Lecture 5:
Designing efficient systems

Measuring and optimising human performance through quantitative experimental methods.

Overview of the course

- Theory driven approaches to HCI
- Design of visual displays
- Goal-oriented interaction
- Designing smart systems (guest lecturer)
- **Designing efficient systems**
- Designing meaningful systems (guest lecturer)
- Evaluating interactive system designs
- Designing complex systems

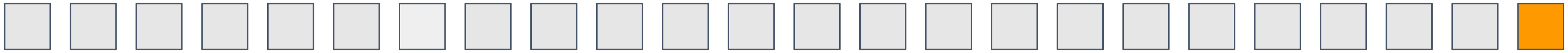
Lessons from text entry - recap

- It’s possible to model human action
- It’s possible (in part) to predict human action
- Efficiency can be predicted, and also measured
- A really fundamental trade-off:
    - Speed versus accuracy

Fitts’ Law

User actions are information-constrained

How many bits of information to select one of these choices?



How many bits of information to select one of these choices?

The human neuromotor system is limited by information rate - size of target relative to movement

Demonstration of Fitts' Law

Fitts' Law – the only equation in HCI!

- How long does it take to point at something?

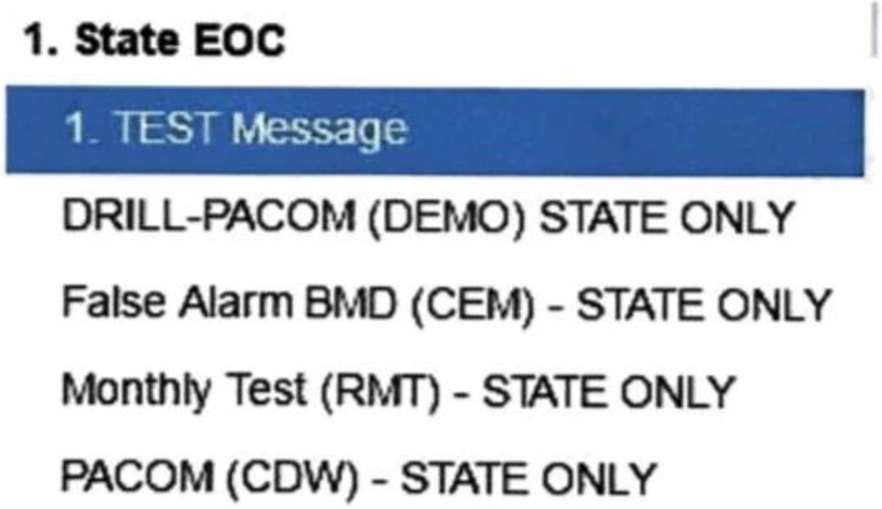
- Proportional to the **D**istance to target
- Inversely proportional to **W**idth of target
- Like most human performance (and most things in information theory), it's a log function:

- Time = k log (2D/W)

Speed-accuracy tradeoff

- Users are capable of doing things faster
- But making more mistakes as a result
- Did your application need speed, or accuracy?

By Anna Frodesiak - Own work, Public Domain, https://commons.wikimedia.org/w/index.php?curid=11443870



https://theoutline.com/post/2954/user-interface-designers-are-horrified-by-hawaii-s-missile-alert-system?zd=1

Hacking Fitt's Law: "semantic pointing"

Renaud Blanch, Yves Guiard and Michel Beaudouin-Lafon. **Semantic Pointing: Improving Target Acquisition with Control-Display Ratio Adaptation.** In *Proceedings of CHI 2004*, pages 519-526, Vienna - Austria, April 2004.

Small changes can have a big effect (1972)

Psychological Evaluation of Two Conditional Constructions Used in Computer Languages

M. E. SIME, T. R. G. GREEN AND D. J. GUEST

NEST solution:

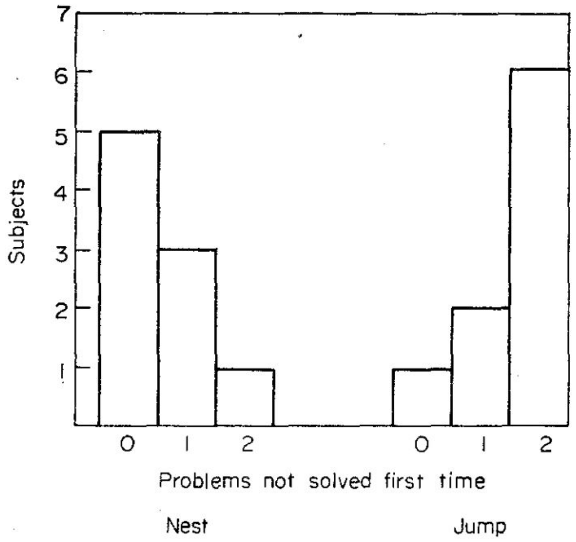
```
IF JUICY THEN
  IF LEAFY THEN
    IF GREEN THEN GRILL
    OTHERWISE BOIL
  OTHERWISE FRY
OTHERWISE
  IF HARD THEN ROAST
  OTHERWISE REJECT
```

vs

JUMP solution:

```
  IF JUICY GOTO L1
  IF HARD GOTO L2
  REJECT
L2 ROAST
L1 IF LEAFY GOTO L3
  FRY
L3 IF GREEN GOTO L4
  BOIL
L4 GRILL
```

=>



KLM/GOMS: Predicting time

Keystroke Level Model (KLM)

Model an interaction as series of operators, to predict the time an expert takes to do something

Operator	**Time/s**	**Description**
K	0.2	Key or button press
P	1.1	Pointing
H	0.4	Homing, switching hand between keyboard/mouse
M	1.35	Mental preparation
R	?	System response time

