

Western researchers can and should be accessing Eastern samples: scientific validation and practical guidance

Liman Wang, Randy Y. Gao, Minah H. Jung, Iris W. Hung, Leif D. Nelson

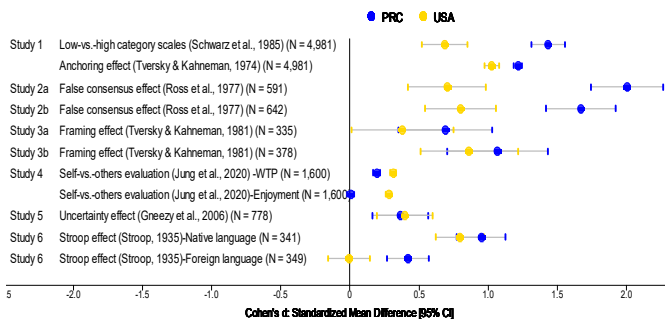
Fudan University, New York University, The Chinese University of Hong Kong, Shenzhen, University of California – Berkeley

BACKGROUND Psychology and Behavioral Science has historically restricted its research sample to Western, Educated, Industrialized, Rich, Democratic (**WEIRD**) populations (Henrich et al., 2010). This lack of diversity **undermines the generalizability** of research findings (McNemar, 1946; Arnett, 2008).

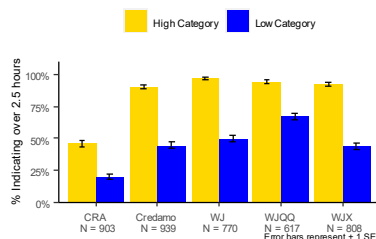
SUMMARY Through eleven pre-registered experiments (N = 18,534), we demonstrate that **Chinese online platforms yield high-quality data**. Furthermore, **language should not be a significant barrier** for Western scholars seeking access to these Asian samples.

- Studies 1-6 replicated seven experiments from diverse fields such as psychology, consumer research, behavioral economics, and cognitive psychology. The results offer compelling evidence of the reliability of data from Chinese platforms.
- Studies 7-8 specifically replicated cross-cultural research, showcasing that Chinese online platforms can be invaluable data for cultural psychologists.
- Study 9 highlighted that Chinese online platforms furnish data of high psychometric quality, as evidenced by a test-retest design.
- Studies 10-11 delved into the impact of varying translation versions on participants' responses. Although there were noticeable differences in responses between translation versions, most of the variations were generally deemed acceptable.

Overview of studies 1-6



Low vs. high category scales (Study 1) (N = 4,981)



Question: How many hours do you spend on your mobile phone for entertainment each day? (Low category: from “up to a half hour” to “more than two and a half hours”; high category: from “up to two and a half hours” to “more than two and a half hours”)

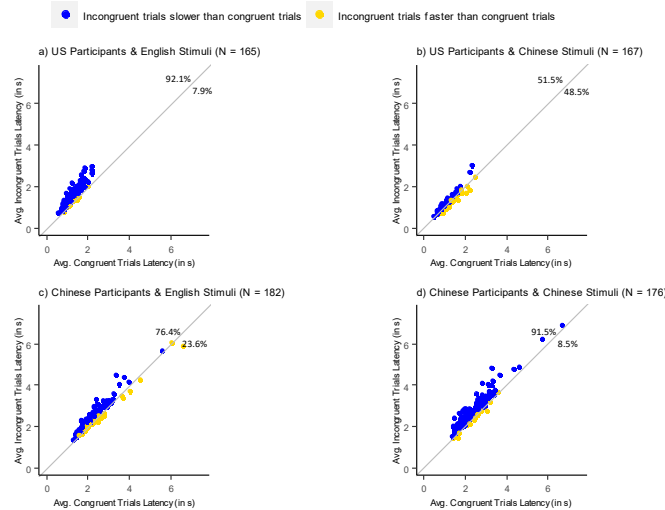
Result: The Effect was replicated with each sample.

Stroop effect (Study 6) (N = 690)

Stimuli

Congruent & English
Red, Green, Blue, Orange
Congruent & Chinese
红色, 绿色, 蓝色, 橙色

Incongruent & English
Red, Green, Blue, Orange
Incongruent & Chinese
红色, 绿色, 蓝色, 橙色



Measurement

Response time was measured using the timestamps from Qualtrics and Credamo.

