

# Massive Online Experiments

Joshua K. Hartshorne  
Boston College



# Cognitive Science Bottleneck



## **Brick-and-Mortar labs**

### **Pros**

- Control over environment
- Specialized equipment
- 100 yrs institutional knowledge

### **Cons**

- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals

# Cognitive Science Bottleneck



## Brick-and-Mortar labs

### Pros

- ~~Control over environment~~
- Specialized equipment
- 100 yrs institutional knowledge

### Cons

- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals

# Cognitive Science Bottleneck



## Brick-and-Mortar labs

### Pros

- ~~Control over environment~~
- ~~Specialized equipment~~
- 100 yrs institutional knowledge

### Cons

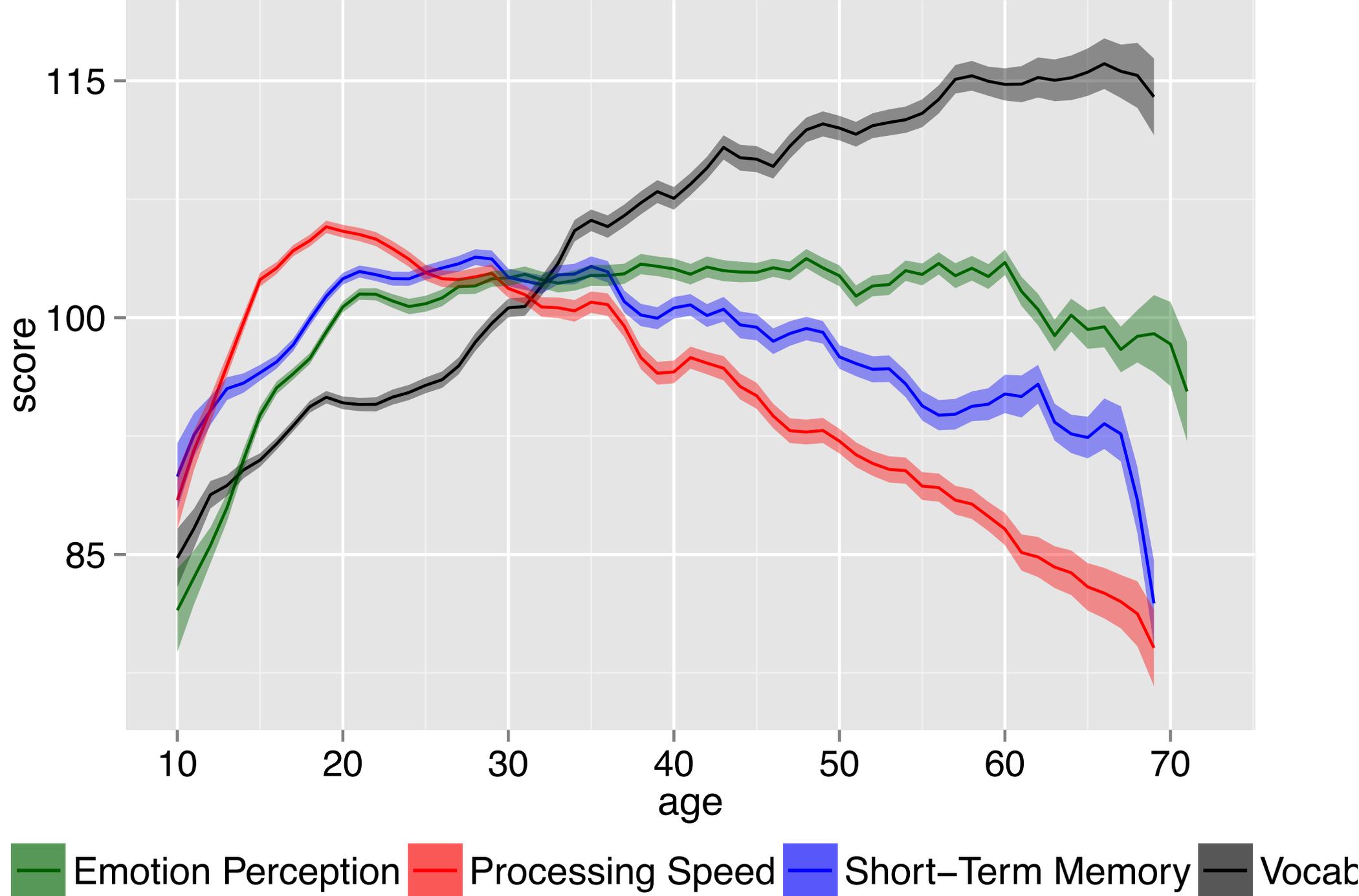
- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals

