Some Reflections on Klein et al.’s replication of Study 2, “Cultural preferences for formal versus intuitive reasoning.”

Ara Norenzayan

*University of British Columbia*

Address correspondence to:
Ara Norenzayan
Professor
Department of Psychology
University of British Columbia
2136 West Mall
Vancouver, BC V6T 1Z4 Canada
email: ara@psych.ubc.ca
Abstract

Klein et al replicated Study 2 of Norenzayan, Smith, Kim, & Nisbett (2002). Here I reflect on their effort in relation to the original study, and offer thoughts and clarifications.
In an ambitious effort to assess the evidentiary value of 28 published findings across 36 countries, Klein et al replicated Study 2 of Norenzayan, Smith, Kim, & Nisbett (2002). Here I reflect on their effort in relation to the original study, and offer thoughts and clarifications.

The original study had two hypotheses: (1) overall greater preference for the rule-based strategy in the categorization (“belong to”) condition and overall greater preference for the family resemblance-based strategy in the similarity judgment (“similar to”) condition. (2) Less preference for the rule strategy and more preference for the family resemblance strategy in the East Asian sample relative to the European American sample in both experimental conditions (with Asian Americans in-between), a hypothesis derived from the theory of holistic vs. analytic cognition from cultural psychology (Nisbett, Peng, Choi, & Norenzayan, 2001). The results of Norenzayan et al showed evidence consistent with hypothesis (1), except for the European American sample which showed equally high levels of preference for the rule strategy in both experimental conditions. As to hypothesis (2), as stated in the article, a cultural difference was found in the “similar to” condition consistent with the hypothesis, but contrary to hypothesis, there was no cultural difference in the “belong to” condition.

Klein et al preregistered and tested the hypothesis that the percentage of rule-based responses would be greater in the “belong to” condition compared to the “similar to” condition, as we had hypothesized in the original study. This aspect of Klein et al.’s efforts follows closely the original study design, and consistent with the original hypothesis (1), this experimental effect was found in their vastly larger and more diverse cross-cultural sample.
While the stimuli and experimental instructions were closely matched to the earlier study, there were notable differences in sampling and implementation in the two studies. In the study we conducted, a narrow sampling strategy was used -- the participants were university students at an elite American public university closely matched on cognitive abilities and education levels, but differing in cultural background. All participants completed this one judgment task on desktop computers in individual cubicles in a standardized environment. These steps were taken to ensure a high degree of experimental control. But due to the fact that Klein et al.’s study was pursuing a different set of research questions, 13-15 different tasks were administered to each participant in a randomized order, and the samples came from a range of countries, varying in age, education, and socio-demographic characteristics, with a correspondingly large sample size with very high statistical power.

Moreover, there were two additional critical divergences in the cross-cultural comparisons, rendering their findings in this regard particularly difficult to interpret in relation to Hypothesis (2) of the original study. First, Klein et al. introduced a novel procedure in their cross-cultural comparisons: they computed a WEIRDness score by country, drawing on the letters of the WEIRD acronym from Henrich, Heine, & Norenzayan (2010). This approach is not derived from Henrich et al.’s conceptualization of WEIRD sampling in psychology, nor is it found in the Norenzayan et al. study. There are many reasons to be skeptical of its merits -- instead, WEIRDness can be more appropriately and precisely conceptualized as a cultural distance measure; moreover, the variation in WEIRD is multidimensional, which is why it is crucial to theoretically identify the precise basis and level of cultural variability as they apply to a particular
psychological phenomenon (see Henrich et al., 2010; Norenzayan & Heine, 2005). This is a broader issue that deserves detailed discussion, but it is beyond the scope of this brief commentary.

Second, the hypothesis we tested was whether European Americans would be more likely to prefer the rule than East Asians in both the “belong to” and the “similar to” conditions. But Klein et al tested a different hypothesis: whether the magnitude of the effect of experimental instruction (“belong to” vs. “similar to”) showed heterogeneity across all samples, and in an exploratory test, whether this effect was different for samples classified as WEIRD vs. less WEIRD. They found that this effect was stronger for WEIRD than for less WEIRD. Although this was not a question of interest in the original study, it is a noteworthy finding in and of itself, considering that a recent paper by Murphy, Bosch, & Kim (2017) failed to find this effect of experimental instruction in 5 American samples (consistent with what was found in Norenzayan et al.’s European American sample); but Murphy et al. did observe this effect in a Korean sample. This pattern of results is at odds with the findings of Klein et al. (Note, however, Murphy et al.’s sample size, like Norenzayan et al.’s, dwarfs in comparison to the current study’s massive sample size). As far as I am aware, there are no a priori theoretical reasons from cultural psychology that would have predicted a cultural difference in responsiveness to experimental instruction; it is an unexpected finding that might be of interest to researchers studying cultural variability in cognition.
Action Editor

Daniel J. Simons served as action editor for this article.
References


