COGS 105

Research Methods for Cognitive Scientists

Week 3, Class 2:
Behavioral Methods I: Reliability and Validity

Last Class

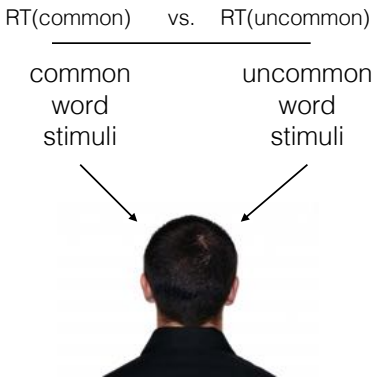
- In any behavioral research we need to **design measures, develop tasks, and recruit people** to participate in them.
- Lots of sampling methods; usually we are stuck with **nonprobability "haphazard" sampling**, and we often assume that our recruitment (e.g., SONA) is "effectively random."

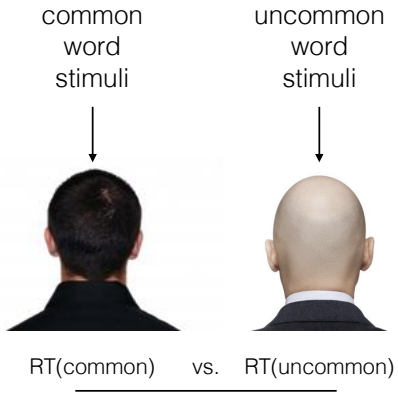
Our LDT Task

- We started with a simple Lexical Decision Task: **Are you faster at processing uncommon or common words?**
- General thrust of the result: **common words** ("higher frequency") are faster to process than **uncommon words** ("lower frequency")

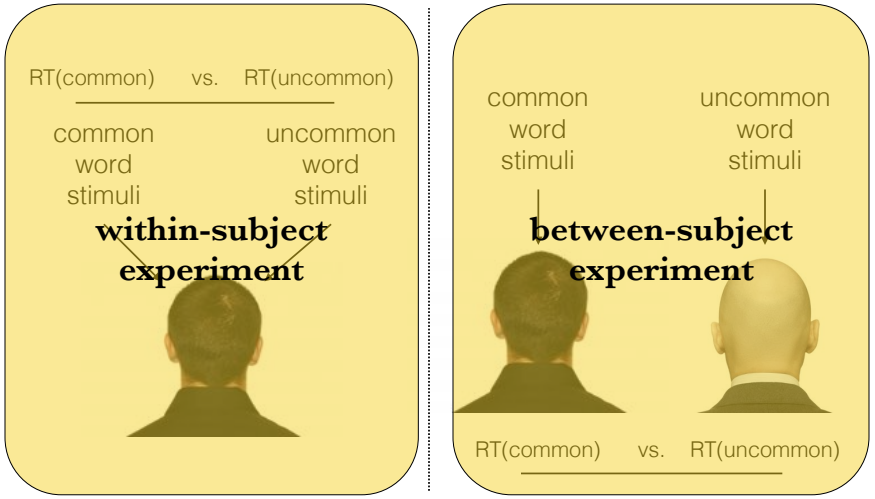
Word	Ranking
The	1st[3]
At	20th
So	50th
Did	70th
Got	100th
Mind	300th
Chaos	5,000th
Falkland	20,000th
Marche	45,000th
Tisane	85,000th

Between vs. Within





Between vs. Within



Pervasive Sampling Issues

- We **sample subjects**, we **sample words** as stimuli, and for each participant in our task we have to **sample the stimuli** we chose for presentation in a given order.
- All of these **can involve biases**.
  - **Participant biases**: e.g., WEIRD
  - **Stimulus biases**: e.g., you choose words that are not perfectly comparable only in the variable of interest (commonality, aka frequency)
  - **Presentation biases**: you order the words in a way that influences responses.