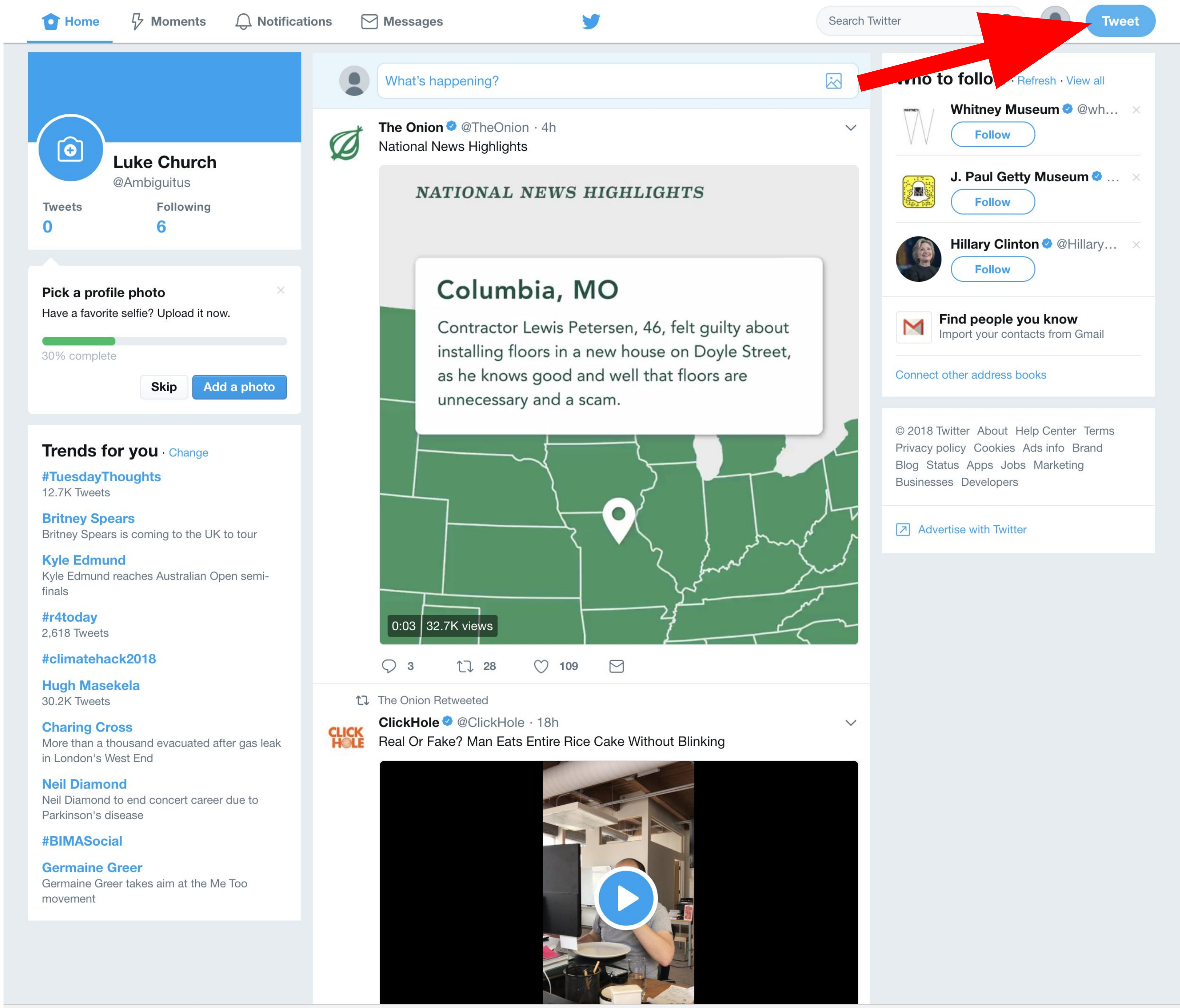
Keystroke Level Model (KLM)

Rules for when you should insert operators (NOT EXAMINABLE)

1. Insert Ms in front of Ks and Ps that select commands
2. Remove any Ms that are fully anticipated
3. Remove all by the first M from runs of MK that are a single cognitive unit
4. Remove any Ms where the K is a redundant terminator
5. Remove Ms from terminate constant strings

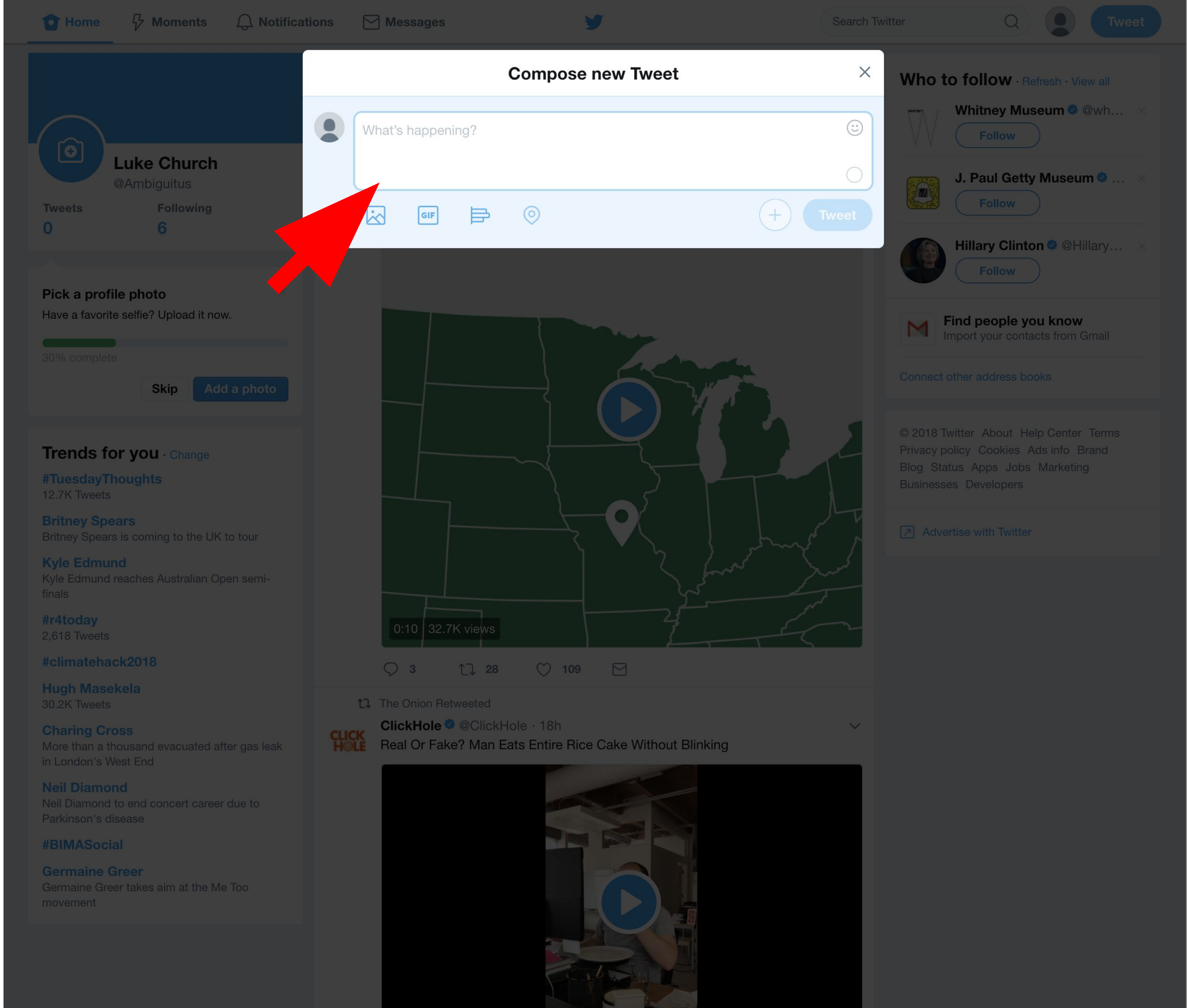
Example

M (before command)
H (hand -> mouse)
P (point at “Tweet”)
K (Click)
R (wait for response)