## The Dream: Massive Online Experiments

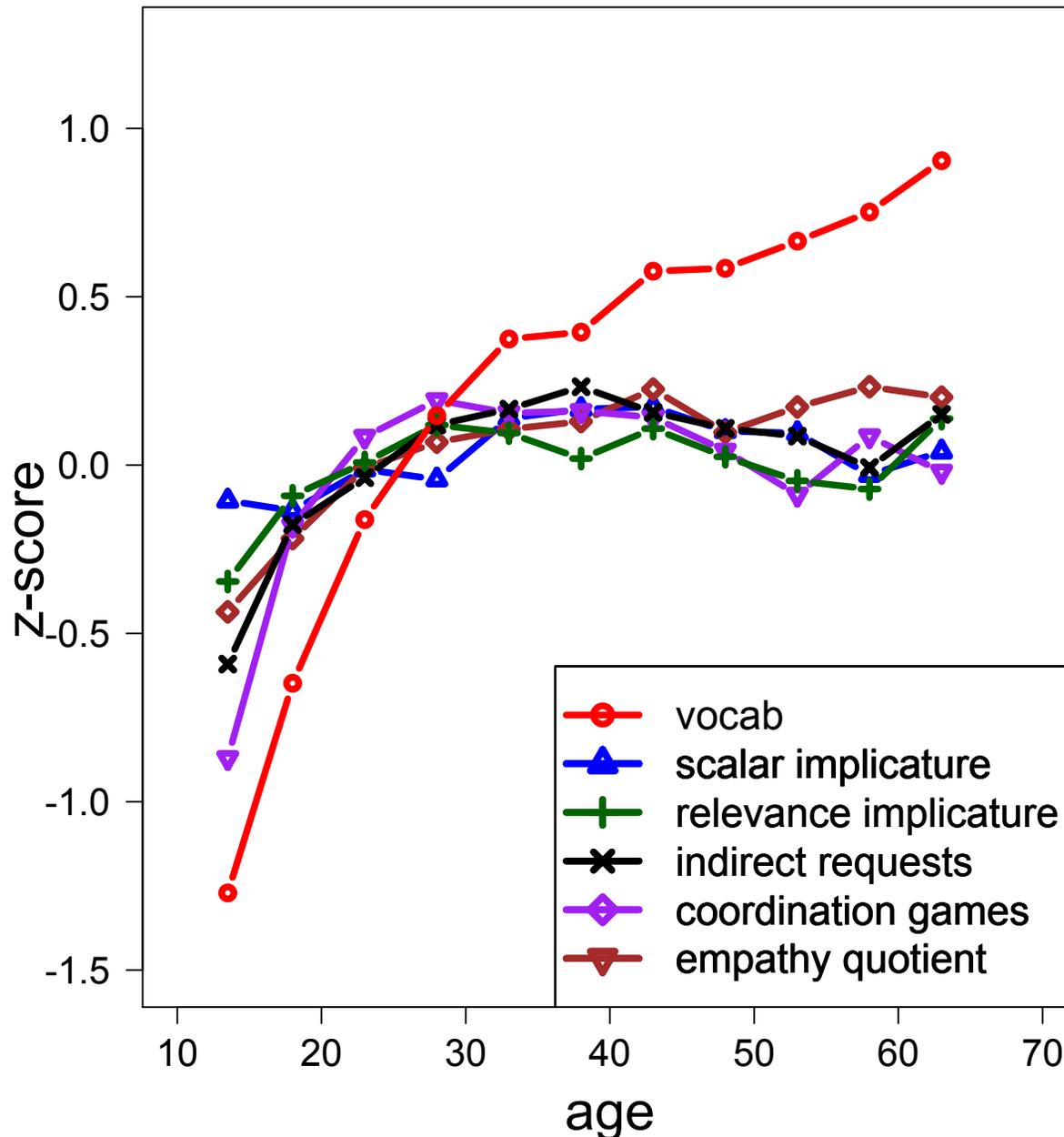
- Many subjects, world-wide
- Many items
- Many contexts/manipulations
- Longitudinal
- Social networks
- “A whole literature in one experiment”