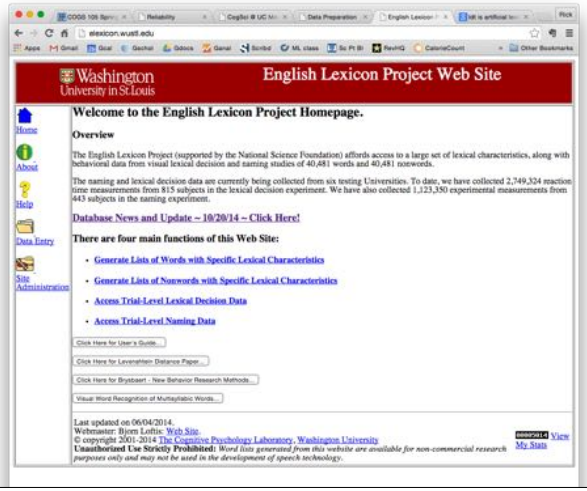
E.g., Stimulus Biases

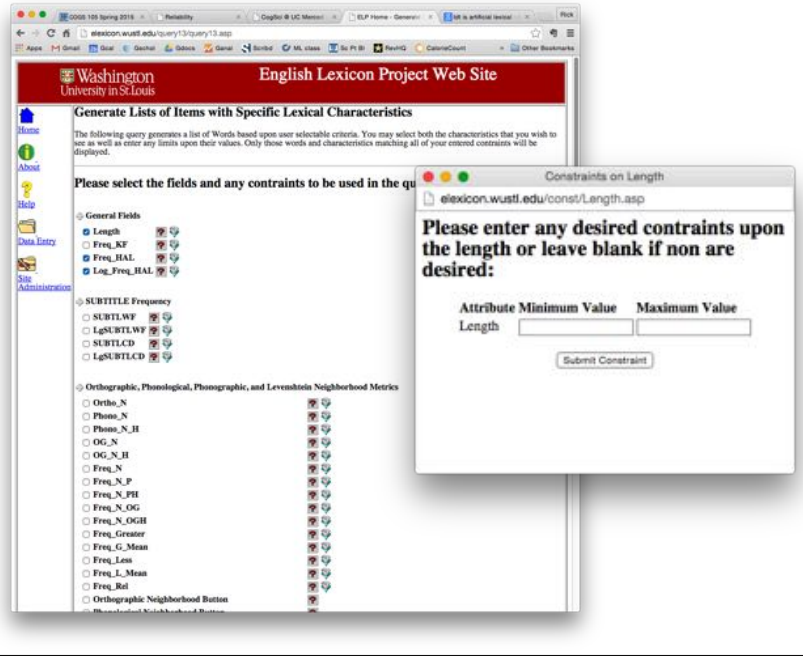
- If we want to compare common vs. uncommon words, we need to **isolate this one difference**, and our **target stimuli** (common vs. uncommon) should be:
  - Overall matched for length
  - Overall matched for pronounceability
  - Overall matched for concreteness in meaning
  - Etc.
- Such extensive controls are difficult to achieve but possible with some available tools.