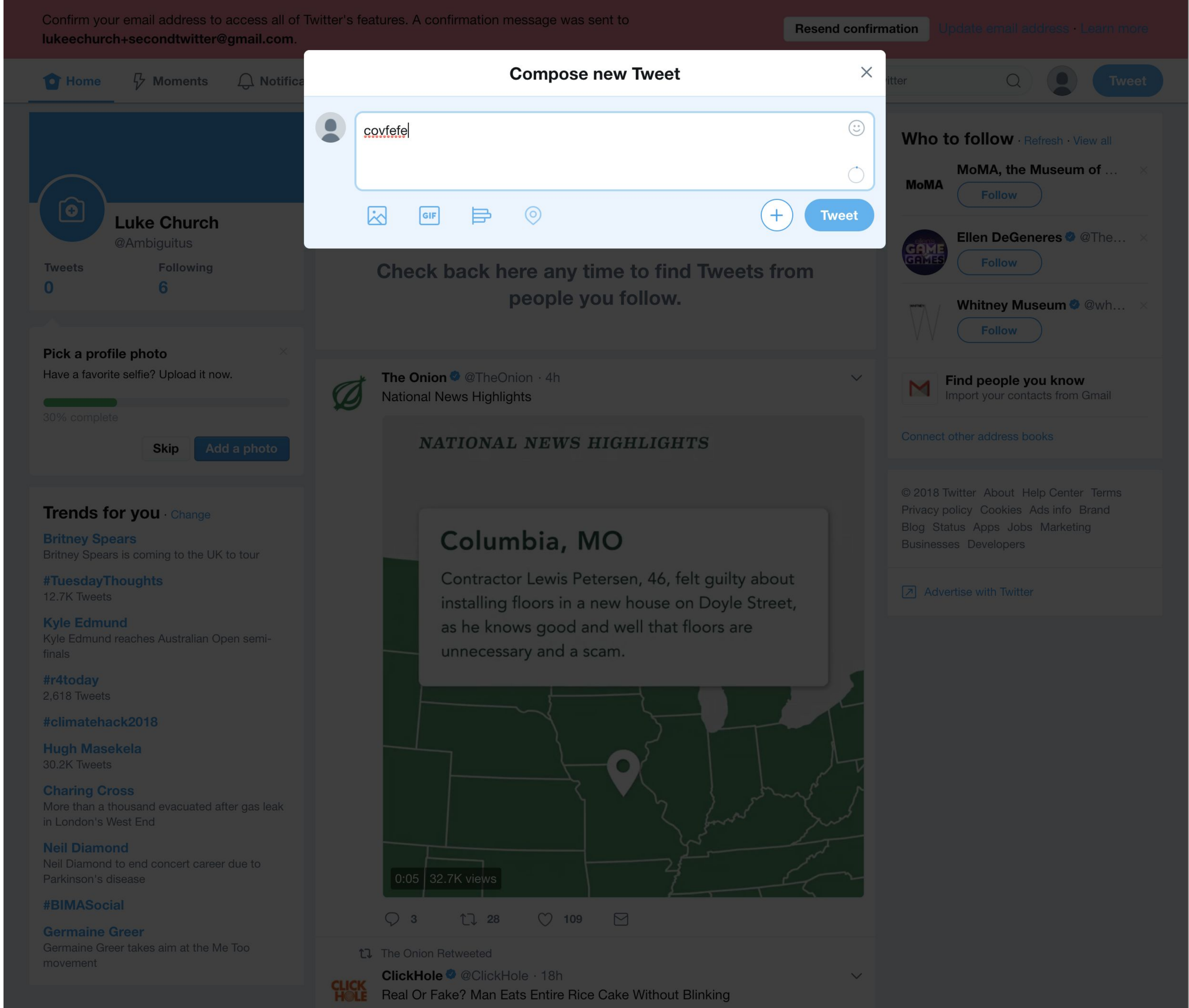
Example

M (before command)
P (point at "What's happening?")
K (Click)



Example

M (Prepare to type)
K
K
K
K
K
K
K
M (Prepare to click)
P (Point at “Tweet”)
K (Click)
R (Wait for response)