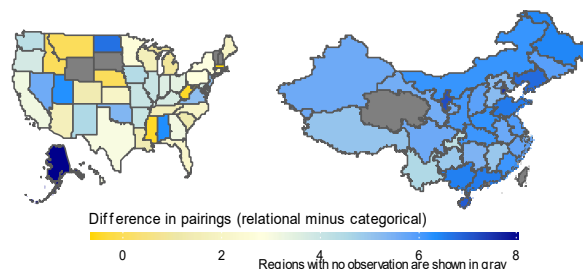
Result

The Stroop Effect was replicated across all conditions except for the “American participants with Chinese stimuli” group. The effect sizes for the four conditions (a, b, c, d) are reported as Cohen's d values of 0.80, 0.00, 0.42, and 0.95, respectively.

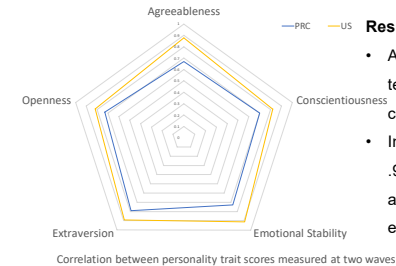
Relational vs. Categorical reasoning (Study 7) (N = 3,535)

(Ji, Zhang, & Nisbett, 2004)

A categorical pairing: Panda, Monkey, Banana. A relational pairing: Panda, Monkey, Banana.



Test-retest reliability (Study 9) (N = 348)

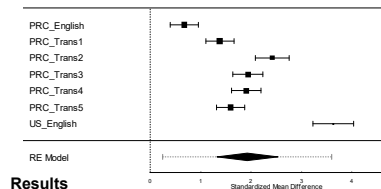


Results

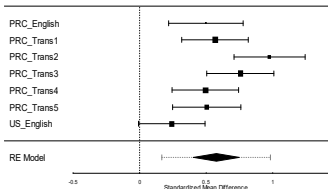
- All dimensions OF the Big Five showed high test-retest reliability within samples from both countries.
- In addition, participants' rating for age ($r_{US} = .99$, $r_{PRC} = .99$), gender ($K_{US} = .95$; $K_{PRC} = .99$), and education level ($K_{US} = .92$; $K_{PRC} = .92$) exhibited high test-retest reliability.

The effect of translation variations (Study 11) (N = 1,969)

The Knobe effect (Knobe, 2003)



The Soda Pricing task (Thaler, 1985)

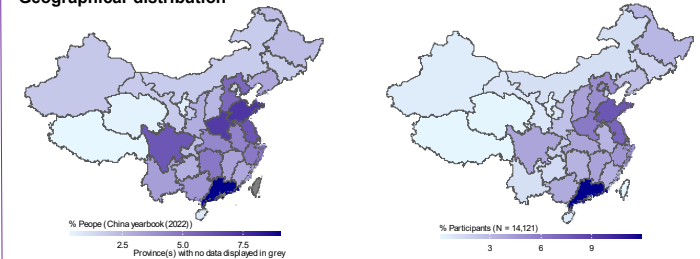


Results

- All conditions showed the Knobe effect, with the chairman seen as more intentional in harming than helping the environment.
- Chinese participants, across all translations, agreed to pay more for soda at a resort hotel than a grocery store, an effect not seen with Americans.

Other evidence

Geographical distribution



Attentiveness

Across studies, for Instrumental manipulation checks 89.9% of Chinese and 89.4% of Americans passed; for the stimulus recall check, 99.8% of Chinese and 94.1% of Americans passed.

Reference

1. Arnett, J. J. (2008). The Neglected 95%: Why American Psychology Needs to Become Less American. *The American psychologist*, 63(7), 602-614.
2. Berinsky, A. J., Huber, G. A., & Lenz, G. S. (2012). Evaluating online labor markets for experimental research: Amazon.com's Mechanical Turk. *Political analysis*, 20(3), 351-368.
3. Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A New Source of Inexpensive, Yet High-Quality, Data?. *Perspectives on Psychological Science*, 6(1), 3-5.
4. Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature*, 466(7302), 29-29.
5. McNemar, Q. (1946). Opinion-attitude methodology. *Psychological bulletin*, 43(4), 289.
6. Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making*, 5(5), 411-419.