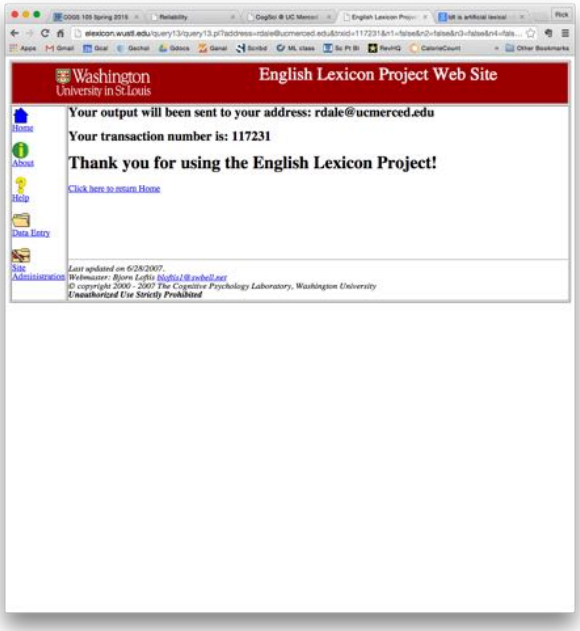
Example Tool

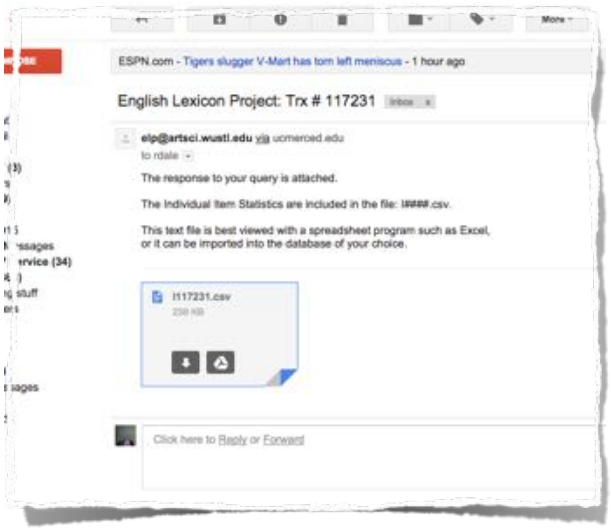
- **English Lexicon Project**!
- Large-scale project helping you select stimuli for your word experiments (used often for LDT).
  - Can help you avoid certain “stimulus biases,” to make sure words are differ only on one dimension.
- Completely free to use; you can use it next week for your lab!
  - http://elexicon.wustl.edu/

E.g., Control for Length

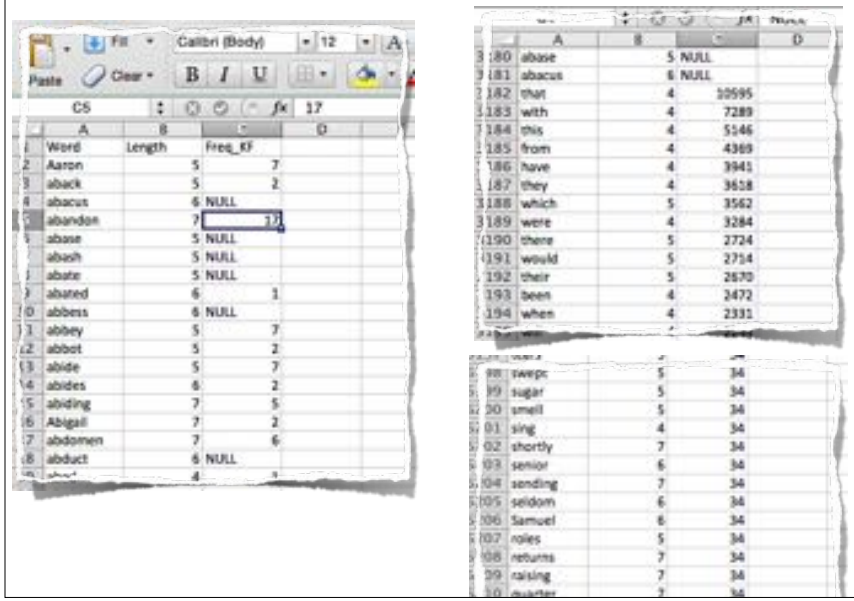








sort the data that is emailed to you



Another Bias: Order

Respond with your dominant hand if you see a real word.

fludl ← "no-go trial"

made

pragrl

walking

suggest

fort

uncommon, but faster because they occur in order?

order bias!

Construct Validity

- Now that you have the task in mind… consider… **construct validity**.
- We wish to make an **inference about how people process words**.
- Thus LDT is a method (an **operationalization**) of mental processing that is supposed to tell us something about a **construct: word processing**.
- You typically **cannot directly observe the construct**; your operationalization (your measures) help you make inferences about it.

