

## Sex differences in spatial abilities: Methodological problems in Hoffman et al.

Hoffman et al. (1) claimed to provide evidence that “nurture” (i.e., residing in a patrilineal vs. matrilineal tribe in India) critically affects sex differences in spatial abilities. Unfortunately, their conclusion is undermined by major problems with their measures of spatial ability and sex equality.

The first and biggest problem is with their measure of spatial abilities. “Spatial abilities” are a complex cognitive domain, with facets ranging from location memory (favoring women) to navigation in 3D virtual space (favoring men) (2). The puzzle used by Hoffman et al. (1) is similar to the Object Assembly subtest of the Wechsler Adult Intelligence Scale (3); sex differences on this task are extremely small ( $d = 0.10$ ), at least 10-fold smaller than those found for spatial measures showing the largest sex differences. It is odd that Hoffman et al. (1) chose to investigate sex differences with this kind of sex-insensitive task.

The second problem is the lack of a control task. The insensitivity of the task used by Hoffman et al. (1) suggests that their finding that men outperform women in a patrilineal tribe but not a matrilineal tribe is not related to sex differences in spatial abilities per se but to other factors instead. Education, as they noted, is likely one of these. The use of a cognitive control task tapping nonspatial abilities would have allowed for an assessment of the specificity of the effect, but, unfortunately, such a task was not included.

Third, defining sex equality as matrilineality is problematic, because cross-cultural studies generally show that equality (a multidimensional construct) is not systematically correlated

with descent system (4). From the descriptions of Hoffman et al. (1), it appears that women in the matrilineal Khasi have more economic power and better education, but this ignores other sex equality dimensions, such as positions of political and religious leadership, domestic authority, and autonomy. Without such measures, it is unclear whether the Khasi are, in fact, more sex-egalitarian than the Karbi. Furthermore, a recent 53-nation cross-cultural study has shown that sex differences favoring men on validated, reliable, multi-item spatial measures are positively associated with United Nation indices of sex development and empowerment (5), a pattern opposite to that reported by Hoffman et al. (1). For all these reasons, the study by Hoffman et al. (1) failed to support their conclusions.

**Drew H. Bailey<sup>a,1</sup>, Richard A. Lippa<sup>b</sup>, Marco Del Giudice<sup>c</sup>, Raymond Hames<sup>d</sup>, and Dave C. Geary<sup>a,e</sup>**

<sup>a</sup>Department of Psychological Sciences and <sup>c</sup>Interdisciplinary Neuroscience Program, University of Missouri, Columbia, MO 65211; <sup>b</sup>Department of Psychology, California State University, Fullerton, CA 92834; <sup>c</sup>Center for Cognitive Science, Department of Psychology, University of Turin, 10123 Torino, Italy; and <sup>d</sup>Department of Anthropology, University of Nebraska, Lincoln, NE 68588

1. Hoffman M, Gneezy U, List JA (2011) Nurture affects gender differences in spatial abilities. *Proc Natl Acad Sci USA* 108:14786–14788.
2. Geary DC (2010) *Male, Female: The Evolution of Human Sex Differences* (American Psychological Association, Washington, DC).
3. Wechsler D (1981) *Manual for the Wechsler Adult Intelligence Scale-Revised. (WAIS-R)* (Psychological Corporation, San Antonio, TX).
4. Whyte M (1978) *The Status of Women in Preindustrial Societies* (Princeton Univ Press, Princeton).
5. Lippa RA, Collaer ML, Peters M (2010) Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Arch Sex Behav* 39:990–997.

Author contributions: D.H.B., R.L., M.D.G., R.H., and D.C.G. wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. E-mail: drewhalbailey@gmail.com.

## Reply to Bailey et al. and Daly: Indigenous societies enable identification of nurture but require nonstandard measures

Policy-makers and activists have control over nurture but not nature. Thus, identifying the effect of changing nurture without directly altering nature is imperative for understanding whether we can reduce gender differences. The Khasi and Karbi offer a unique “natural experiment,” allowing us to identify the effect of “changing” nurture without directly altering nature. Standard cross-cultural studies cannot achieve such identification because nurture often covaries with, for example, genetics. Although experiments, such as those that manipulate stereotype salience or training, enable us to identify the role of nurture, they do not enable us to know if changes to nurture could have a dramatic impact on an entire society. This natural experiment is the main methodological innovation of our paper. To our knowledge it is the first comparison of two matched societies, used to identify the role of nurture while holding constant nature.

