Lecture 2 - WEIRD-os

Jesse Hoey
School of Computer Science
University of Waterloo

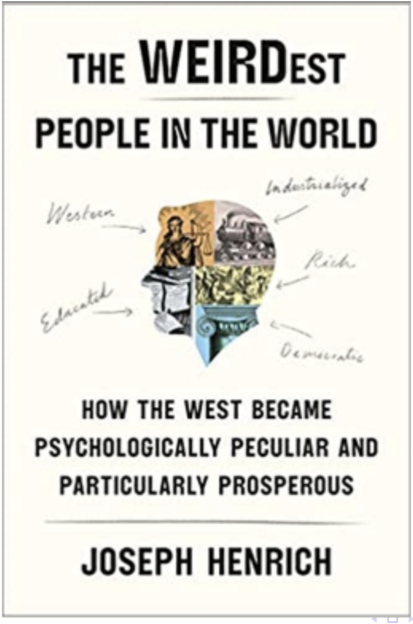
September 11, 2024

Readings:

- Henrich, J., The WEIRDest People in the World, *Beh. Br. Sci*, 33, 2010, 61-135.
- Henrich, J., *The WEIRDest People in the World*, 2010, Farrar, Strauss, Giroux, NY.

WEIRD

- Western
- Educated
- Industrialized
- Rich
- Democratic



WEIRD

from the abstract:

- *"Behavioral scientists routinely publish broad claims about human psychology and behavior in the world's top journals based on samples drawn entirely from Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies."*
- *"there is substantial variability in experimental results across populations and that WEIRD subjects are particularly unusual compared with the rest of the species – frequent outliers."*
- *"findings involve domains that are associated with fundamental aspects of psychology, motivation, and behavior"*
- *"We close by proposing ways to structurally re-organize the behavioral sciences to best tackle these challenges."*

Am I in the wrong room?

- This is
  not and anthropology course,
- but, knowing anthropology can help us design
  better affectively aware agents,
- and helps build better models of fairness in AI.
- We cover this first as it is foundational.

How do humans work?

First, we study how humans would do it, as in Henrich:

- Industrialized vs. Small Scale
- Western vs. Non-Western
- Contemporary Americans vs. the West
- Differences in America

Meta-analysis: only study phenonmenon which have led to claims of universality

Background: key assumptions

1. The database in the behavioral sciences is drawn from an extremely narrow slice of human diversity
2. behavioral scientists routinely assume, at least implicitly, that their findings from this narrow slice generalize to the species.

Contrast I: Industrialized vs. Small Scale

- Visual perception
- Social Dilemmas
- Folk-Biological Theories
- Spatial Cognition
- Other differences

Scale: Müller-Lyer Illusion

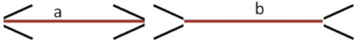


Figure 1. The Müller-Lyer illusion. The lines labeled “a” and “b” are the same length. Many subjects perceive line “b” as longer than line “a”.

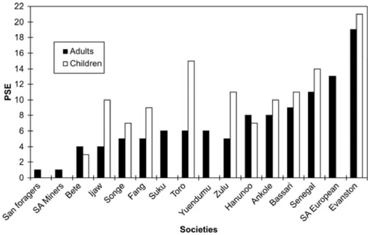


Figure 2. Müller-Lyer results for Segall et al.’s (1966) cross-cultural project. PSE (point of subjective equality) is the percentage that segment a must be longer than b before subjects perceived the segments as equal in length. Children were sampled in the 5-to-11 age range.

- People handle this illusion differently
- Americans are the most susceptible
- Some people don’t perceive the illusion at all
- Architectural bias?

Scale: Social Dilemmas

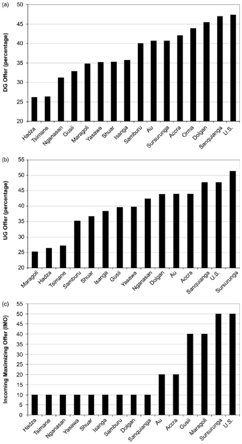


Figure 3. Behavioral measures of fairness and punishment from the Dictator and Ultimatum Games for 15 societies (Phase II). Figures 3A and 3B show mean offers for each society in the Dictator and Ultimatum Games, respectively. Figure 3C gives the income-maximizing offer (IMO) for each society.

- Dictator (DG) and Ultimatum (UG) games
- measures "altruism"
- different (cultural) notions of fairness
- Nowak found cooperation with reputation, inferred reputation is needed for cooperation
- but other groups (non-WEIRD) don't need reputation for cooperation, so the inference is incorrect
- IMO: willingness to punish

Scale: Folkbiological reasoning

what species are in this picture?

- Prototype WEIRD pattern "tree", "bird"
- Different cultures see different patterns e.g. "pine," "robin"
- How we relate (metaphorically) the world to the animal/natural world is a foundational part of how we reason
- cultural differences in this mapping lead to differences in behaviours and emotions

Scale: Folkbiological reasoning

the rock is [spatial relation]? the tent? the tree?

- Different cultures see different relations, directions, colors, integer amounts
- How we relate (metaphorically) the world to the natural world is a foundational part of how we reason
- cultural differences in this mapping lead to differences in behaviours and emotions
- Example: street addresses

Scale: Other Differences and Similarities

Differences:

- Risk aversion
- temporal discounting
- ...

Similarities:

- Color, other illusions
- Facial expressions under conditions - but only expressions, not the underlying emotion
- Essentialism (the belief that life is deep)
- ...