# Lifespan Development



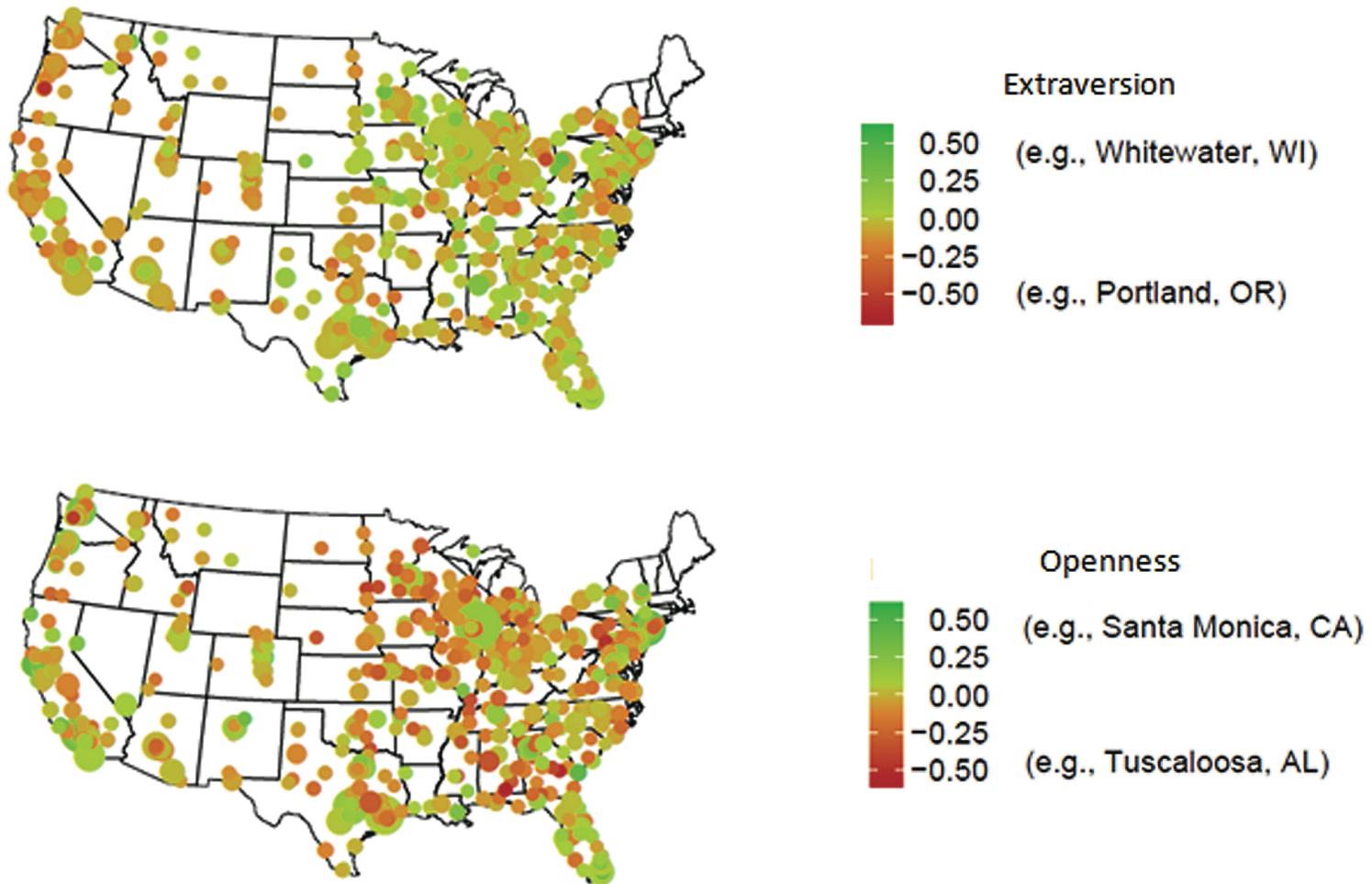
# Lifespan Development



(3,224 subjects)