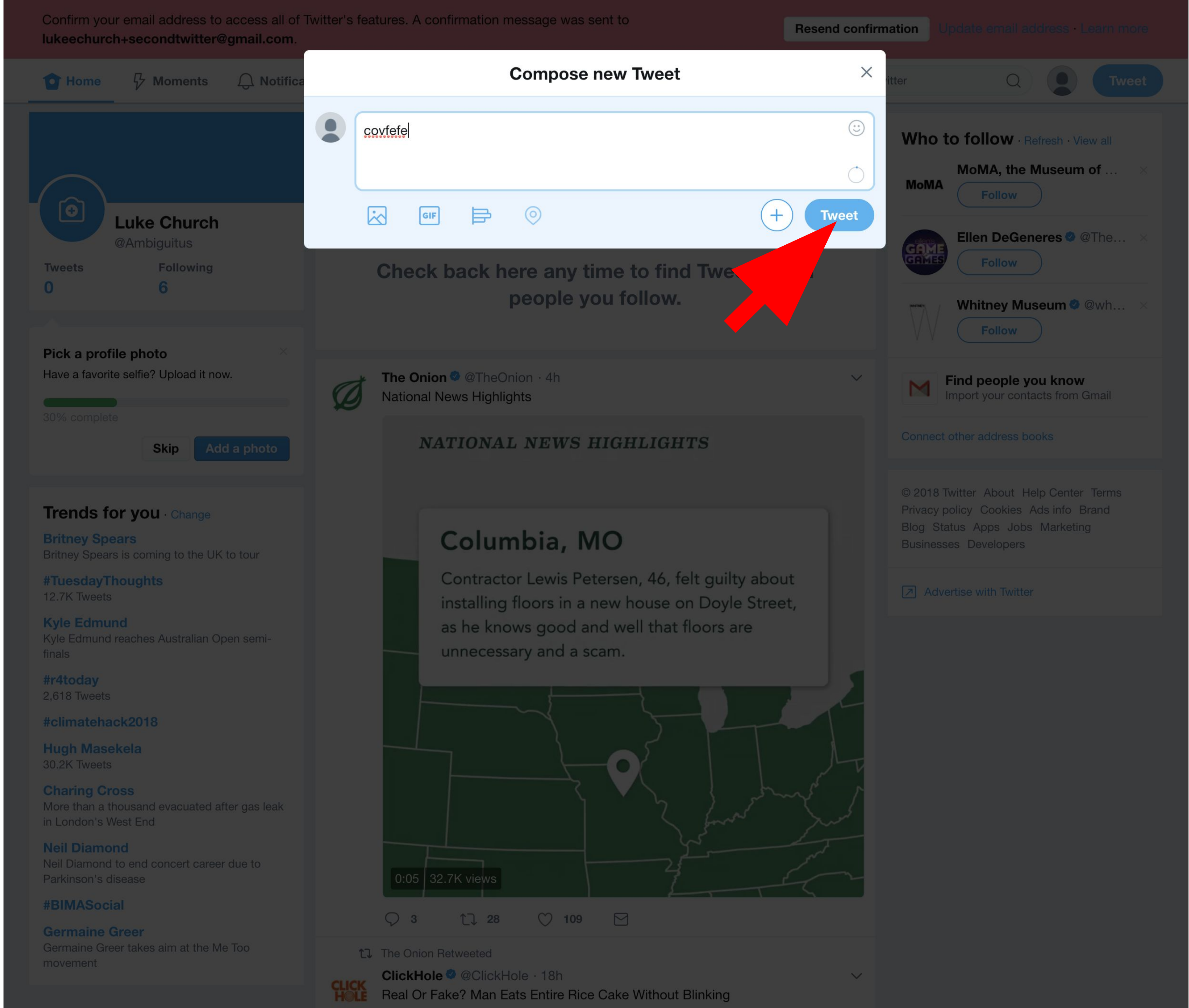


Example

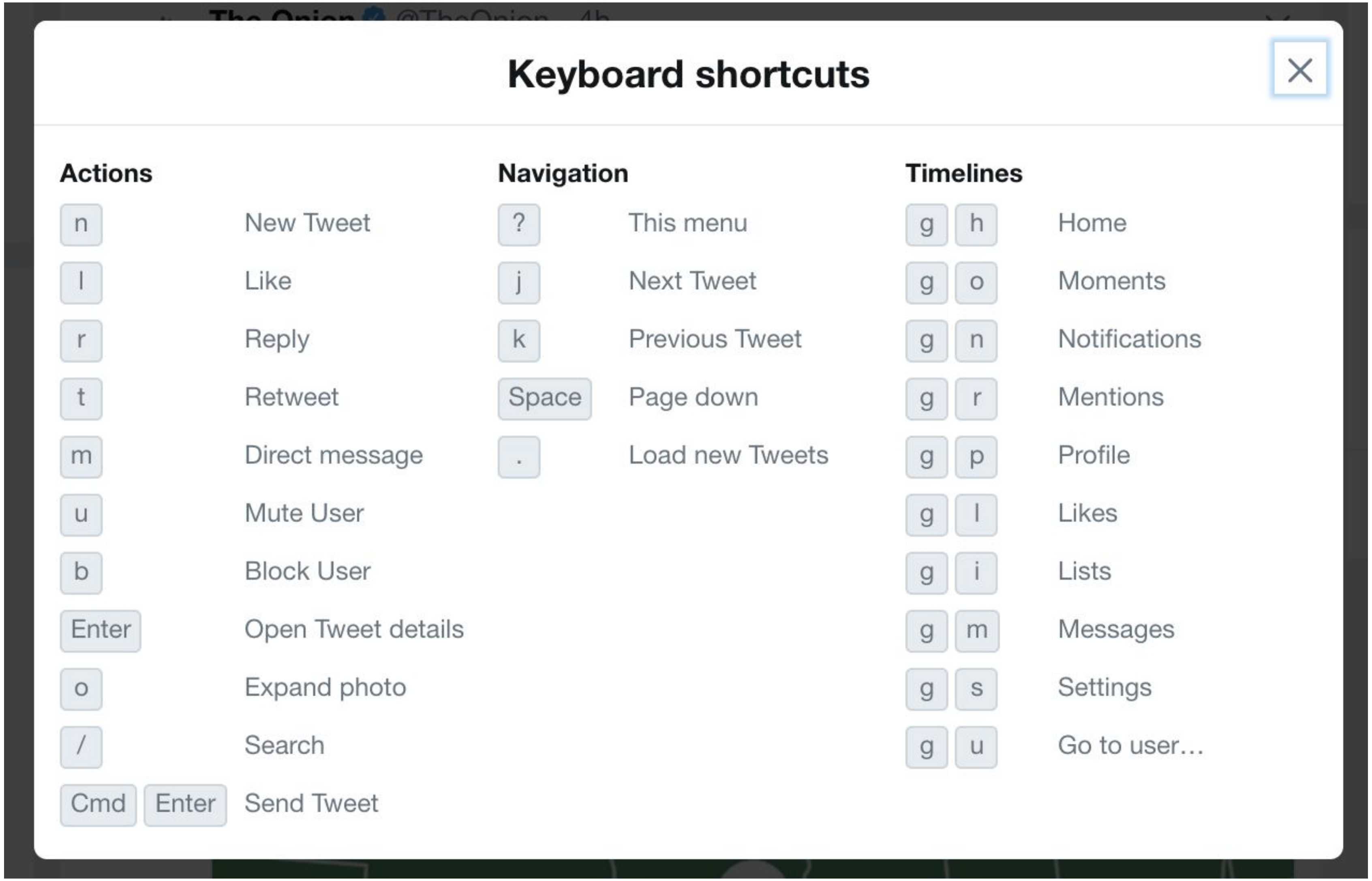
MHPKR
MPK
MKKKKKKKMPKR

1.35 + 0.4 + 1.1 + 0.2 + ~0.2
1.35 + 1.1 + 0.2
1.35 + 7*0.2 + 1.35 + 1.1 + 0.2 + ~0.2