Validity of What?

method

LDT

the method is a proposed operationalization of construct

construct

language processing

theoretical concept or proposal under study

Strategic Workforce Planning

As you are probably aware, the campus is engaged in a workforce planning exercise that seeks to identify the most critical administrative and staff positions the campus will need as it grows to 10,000 students. This exercise is also an opportunity for the campus to create efficiencies that allow for new investments, drive innovation and invent new and more effective ways of working.

Some of you have asked why we are doing this now given all of the other initiatives the campus has undertaken. The answer is simple: We can't afford not to.

We know the current staff-to-faculty ratio at UC Merced is higher than other University of California campuses, and this is typical of a new campus in its early stages. Even so, we are weak in some staffing areas, including support for our faculty and schools. Given we will not have the fiscal resources to continue to grow staff in the future at the same rate we have enjoyed in the past, it is
ns reflect our most critical needs and also
l administrative operations.

Kinds of Validity

- "In **face validity**, you look at the operationalization and see whether "on its face" it seems like a good translation of the construct."

LDT
carefully choose a bunch of words
show 'em one at a time
separated by carefully controlled time intervals
in a quiet room in front of a computer
and you're asked to "just recognize them"

also: ***ecological validity***

Kinds of Validity

- In **predictive validity**, "we assess the operationalization's ability to predict something it should theoretically be able to predict."
  - E.g., can LDT be used to measure other aspects of language processing? For example, can it demonstrate that positive vs. negative words are processed differently? Can it show that longer words and processed more slowly than shorter words? Etc.

Kinds of Validity

- "In **convergent validity**, we examine the degree to which the operationalization is similar to (converges on) other operationalizations that it theoretically should be similar to."
  - Eye movements while reading?
  - Naming times? Rather than responding to word/ nonword, respond by speaking the sequence of letters (common words also faster!).
  - LDT should "converge" with these tasks.

Word-Naming Task

- WNT is a variant of LDT that is often used for similar purposes. Let's give it a try. Just speak these words as you see them as quickly, but naturally, as you can.

symbol plenty other also

We expect WNT to have "convergent validity" with results in LDT.

Unreliable & Unvalid

Unreliable, But Valid

Reliable, Not Valid

Both Reliable & Valid

reliability

=

"consistency"

Reliable, Not Valid

Both Reliable & Valid

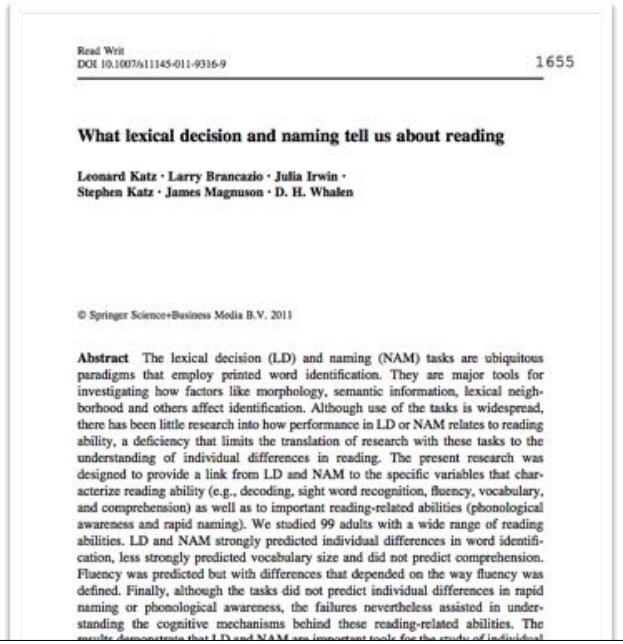
Unreliable, But Valid

Both Reliable & Valid

validity

=

"accuracy"



Read Writ
DOI 10.1007/s11145-011-9316-9 1655

What lexical decision and naming tell us about reading

Leonard Katz · Larry Brancazio · Julia Irwin · Stephen Katz · James Magnuson · D. H. Whalen