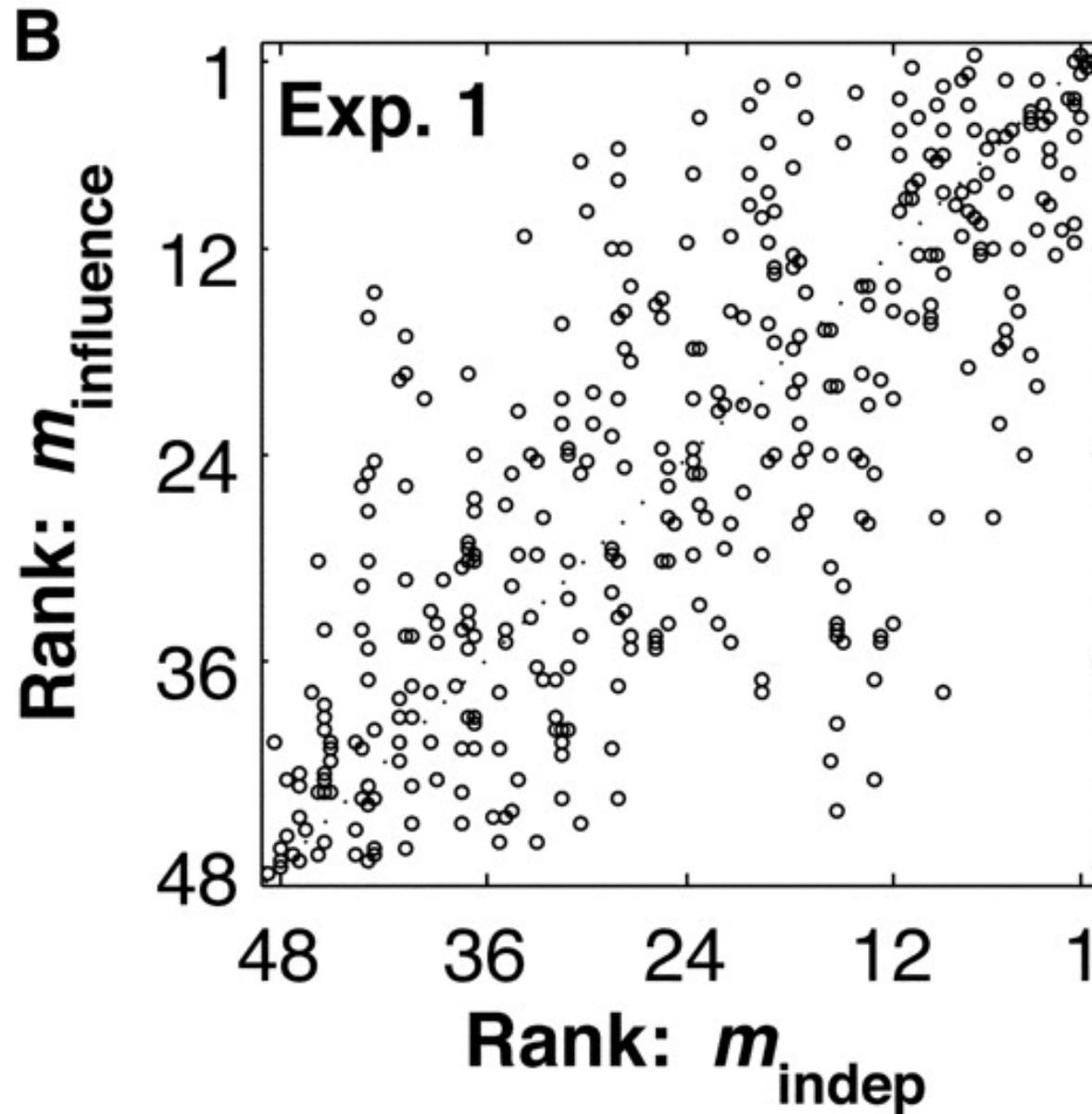
(Hartshorne & Pinker, in prep)