= 11.5s

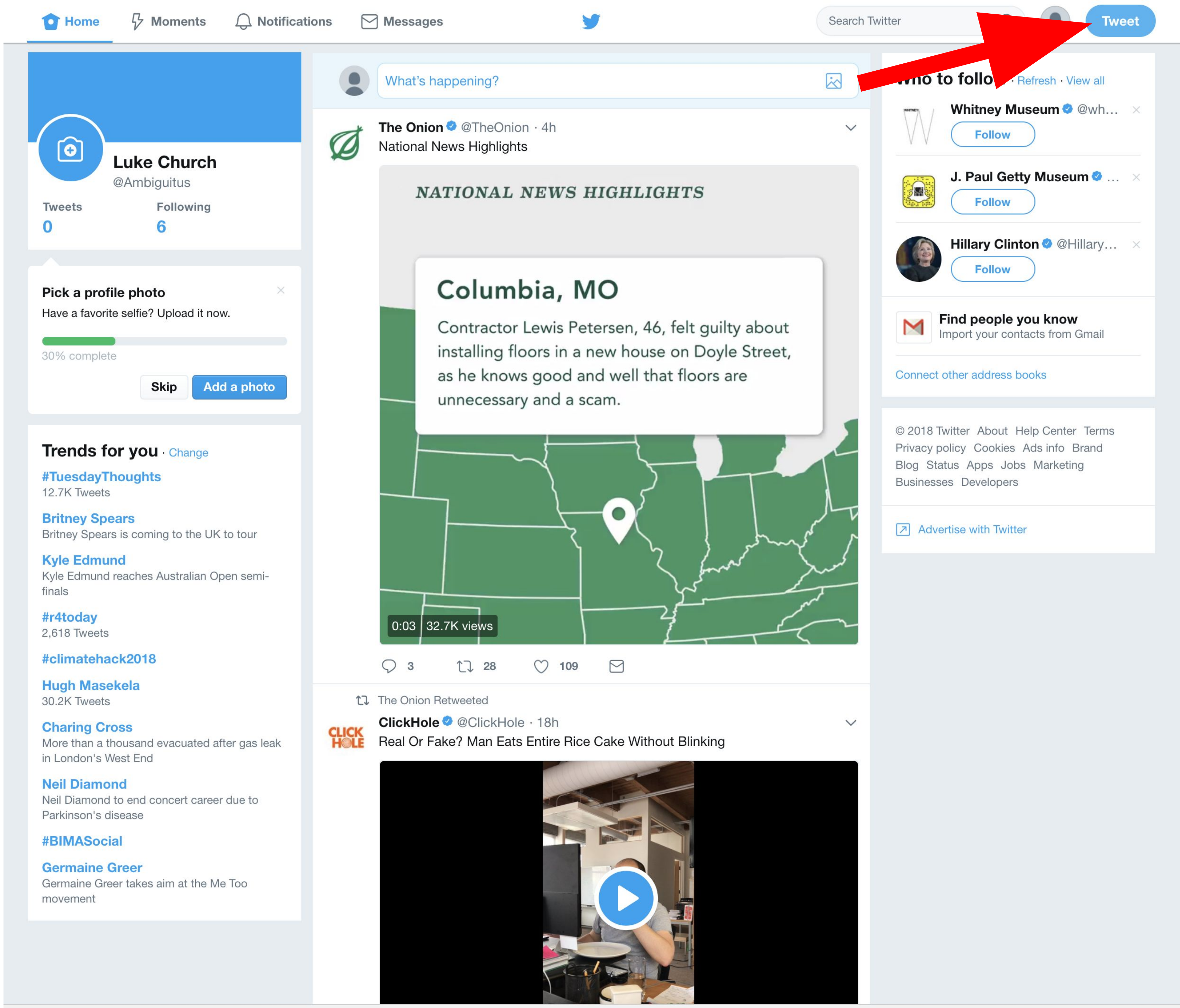


Keyboard shortcuts



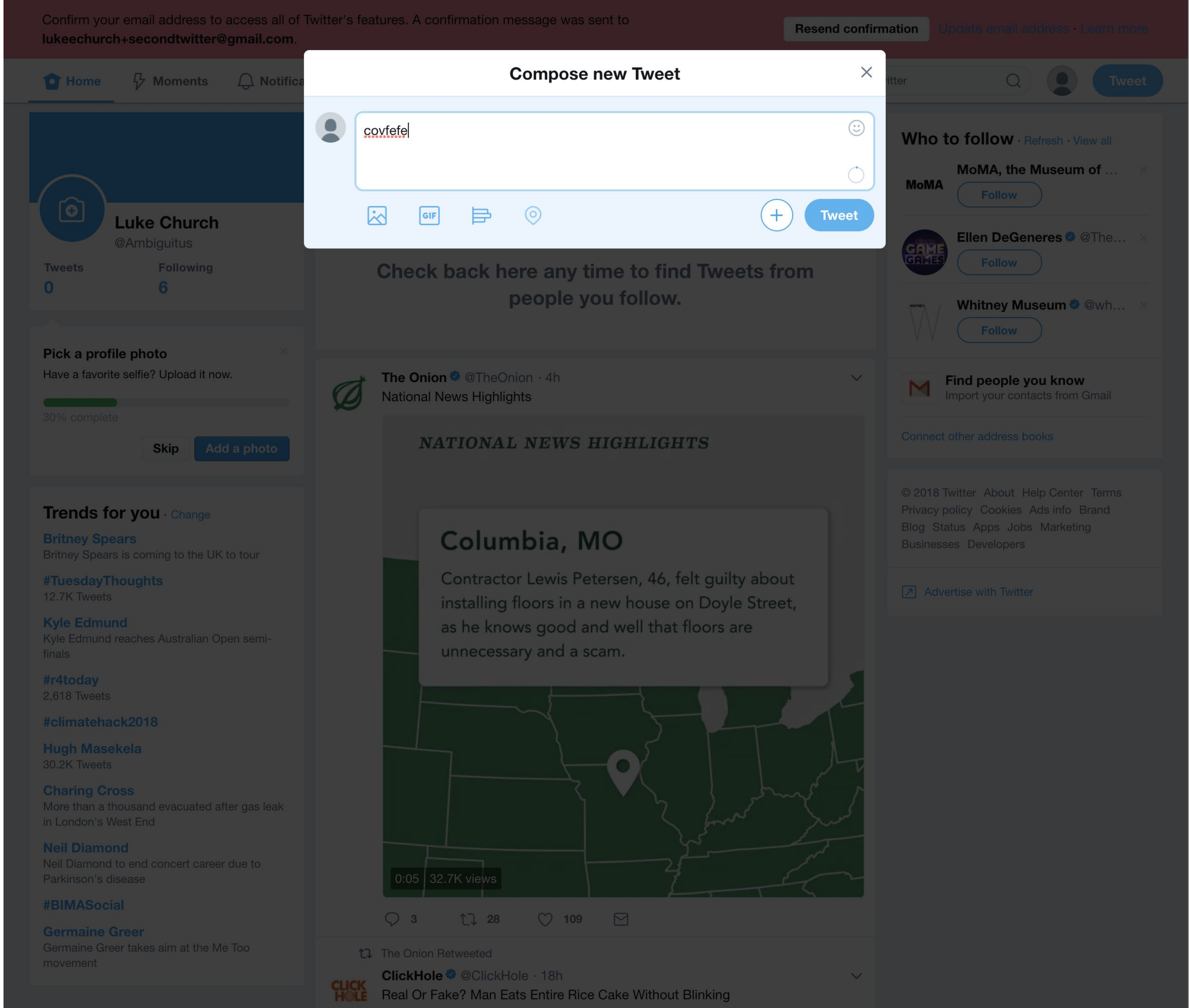
Example

M (become command)
K (‘n’)
R (wait for response)



Example

M (Prepare to type)
K
K
K
K
K
K
K
M (Prepare to click)
K (cmd)
K (enter)
R (Wait for response)



