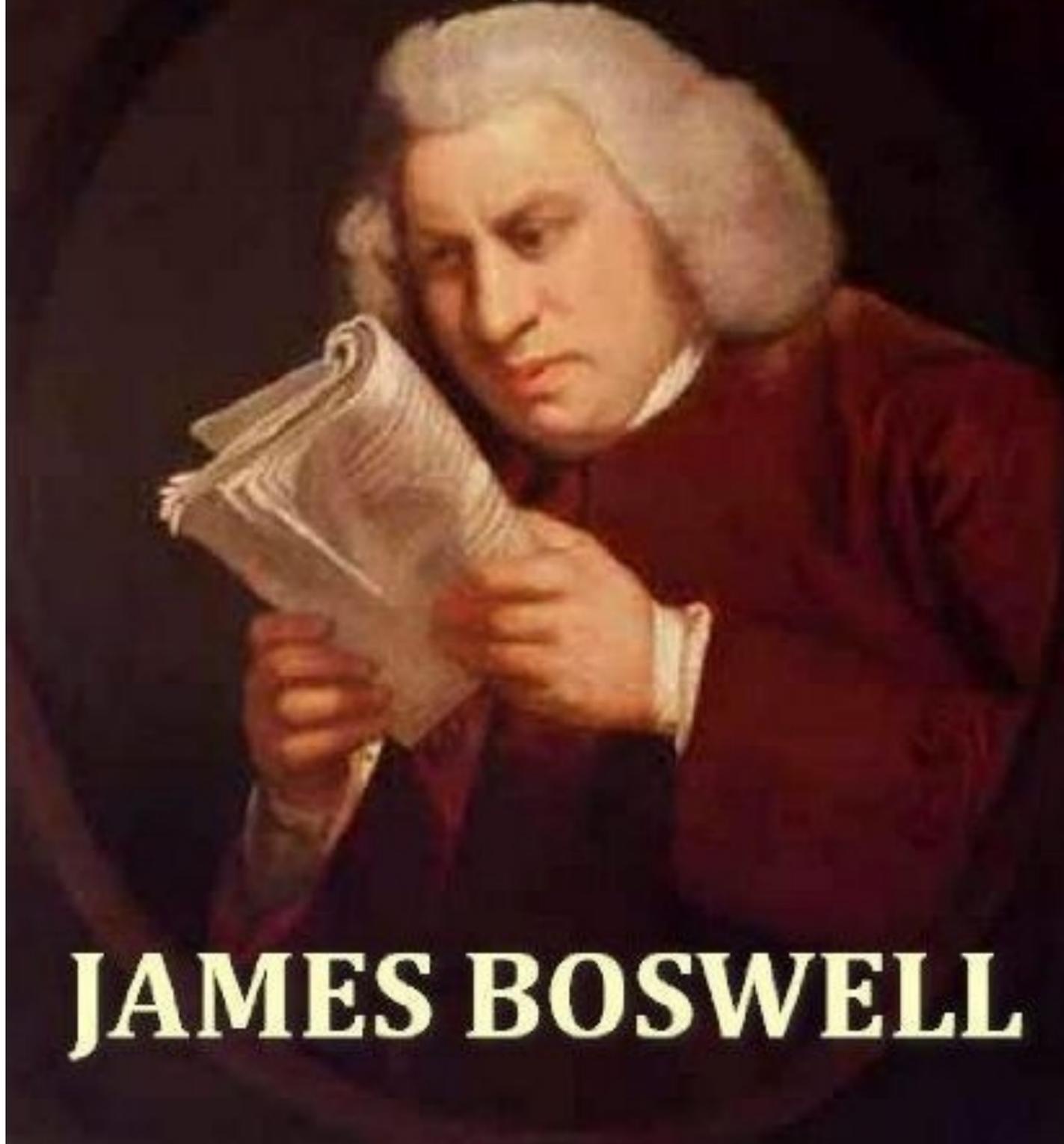
EN

DE

$$V_{oc} = V_{ind} \cos(\phi)$$



# LIFE OF JOHNSON



**JAMES BOSWELL**

Lectures were once useful, but now, when all can read, and books so numerous, lectures are unnecessary. If your attention fails and you miss part of a lecture it is lost; you cannot go back as you do upon a book.