# Geography of Personality



(Bleidorn et al., 2016)

# Music popularity vs. quality



(Salganik, Dodds, & Watts, 2006)

# Demographics & Visual Preferences

Preferred by under-20s

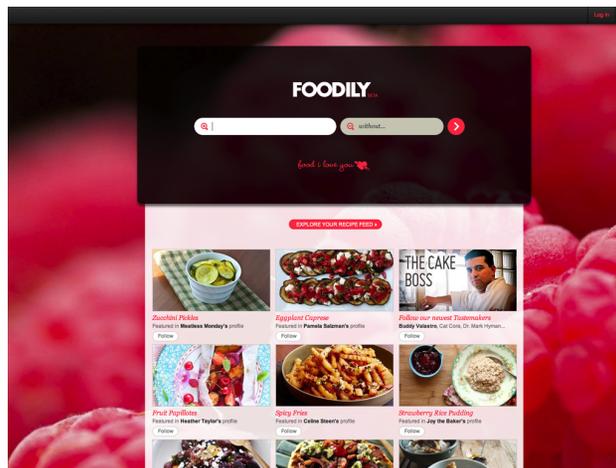


(a) under 20=7.2, over 51=5.4

Preferred by over-51s



(b) under 20=3.0, over 51=4.8

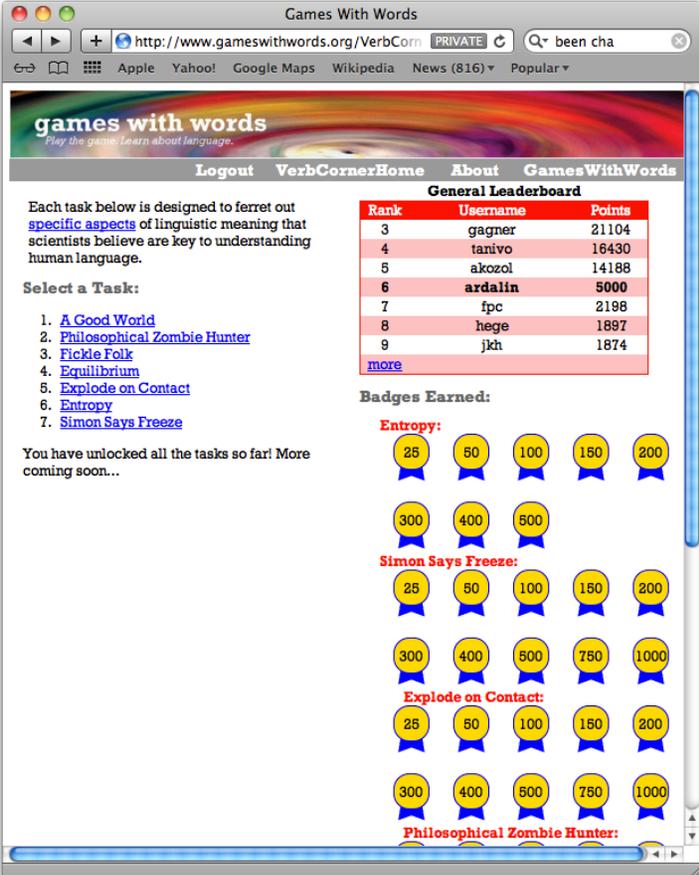


(c) under 20=7.1, over 51=5.7



(d) under 20=3.5, over 51=5.0

# Crowdsourcing Linguistic Judgments



The screenshot shows the 'Games With Words' website interface. At the top, there's a navigation bar with 'Logout', 'VerbCornerHome', 'About', and 'GamesWithWords'. Below this is a 'General Leaderboard' table with columns for Rank, Username, and Points. The table lists users like 'gagner', 'tanivo', 'akozol', 'ardalin', 'fpc', 'hege', and 'jkh'. To the left of the table, there's a section titled 'Select a Task:' with a list of seven tasks: 'A Good World', 'Philosophical Zombie Hunter', 'Fickle Folk', 'Equilibrium', 'Explode on Contact', 'Entropy', and 'Simon Says Freeze'. Below the tasks, a message states 'You have unlocked all the tasks so far! More coming soon...'