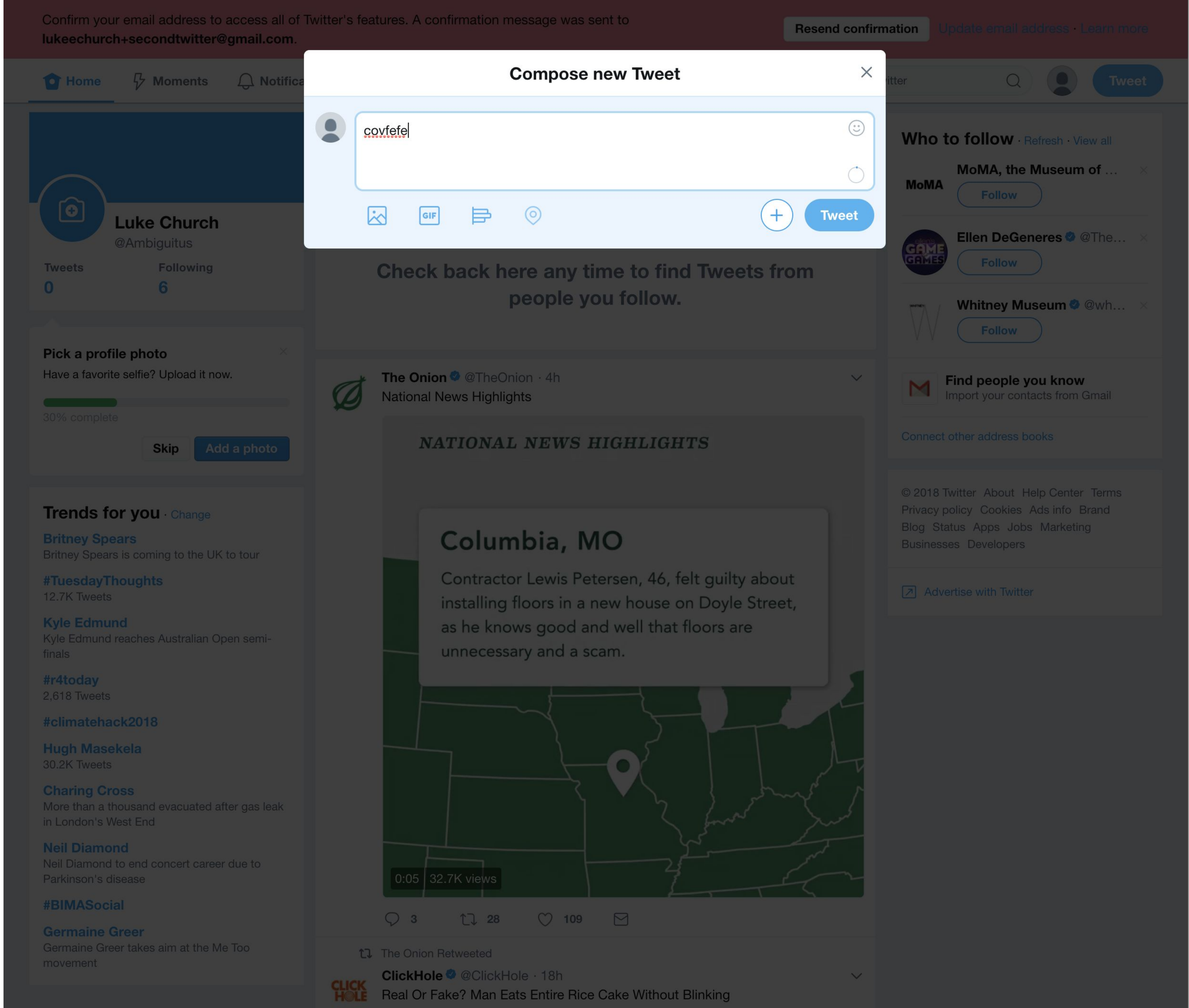
Example

MKR
MKKKKKKKMKKR

1.35 + 0.2 + ~0.2
1.35 + 7*0.2 + 1.35 + 0.2 + 0.2 + ~0.2

= 6.45s

(Compared to 11.5s before)



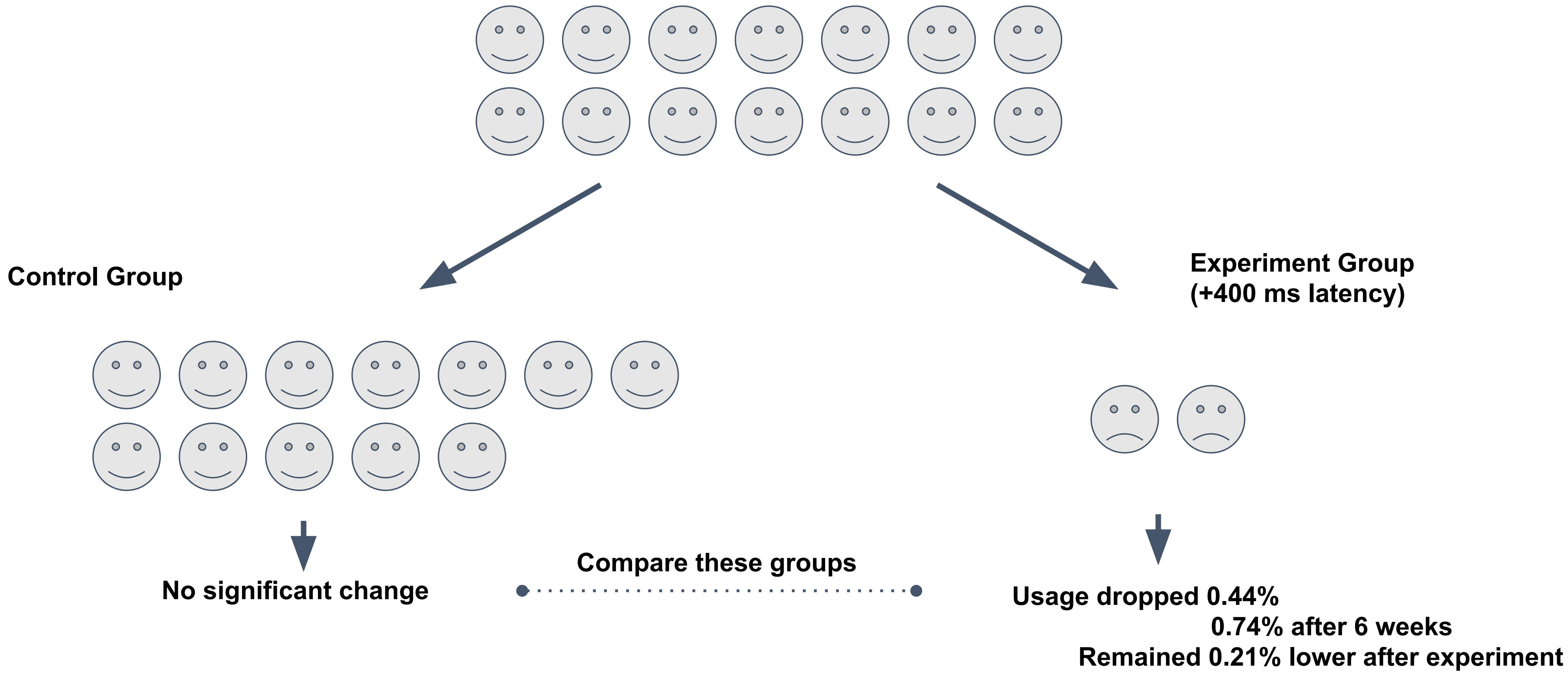
Experiments: Measuring time/usage

How many links should be on a search result page? (10, 20 or 30?)

- User studies: More is better
- When given 30, usage fell - why?
  - Analysis showed 400ms extra latency

Marissa Mayer, http://assets.en.oreilly.com/1/event/29/Keynote%20Presentation%202.pdf
https://www.youtube.com/watch?v=WFsQvcdmLxc&feature=channel