- 1791

- Challenge 2 E
- Challenge 3 M
- Challenge 4 M
- Challenge 5 M
- Challenge 6 H
- Challenge 7 E
- Challenge 8 E
- Challenge 9 M
- Challenge 10 M
- Challenge 11 M
- Challenge 12 H

## Step 1: Explore skills relevant for this challenge.

### Lessons

- Google Scholar ▶ Video (3:10) ▶ Text
- Legal Scholar ▶ Video (6:21) ▶ Text
- Google News / News Archive ▶ Video (4:36) ▶ Text
- Using site structure ▶ Video (3:21) ▶ Text
- Comparing multiple sources ▶ Video (6:43) ▶ Text
- All skills ▶ Videos ▶ Texts
- Quick Reference Guide

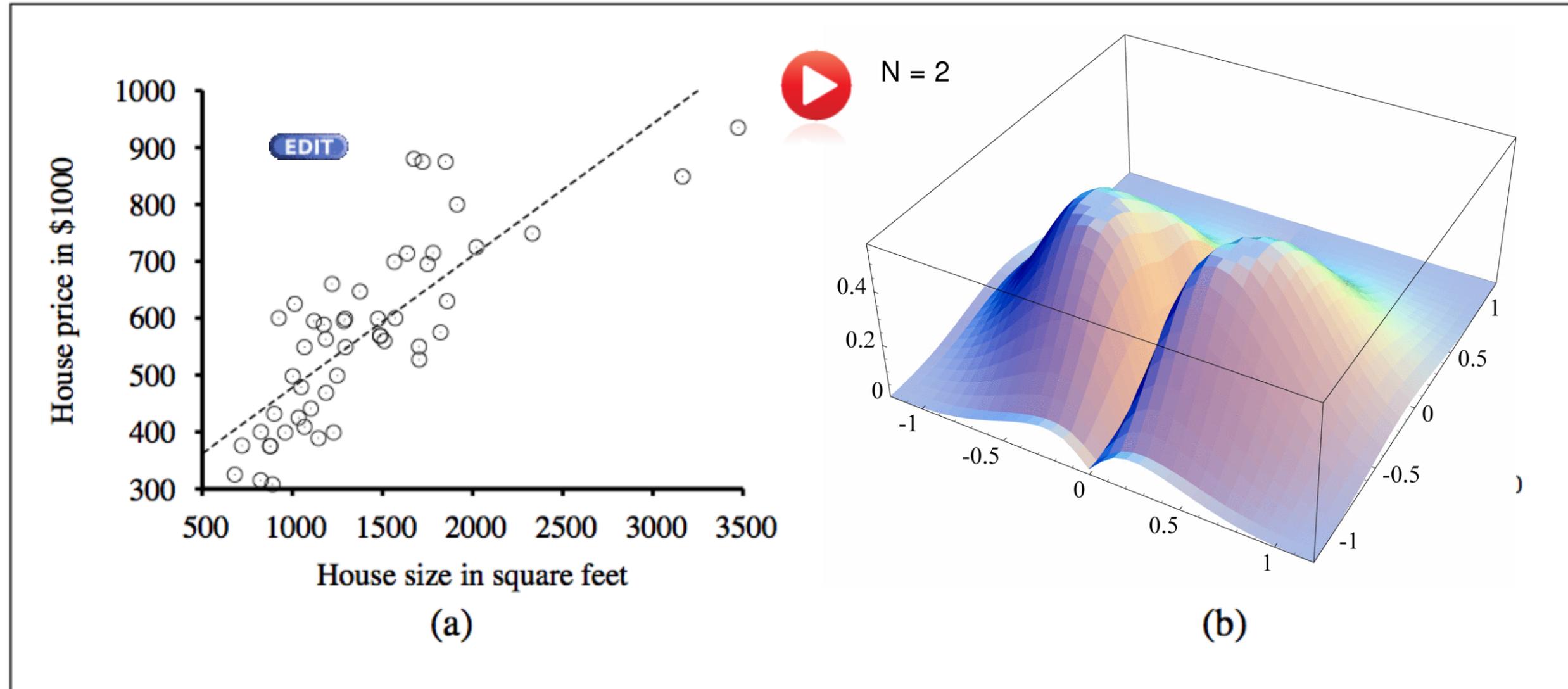
### Examples

- ▶ King Tut's curse
- ▶ Which anthropologists?
- ▶ PT Barnum's notebook

## Step 2: Solve the challenge and check your answer.

Hints

Show Answer



**Figure 18.13** (a) Scatter plots of price versus floor space of houses for sale in Berkeley, CA, in July 2009, and the linear function hypothesis that minimizes squared error loss:  $y = 0.232x + 200$ ; (b) Plot of the loss function  $\sum_j (w_1 x_j + w_0 - y_j)^2$  for various values of  $w_0, w_1$ . Note that the loss function is convex, with a single global minimum.