Unfortunately, this sample comes at a cost: Standard measures of spatial abilities are too abstract for our subjects, who are near-subsistence indigenous villagers, and whose education consists of practical know-how, such as how often crops should be rotated, and a fact-based curriculum required by the state, such as what is the capital of India. In fact, we tried to have our subjects play a children’s memory game, Simon (Fig. 1*A*), but we were unable to train subjects. Our puzzle measure actually began as a 12-piece puzzle with the five unrelated sides uncovered (Fig. 1*B*); however, after subjects took more than 15 min to complete the puzzle, we covered five of the sides, then removed 6 of the pieces, and then removed 2 more pieces.

We therefore agree with the sentiments of Bailey et al. (1) and Daly (2): Our study should not be viewed as the definitive study on this topic but as a proof of concept, which should propel researchers to exploit this unique sample to its fullest advantage. Further research using noncognitive measures, as well as alternative spatial measures would prove invaluable in addressing some of the shortcomings pointed out by Bailey et al. (1) and Daly (2). Moreover, such research would reveal the generality of our results and could focus activist efforts on traits that are most amenable to nurture. Also, if the measures are chosen to be more or less gender-dimorphic and more or less influenced by motivation, stereotype threat, and training, for example, this research, in addition to addressing some of the astute criticisms of Bailey et al. (1) and Daly (2), could also reveal mechanism, which would likewise be invaluable for focusing activists’ efforts. We welcome collaboration with psychologists and anthropologists, such as the experts to whom this letter replies, to help us develop such measures of spatial and nonspatial cognitive



**Fig. 1.** Two measures that were too complex for our sample. (*A*) Game of Simon requires individuals to repeat a sequence of flashing lights (e.g., red, blue, blue, yellow). We attempted to use this game as a measure of short-term memory, a nonspatial cognitive ability. Subjects consistently pressed the most recent light, (e.g., yellow), despite several different research assistants’ attempt to explain the notion of a “sequence.” After a handful of subjects, we dropped the measure and could not come up with a replacement in time. (*B*) This 12-piece, six-sided puzzle served as our original measure of spatial abilities. Subjects were taking longer than 15 min to solve the puzzle, leading us to cover the five unrelated sides. When the puzzle still proved too time-consuming, we removed half of the pieces. When the puzzle still proved too time-consuming, we removed 2 more pieces, thereby settling on our final measure.

abilities that are easy to explain and quick to implement, to take full advantage of this unique sample.