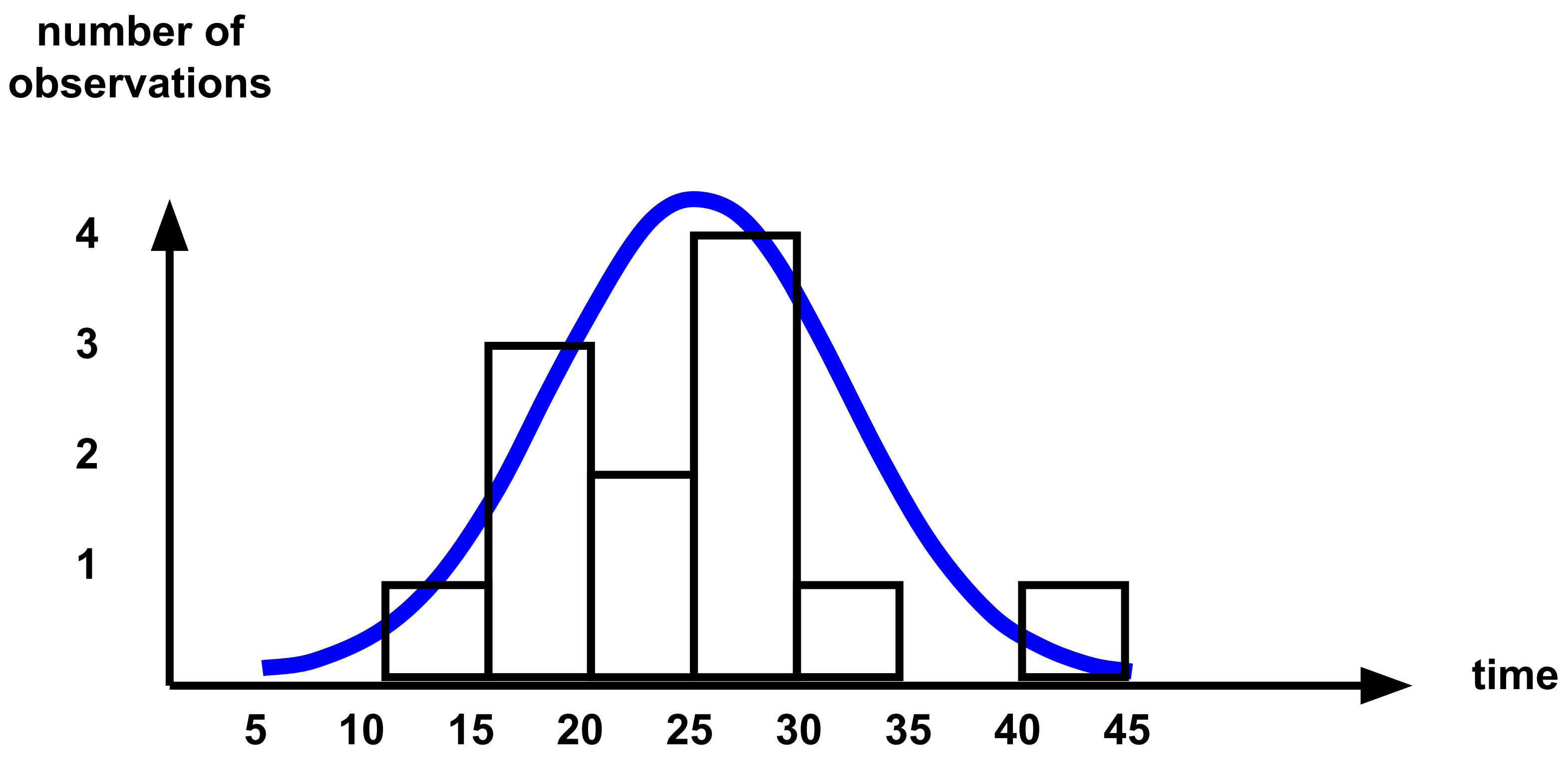
Latency experiment



Marissa Mayer, http://assets.en.oreilly.com/1/event/29/Keynote%20Presentation%202.pdf

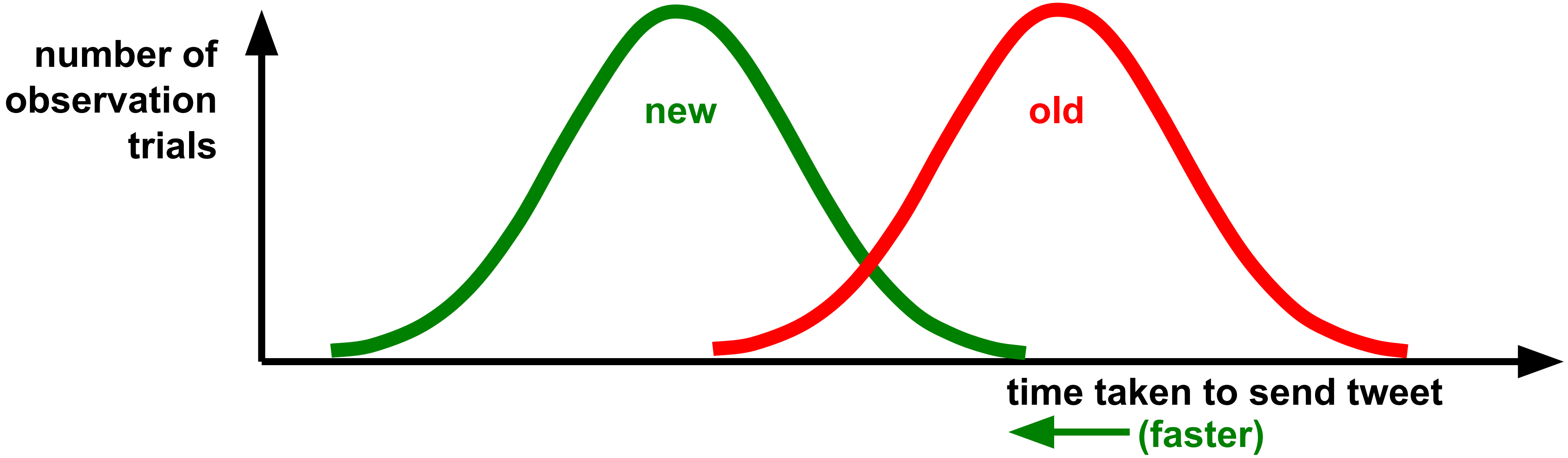
These are A/B experiments

(statistics: histograms & distributions)



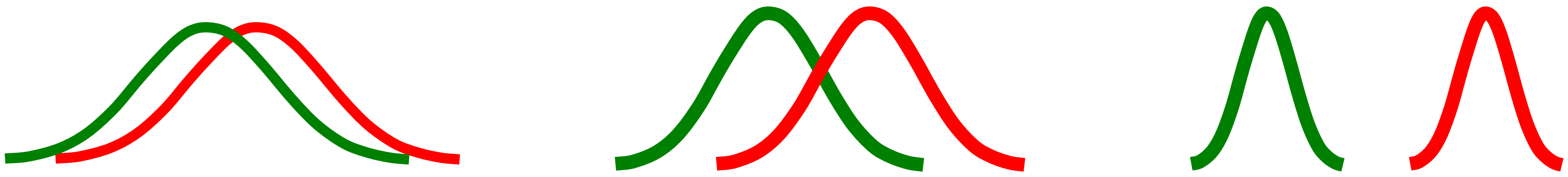
Experimental treatments

- A *treatment* is some modification that we expect to have an effect on usability:
  - How long does Donald take to send his tweet using this great new interface, compared to the crummy old one?
  - Expected answer: *usually* faster, but not *always*



Hypothesis testing

- *Null hypothesis:*
  - What is the probability that this amount of difference in means could be random variation between samples?
  - Hopefully very low ($p < 0.01$, or 1%)
  - Use a statistical *significance test*, such as the *t-test*.