# Programming as a Media

```
def qsort(A):
    if len(A) <= 1:
        return A
    LT, EQ, GT = []
    pivot = A[randrange(len(A))]
    for x in A:
        if x < pivot:
            LT.append(x)
        elif x > pivot:
            GT.append(x)
        else:
            EQ.append(x)
    return qsort(LT) + EQ + qsort(GT)
```

```
def qsort(A):
    if len(A) <= 1:
        return A
    LT = EQ = GT = []
    pivot = A[randrange(len(A))]
    for x in A:
        if x < pivot:
            LT.append(x)
        elif x > pivot:
            GT.append(x)
        else:
            EQ.append(x)
    return qsort(LT) + EQ + qsort(GT)
```

```
def qsort(A):
    if len(A) <= 1:
        return A
    LT, EQ, GT = [], [], []
    pivot = A[randrange(len(A))]
    for x in A:
        if x < pivot:
            LT.append(x)
        elif x > pivot:
            GT.append(x)
        else:
            EQ.append(x)
    return qsort(LT) + EQ + qsort(GT)
```

(6) We can build big things;  
Can we build a big course?

# Authored Multipath Sequence

## Basic Genetics Refresher (OPTIONAL)

Genes and alleles (10 min)

Mendelian inheritance (13 min)

What is DNA? (12 min)

## Modern Genomics

Modern sequencing methods (11 min)

History: The Human Genome Project (12 min)

Genomic economics (8 min)

Personalized medicine (13 min)

## Commercial Genomics (OPTIONAL)

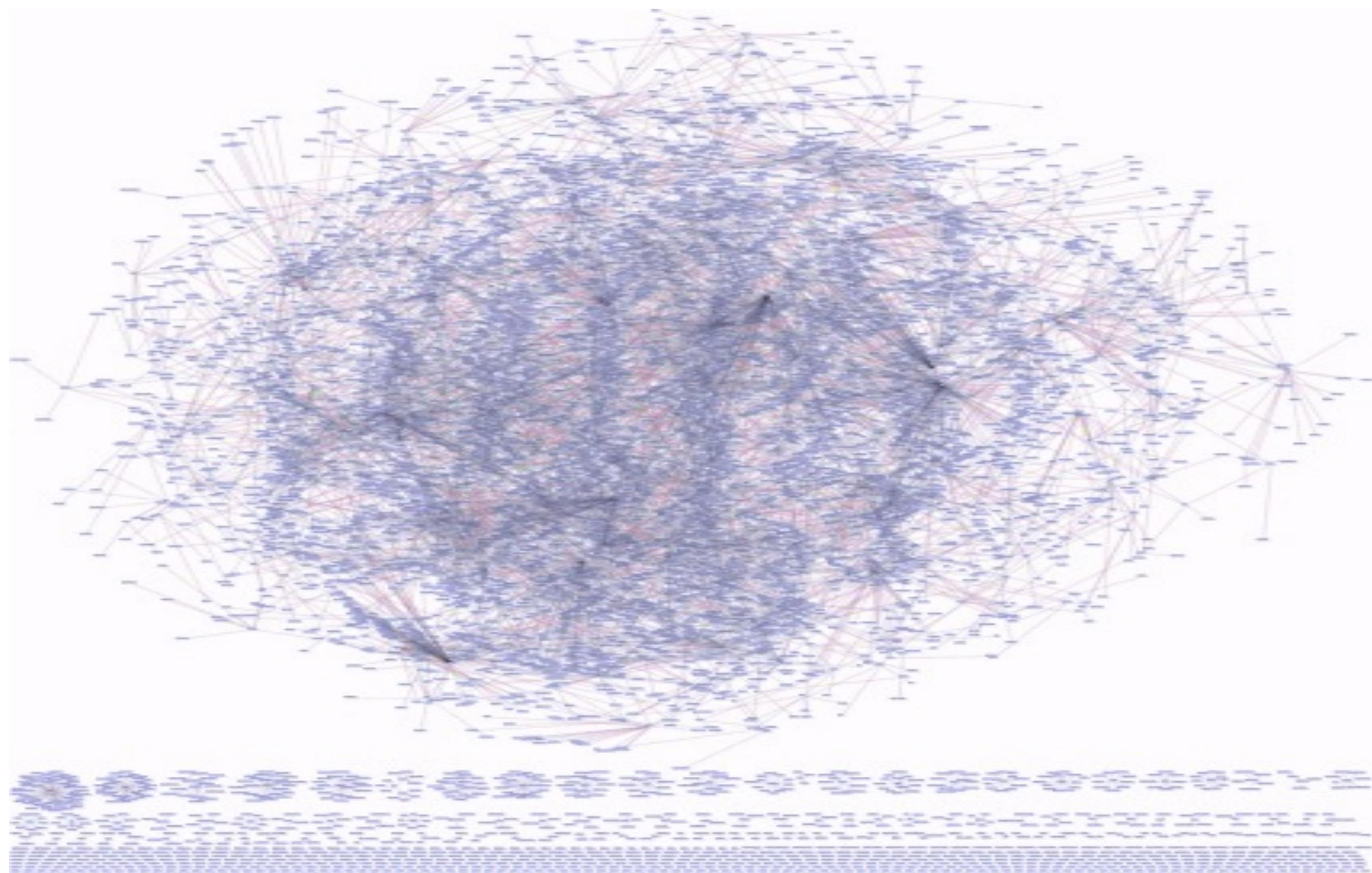
Genetic testing in the commercial world (11 min)