. To the right of the tasks, there's a 'Badges Earned:' section with four categories: 'Entropy', 'Simon Says Freeze', 'Explode on Contact', and 'Philosophical Zombie Hunter'. Each category has a grid of yellow circular badges with blue ribbons, representing different point values (25, 50, 100, 150, 200, 300, 400, 600, 750, 1000).

Rank	Username	Points
3	gagner	21104
4	tanivo	16430
5	akozol	14188
6	ardalin	5000
7	fpc	2198
8	hege	1897
9	jkh	1874

**Badges Earned:**

- Entropy:** 25, 50, 100, 150, 200, 300, 400, 600
- Simon Says Freeze:** 25, 50, 100, 150, 200, 300, 400, 600, 750, 1000
- Explode on Contact:** 25, 50, 100, 150, 200, 300, 400, 600, 750, 1000
- Philosophical Zombie Hunter:** (no badges shown)

## Progress (old):

1,247 verbs

7 semantic features

~10,000 volunteers

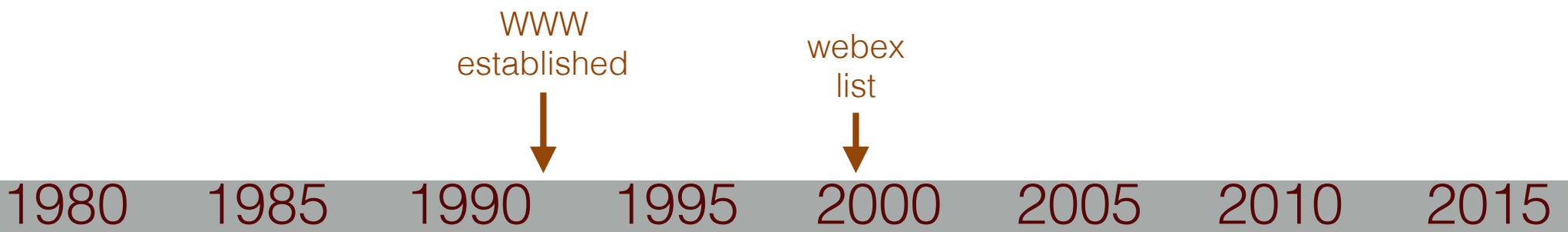
~450,000 judgments

## The Dream: Massive Online Experiments

- Many subjects, world-wide
- Many items
- Many contexts/manipulations
- Longitudinal
- Social networks
- “A whole literature in one experiment”

So are we living  
the dream?