Contrast II: Western vs. Non-Western

- punishment and cooperation
- independence
- positive self-views
- personal choice
- conformity
- analytical/holistic reasoning
- moral reasoning
- other differences and similarities

Cooperative Games

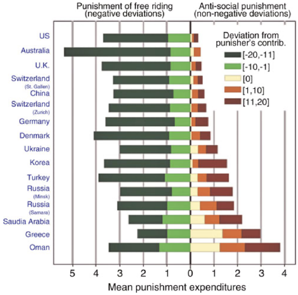
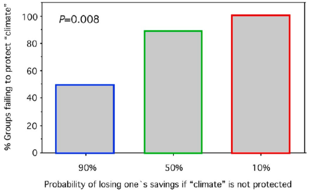
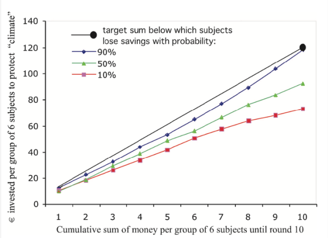


Figure 4. Mean punishment expenditures from each sample for a given deviation from the punisher's contribution to the public good. The deviations of the punished subject's contribution from the punisher's contribution are grouped into five intervals, where [-20,-11] indicates that the punished subjects contributed between 11 and 20 less than the punishing subject; [0] indicates that the punished subject contributed exactly the same amount as the punishing subject; and [1,10] ([11,20]) indicates that the punished subject contributed between 1 and 10 (11 and 20) more than the punishing subject. Adapted from Herrmann et al. (2008).

- public goods game
- contributions to central project are doubled and distributed equally
- even to non-contributors
- free riding
- punishment for non-contributers and for over-contributors
- how much would you pay to punish non-contributors?
- WEIRD punish more
- cultural differences in this mapping lead to differences in behaviours and emotions

Collective Risk Dilemmas





- collective risk game
- contributions to central pile must reach a threshold
- otherwise, everyone loses everything with some probability
- free riding
- used to set climate policy - how do we get people to cooperate?
- one idea: conformity (we will see this more later on)
- emotional signaling plays a role, and this will be different for different groups of people

Millinski *et al.*, 2008 The collective-risk social dilemma and the prevention of simulated dangerous climate change. PNAS, Feb 19, 2008 vol 105 no.7 2291-2294

Other differences

Western (vs. Non-Western):

- more positive self-views
- feel more free
- conform less
- more rule-based, less *resemblance-based*
- mate preferences: males value physical attractiveness more than females
- personality structure (5-factors)
- punishing free-rinding

Other differences (II)

analytic vs. holistic thinking:

- Mind-body duality at its core
- analytic thought is separate from the body and nature
- holistic thought sees integration between nature, mind, body
- perceptions, memory, and language reflect this difference
- personality traits are WEIRD
- situational traits are not
- often WEIRD thinking is challenged by reality: choices actually made disagree with stated preferences (see Mercier and Sperber *the Enigma of Reason*, 2017)

Other differences (II)

Morality and Kohlberg's three levels:

- pre-conventional
  (children - internal morality based on physiology)
- conventional
  (adults - external morality based on social norms)
- post-conventional
  (WEIRD - abstract morality based on philosophy)
- WEIRD: morality based on autonomy
- others: morality based on community or divinity

See also: George Lakoff. *Moral Politics*. 1996

Contrast 3 & 4: Typical contemporary Americans vs.

Rest of the west:

- individualism - Americans are more individualistic, but not necessarily towards family (see Fukuyama)
- many broad similarities

vs other American groups

- many samples in psychology experiments are American university students
- how does this compare with other americans?
- other generations?

Discussion

- Difficult to generalize from WEIRD samples
- Topics chosen by WEIRD researchers are biased
- Human diversity is a key component of human psychological research
- Some research (e.g. existential proofs) are ok with limited samples
- Henrich is mostly is interested in describing human nature
- We are interested in building human nature
- Throughout the class, we will discuss how this may be possible

Main message

- Understanding how situated, embodied, humans behave in general will help us to better understand how to build computer systems that interact with humans
- Emotion is a major factor in human behaviour, and so I label the course "Affective Computing"
- But it could be called "Social Computing" or "Hybrid Human-Machine Problem Solving"
- Computationally modeling how humans behave is therefore an important component in the process of building a computational system that can behave in a human-like way (thus solving the "strong" AI problem).
- But are you stuck in a Chinese Room? [discussion]

- What would you do?

Searle's Chinese Room

- What would you do?
- Start to make mistakes

Searle's Chinese Room

- What would you do?
- Start to make mistakes
- Look for correlations in subsequent inputs

Searle's Chinese Room

- What would you do?
- Start to make mistakes
- Look for correlations in subsequent inputs
- Establish a secondary communication based on the symbols

Searle's Chinese Room

- What would you do?
- Start to make mistakes
- Look for correlations in subsequent inputs
- Establish a secondary communication based on the symbols
- but what are these correlations?

Searle's Chinese Room

- What would you do?
- Start to make mistakes
- Look for correlations in subsequent inputs
- Establish a secondary communication based on the symbols
- but what are these correlations?
- psychology studies: 96% of samples come from 12% of the world

Searle's Chinese Room

- What would you do?
- Start to make mistakes
- Look for correlations in subsequent inputs
- Establish a secondary communication based on the symbols
- but what are these correlations?
- psychology studies: 96% of samples come from 12% of the world
- seek difference across cultures

Our Goal

To build AI systems that can

- get out of the Chinese Room,
- by learning to adapt to a novel population of people,
- thus becoming "fair" AI by definition of learning.

adapting to your society is a hallmark of intelligence

Next Lecture

- Metaphorical Reasoning
- George Lakoff and Mark Johnson *Metaphors we Live By*, 2nd edition (afterword), 2003.
- **optional** Keith J. Holyoak and Paul Thagard *Mental Leaps*, MIT Press, 1995.