© Springer Science+Business Media B.V. 2011

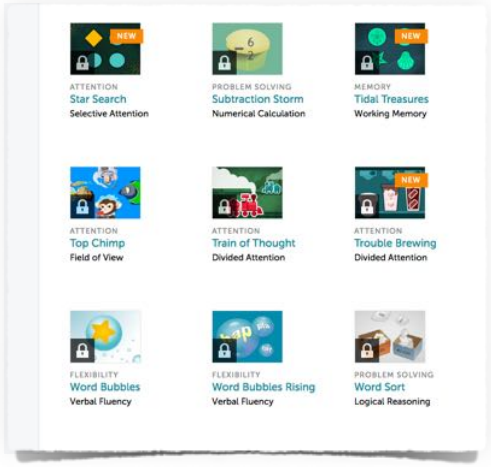
Abstract The lexical decision (LD) and naming (NAM) tasks are ubiquitous paradigms that employ printed word identification. They are major tools for investigating how factors like morphology, semantic information, lexical neighborhood and others affect identification. Although use of the tasks is widespread, there has been little research into how performance in LD or NAM relates to reading ability, a deficiency that limits the translation of research with these tasks to the understanding of individual differences in reading. The present research was designed to provide a link from LD and NAM to the specific variables that characterize reading ability (e.g., decoding, sight word recognition, fluency, vocabulary, and comprehension) as well as to important reading-related abilities (phonological awareness and rapid naming). We studied 99 adults with a wide range of reading abilities. LD and NAM strongly predicted individual differences in word identification, less strongly predicted vocabulary size and did not predict comprehension. Fluency was predicted but with differences that depended on the way fluency was defined. Finally, although the tasks did not predict individual differences in rapid naming or phonological awareness, the failures nevertheless assisted in understanding the cognitive mechanisms behind these reading-related abilities. The

LDT predicts word recognition, vocabulary size, and fluency to some degree

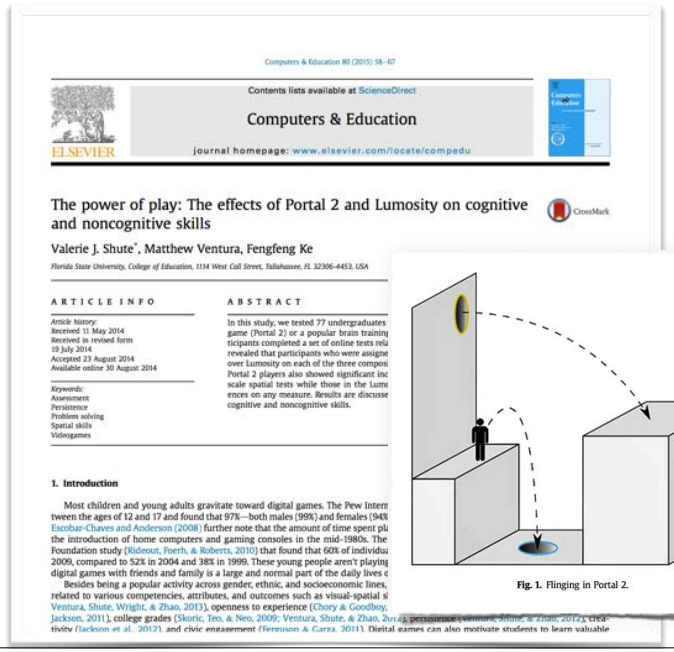
Why RT / LDT?

- These kinds of measures are very simple, and seemingly artificial, however they have massive and **broad applicability**!
- Two case studies:
  - 1) **Lumosity**
  - 2) The **IAT** (as in lab)

Lumosity



Online brain-training system that uses basic cognitive task operationalization



Computers & Education 80 (2015) 58–67

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

The power of play: The effects of Portal 2 and Lumosity on cognitive and noncognitive skills

Valerie J. Shute*, Matthew Ventura, Fengfeng Ke

Florida State University, College of Education, 1114 West Call Street, Tallahassee, FL 32306-4453, USA

ARTICLE INFO

Article history:
Received 11 May 2014
Received in revised form
10 July 2014
Accepted 23 August 2014
Available online 30 August 2014