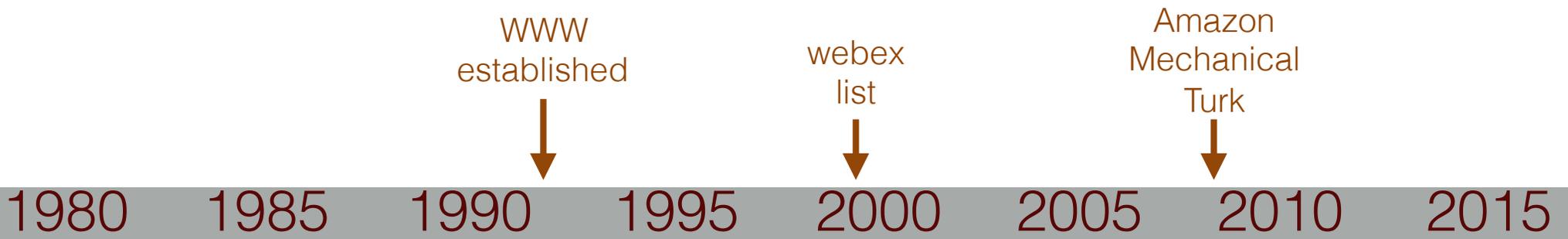
1st online  
psych study

WWW  
established

webex  
list

*In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167*

27% of APA journals  
have published online studies



1st online  
psych study

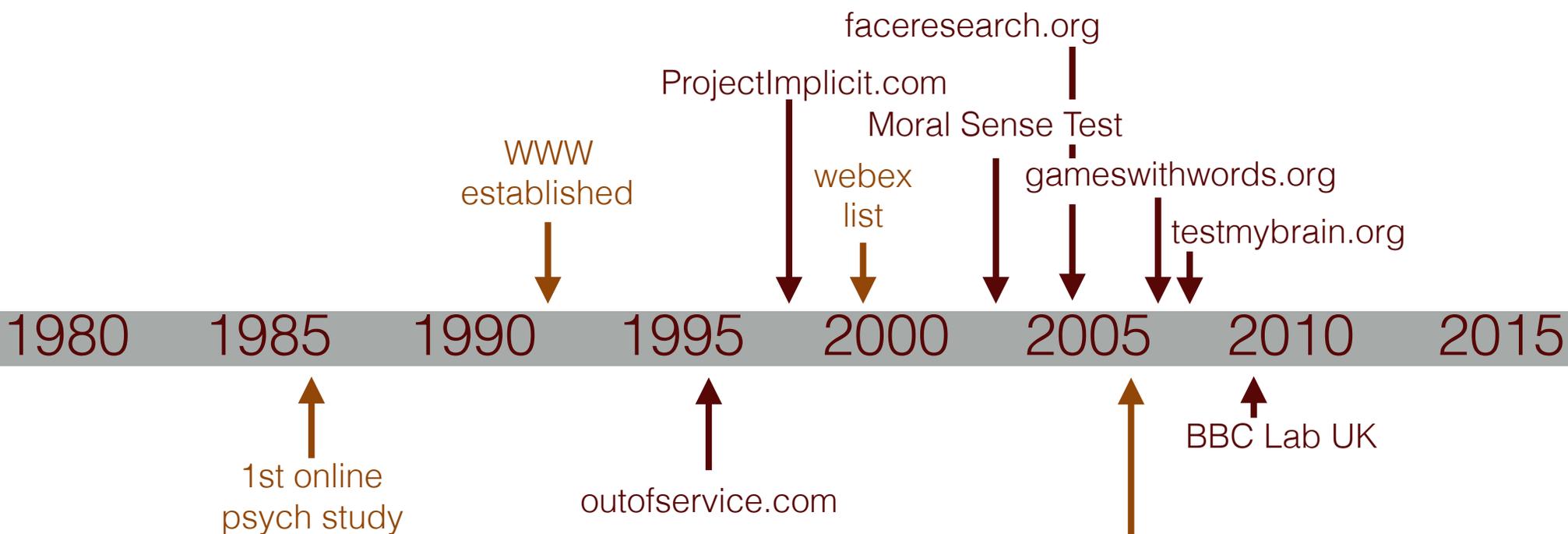
WWW  
established

webex  
list

Amazon  
Mechanical  
Turk

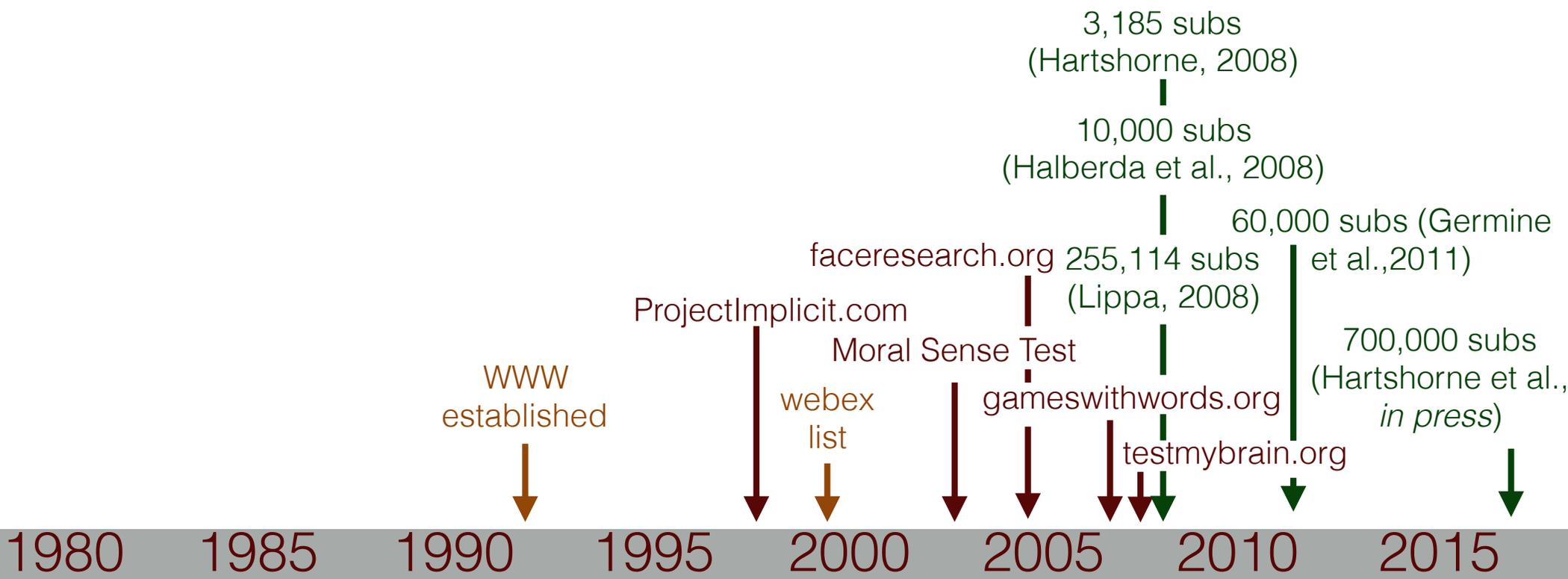
*In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167*

27% of APA journals  
have published online studies



*In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167*

27% of APA journals have published online studies



1980 1985 1990 1995 2000 2005 2010 2015

*In the study reported in this paper, EMS [Electronic Mail Service] was clearly the preferred data-collection method because it produced adequate data, response rates, and willingness for further participation, with little expenditure of researcher time or effort and a high degree of convenience for respondents – Sproull (1986), p. 167*

27% of APA journals have published online studies

# Cognitive Science Bottleneck



## Brick-and-Mortar labs

### Pros

- ~~Control over environment~~
- ~~Specialized equipment~~
- 100 yrs institutional knowledge

### Cons

- Restricted populations
- Low power
- Constrained by lab
  - Num. experimenters