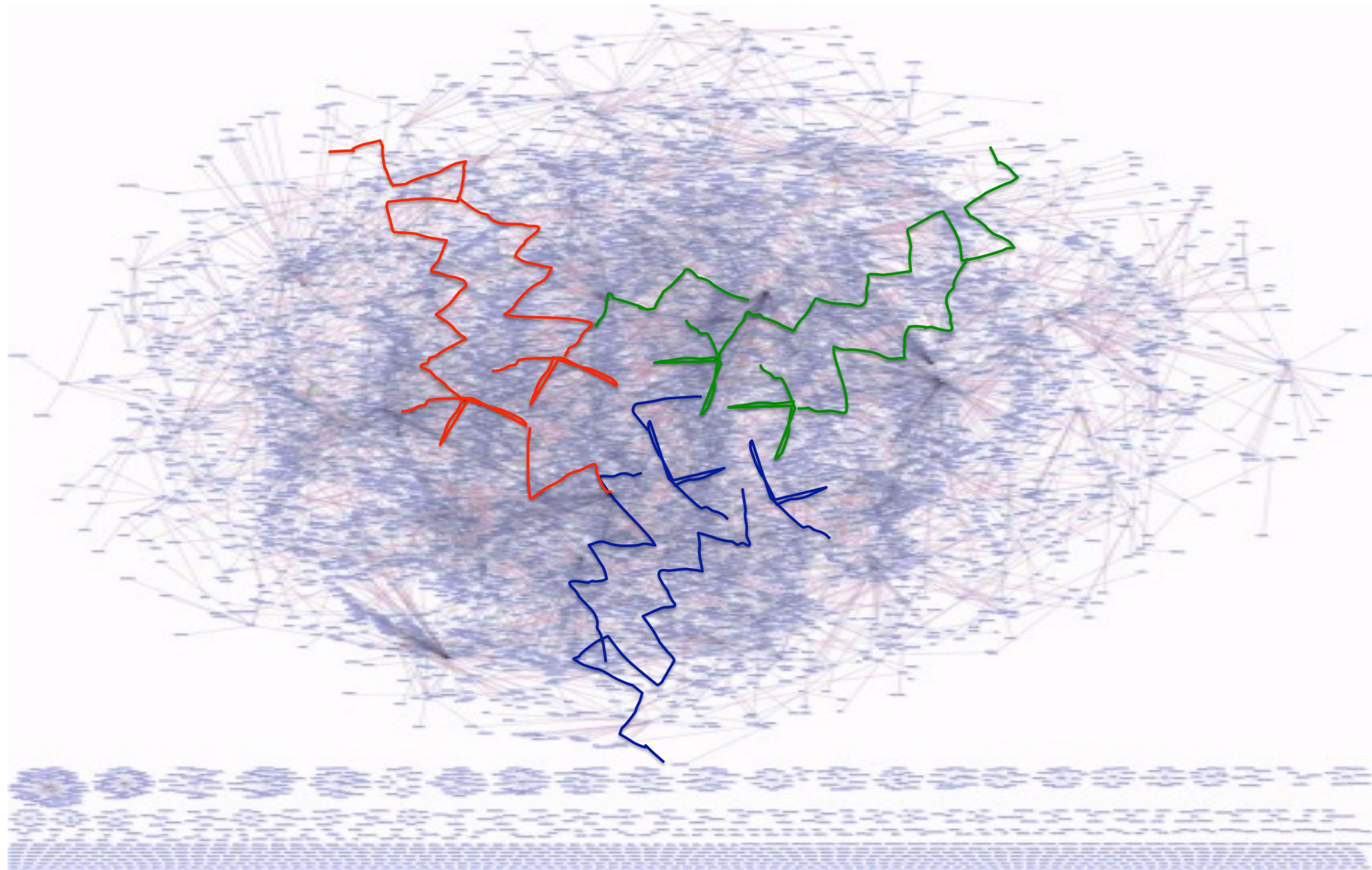
Protecting privacy (9 minutes)

Case study: direct-to-consumer genetics (12 min)

Case study: family planning (13 min)







# Courseware Engineering?