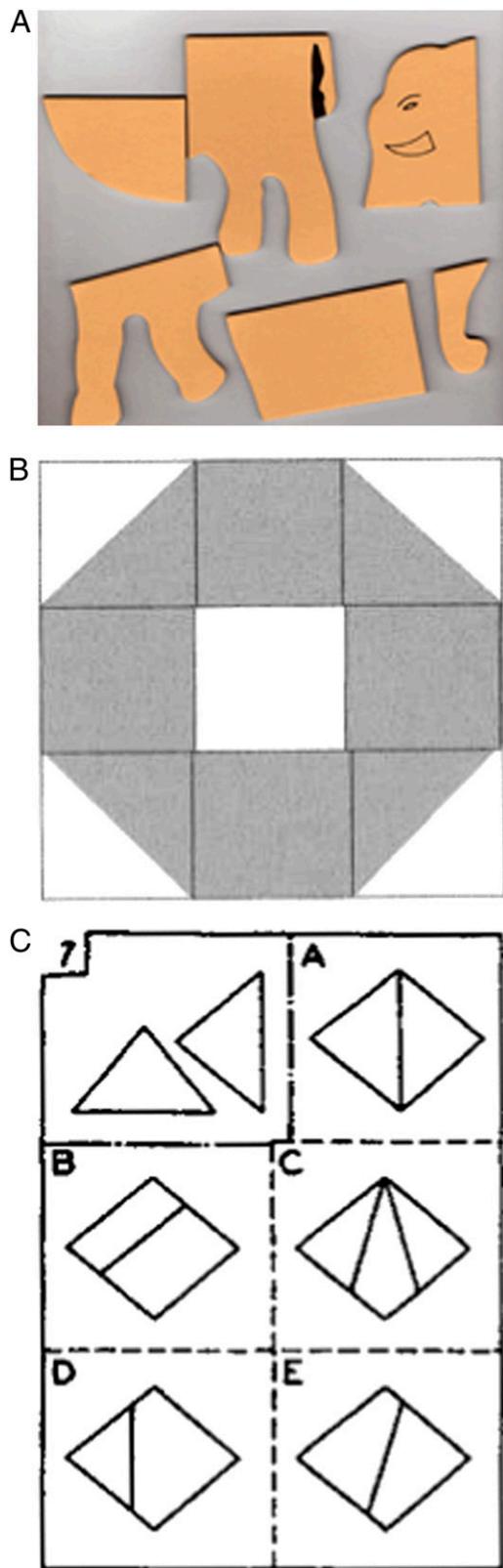
### Response to Specific Criticisms

We disagree with Bailey et al. (1) and Daly (2) that our measure obviously resembles the three they present (Fig. 2). It is not clear to us which measure our puzzle most resembles, nor does it seem clear to Bailey et al. (1) and Daly (2), who mention three different possibilities. Moreover, all measures of spatial abilities involve a complex mixture of cognitive traits, making it unlikely that ours maps directly onto any. For instance, one volume on intellectual assessments (3) notes:

Author contributions: M.H., U.G., and J.A.L. wrote the paper.

The authors declare no conflict of interest.

<sup>1</sup>To whom correspondence should be addressed. E-mail: mohoffman@ucsd.edu.



**Fig. 2.** Three separate measures that Bailey et al. (1) and Daly (2) argue are most similar to ours. (A) In the object assembly measure, subjects form an identifiable object from several differently shaped pieces. (B) In the block design measure, subjects arrange two-colored blocks to repeat a design. (C) In the paper form board measure, subjects pick an image from A–E that can be produced from a focal image (Upper Left).



**Fig. 3.** Billboard in the nearby city of Shillong, where Khasi ex-village men are demanding equal rights.

object assembly is designed to assess perceptual organization, integration and synthesis of part-whole relationships, and use of sensory-motor feedback. It also involves spatial ability, fluid ability, visual-motor reasoning, trial-and-error learning, visual-motor coordination, cognitive flexibility, persistence, motor coordination, and dexterity... nonverbal reasoning, speed of mental processing, and anticipation of relationships.

We share the concern of Daly (2) that motivational/personality confounds may exist in our measure. However, we would have this same concern for other measures as well. The aforementioned volume states “[object assembly is influenced by] cognitive flexibility, persistence, motor coordination, and dexterity” (3). If anything, we are less concerned with our measure because the large financial incentives (1/4 of a day’s wage) should motivate even the most indolent of subjects. Moreover, after our subjects waited in line upward of an hour in the rain, when it came time to solve the puzzle, even the most inattentive could concentrate for the requisite 30–60 s. Far from being a simple task, only 607 of 1,279 subjects earned the bonus. We also see no problem if our results, in fact, were fully driven by motivation/personality, because motivation/personality affects performance in standardized tests, science, technology, engineering, and mathematics (STEM) classes, and STEM careers as well. Thus, the implications of our research for those wishing to reduce the gender gap therein would not be dramatically altered, even if the cognitive interpretation were different.

We agree with Bailey et al. (1) that the Khasi are not gender-equal. Women cannot be religious or political leaders, and the Khasi do show a stark division of labor, with women tending to the home and men to the field.

However, we disagree with Bailey et al. (1) that Khasi women are not treated better than Karbi women, which is all we needed for our purposes. Women, as is legally dictated, are the sole property owners among the Khasi, whereas women are the minority property owners among the Karbi (4). Despite the claim of Bailey et al. (1) to the contrary, the original source provided by these researchers does show that descent system correlates with treatment of women (ref. 5, pp. 133–134 and