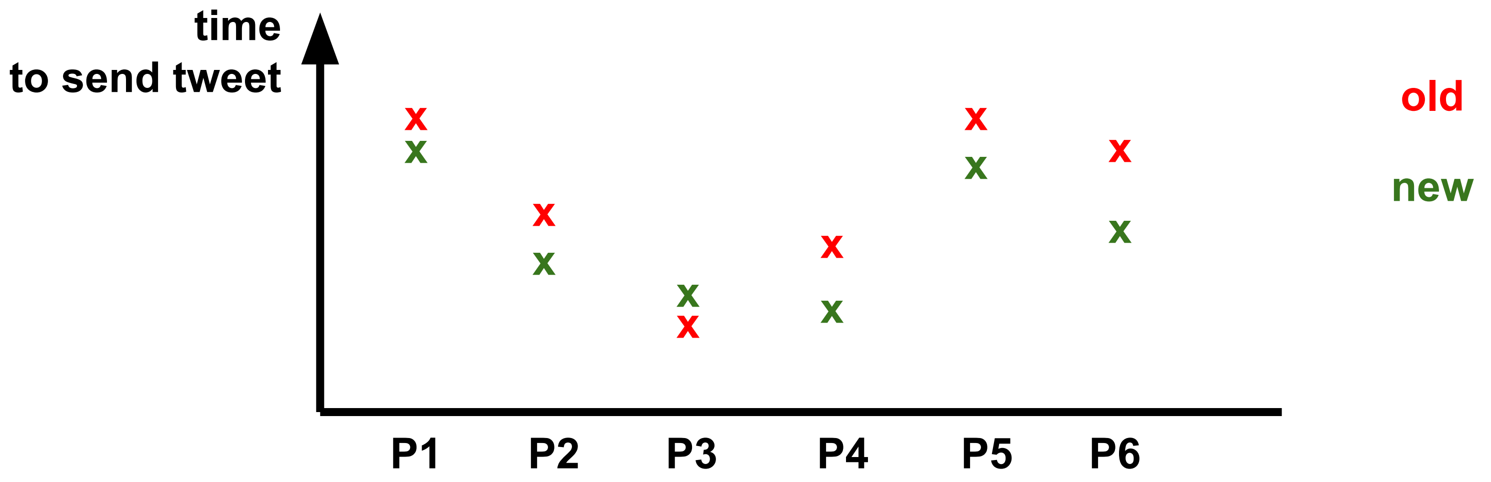
only random variation observed

observed effect probably does result from treatment

very significant effect of treatment

Sign tests

- In a within subjects experiment it’s possible to compare the results
  - Explores the [null] hypothesis that the median of the pairs is zero
  - Means might not be significant, but the sign can be
  - This is a non-parametric test, so doesn’t depend much on the data, but not very powerful (use a paired t-test, or Wilcoxon rank test instead)



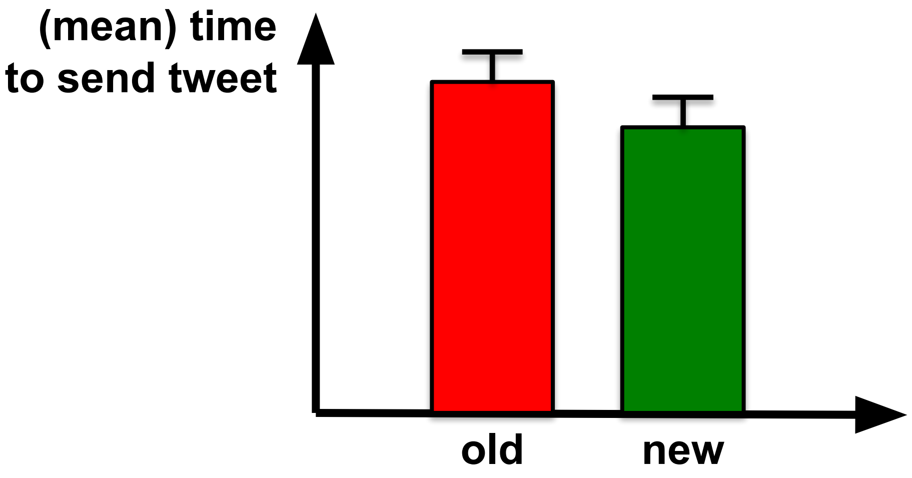
Experiment A: ‘significant’ but boring

Sources of variation

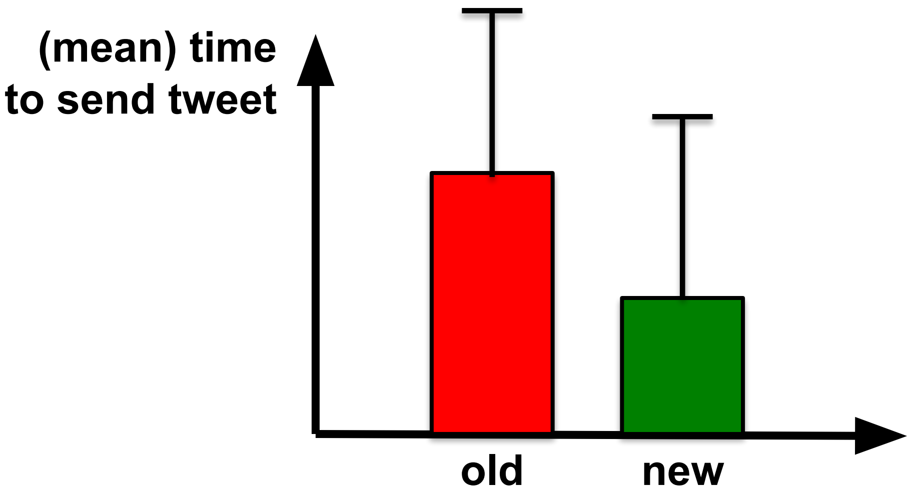
- People differ, so quantitative approaches to HCI must be statistical.
- We must distinguish sources of variation:
  - The effect of the treatment - what we want to measure.
  - Individual differences between subjects (e.g. IQ).
  - Distractions during the trial (e.g. sneezing).
  - Motivation of the subject (e.g. Mondays).
  - Accidental intervention by experimenter (e.g. hints).
  - Other random factors.
- Good experimental design and analysis isolates these.

Effect size – means and error bars

- Difference of two means may be statistically significant (if sample has low variance), without being very interesting.
  - But mean differences must ***always*** be reported with a confidence interval, or plotted with 'error bars'



Experiment A: 'significant' but boring



Experiment B: interesting, but treat with caution

Problems with controlled experiments

- Huge variation between people (~200%)
- Mistakes mean huge variation in accuracy (~1000%)
- Improvements are often small (~20%)
- … or even negative (because new & unfamiliar)
- … and may result from something unrelated to your design!

The Hawthorne Effect

- Studies on productivity in 1924-1932
  - Do lighting levels affect productivity?
  - Studies appeared to show improvements in both directions
  - Results show the motivational effect of being studied, not of the change

Is efficiency always a design goal?

- What if you wanted to encourage thoughtfulness? Creativity?

Taylorism

- F.W. Taylor (1856-1915)
  - Engineer who invented scientific management
  - Measure workers as if parts in a machine
  - Optimise by measurement and correction
- Not so popular with trade unions!
  - Note that 2nd wave HCI (the turn from human factors to social science) involved working closely with trade unions, especially in Sweden and Denmark

Discretionary use systems

If you are not working to someone else's goal, you can decide whether or not to be efficient (or whether you want to use the system at all)

Simone Giertz: "Queen of Shitty Robots"

Efficient creativity?

- What if there isn’t a good measure of productivity?
  - Maximise output of poetry-lines?
  - Maximise musical notes played per second?
  - Maximise Cambridge graduates per year?
- Optimum User Experience
  - What if you wanted people to enjoy what they did?