Keywords:
Assessment
Persistence
Problem solving
Spatial skills
Videogames

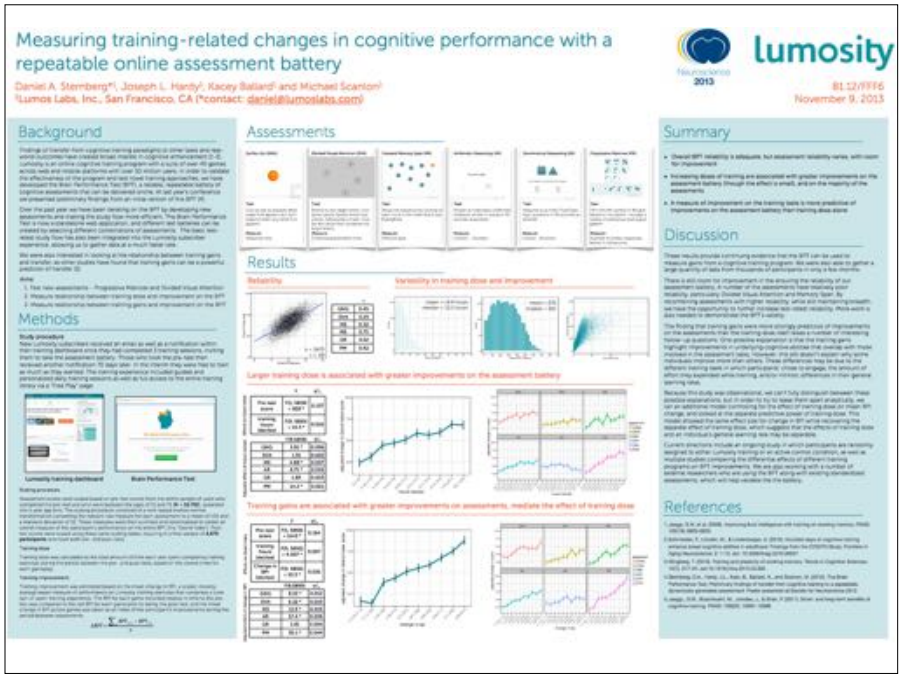
ABSTRACT

In this study, we tested 77 undergraduates
game (Portal 2) or a popular brain training
ticipants completed a set of online tests rela
revealed that participants who were assigne
over Lumosity on each of the three compos
Portal 2 players also showed significant inc
scale spatial tests while those in the Lum
ences on any measure. Results are discusse
cognitive and noncognitive skills.

1. Introduction

Most children and young adults gravitate toward digital games. The Pew Intern
tween the ages of 12 and 17 and found that 97%—both males (99%) and females (94%
Escobar-Chaves and Anderson (2008) further note that the amount of time spent pla
the introduction of home computers and gaming consoles in the mid-1980s. The
Foundation study (Rideout, Foerh, & Roberts, 2010) that found that 60% of individua
2009, compared to 52% in 2004 and 38% in 1999. These young people aren't playing
digital games with friends and family is a large and normal part of the daily lives o
Besides being a popular activity across gender, ethnic, and socioeconomic lines,
related to various competencies, attributes, and outcomes such as visual-spatial s
Ventura, Shute, Wright, & Zhao, 2013), openness to experience (Chory & Goodboy,
Jackson, 2011), college grades (Skoric, Teo, & Neo, 2009; Ventura, Shute, & Zhao, 2012), persistence (Ventura, Shute, & Zhao, 2012), civic
tivity (Jackson et al., 2012), and civic engagement (Ferguson & Garza, 2011). Digital games can also motivate students to learn valuable

Fig. 1. Flinging in Portal 2.