  - Num. subjects that fit
  - Num. rooms, computers, etc.
  - Usually one-off, 30-60 min. intervals

# Barriers to Use

- Paradigm shift
  - “I don’t need that many subjects”
  - “My studies take more than 10 minutes!”
  - Recruitment: I can’t pay 1,000,000 subjects
  - What studies aren’t we running?
- Software
  - Robust scaling
  - Dynamic experiment design
    - Optimal Experimental Design
    - Active Learning
    - “Pipeline” experiments
  - Tracking repeat subjects / longitudinal data
  - Recruitment
    - Feedback, social media integration, mailing lists
    - Citizen Science
    - Forum, badges, etc.
- Analysis



# Pushkin

robust tools for  
massive online  
experiments

- **Completed**

- Stimuli: Text, video, audio, images
- Responses: Keyboard, RTs (within-subject), drag-and-drop
- Mobile-friendly webpages
- Stub website
- Auto-scaling (mostly)
- (Limited) dynamic stimulus selection

- **Available soon**

- “Endless” quizzes
- Support for robust dynamic stimulus section
- Interactive forum
- Badges & leaderboards
- Social media *authentication*
- Profile page
- Simplified social media *integration*
- Eyetracking / preferential looking

- **On deck**

- OED & Active Learning with WebPPL
- Experiment templates
- “Pipeline” experiments
- [Insert your ideas here]