IAT

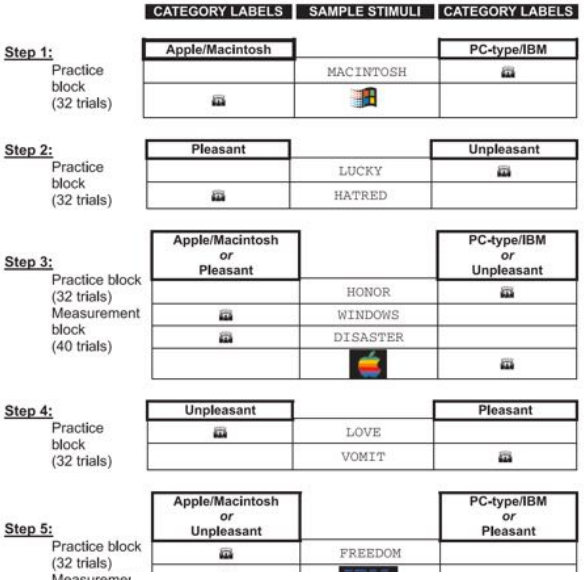
- "Implicit Association Task": Uses basic RT to tap into potential biases or stereotypes you might have.

Demo...

Left hand: Good

Right hand: Bad

smelly
stupid
delicious
friendly
evil
pleasant
Affordable Care Act



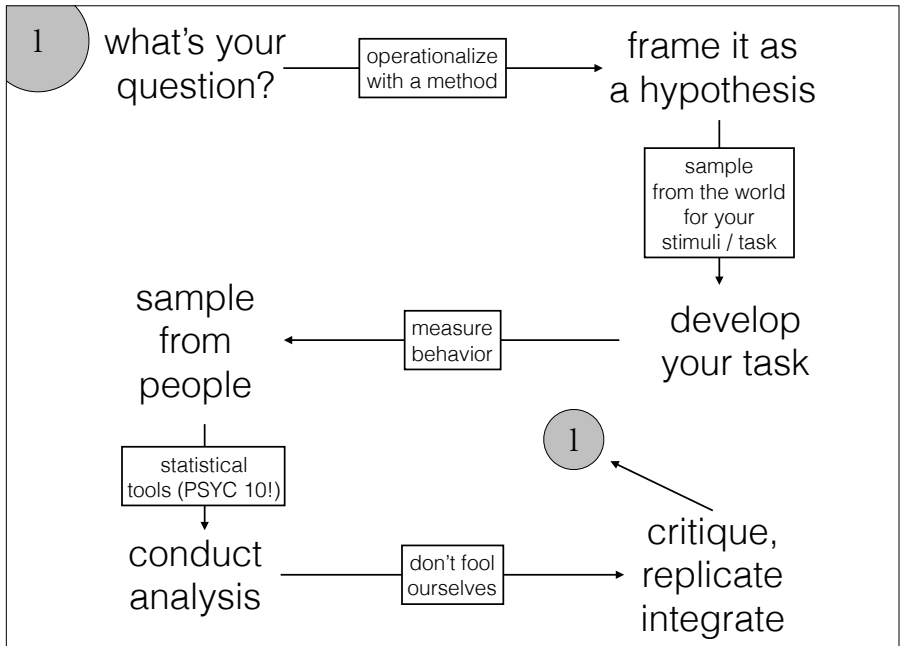
Example use of IAT in business / marketing

"First, explicit measures and IAT measures of attitudes and other marketing constructs converge when consumers are willing and able to report their feelings and beliefs."

Brunel, F. F., Tietje, B. C., & Greenwald, A. G. (2004). Is the implicit association test a valid and valuable measure of implicit consumer social cognition?. Marketing, 4.

Construct?

- **Construct**: Political affiliation, or race?
- **Operationalization**: reaction time (RT) to responses that are mapped onto the same hand.
- **Construct validity**:
  - Face validity?
  - Predictive validity?
  - Convergent validity?



"These subtle distinctions, about sampling, validity, reliability, and so on… really it is becoming clear that the only way to really understand these things is to get in there and do studies…"

Next class…

- Let's move into some methodological specifics: Details of using reaction time.
- Lab: You will build your own reaction-time experiment.
  - You can build your own creative experimental idea using the overall process just described.