

# Presidential Column: How to Spot Bias in Research

November 01, 2006

A 1924 article published in the *American Journal of Psychiatry* reported the results of the following laboratory task: “A meaningless picture was produced by pouring India-ink of different intensities on a piece of thick lined paper and then pressing the paper under a glass plate. In addition some abstract lines were drawn by chance on the picture and a few pieces of white paper cut also by chance pasted on the same.”

Two genetically distinct groups of 25 human participants were shown the meaningless picture and asked to talk, for two minutes, about any objects in the picture that they recognized. One group of participants was more prone to articulate what the researcher deemed “insignificant” statements, such as “I really don’t see any objects” or, in the case of one participant, “If you are a photographer, doctor, I’d like to tell you that you had better change professions.” The other group was more likely to talk about objects such as dogs, elephants, and steamships. Because the first group had a higher ratio of “insignificant” to “significant” statements, that group was also deemed “talkative.”

An article published just last year in the same journal reported the results of the following laboratory task: “A target force was applied to the subject’s left index finger by a torque motor. Subjects were then required to reproduce the force they just experienced, either directly by pressing with the index finger of their right hand or indirectly by using a joystick controlling the torque motor.”

The data from the study’s two groups of 20 participants are shown in Figure 1. Both groups “reproduced the original force much more accurately” when using the joystick to control the torque motor than when using only their index finger. And when using the joystick, the two groups did not differ. In contrast, when using their index fingers to control the torque motor, the group represented in blue was significantly “more accurate at the task;” that group’s ability to match the target force more closely resembled “perfect performance.”

These results were interpreted by the researchers as supporting the hypothesis that one group was characterized by “a dysfunction in their ability to predict the sensory consequences of their actions.” Indeed, the title of the article was “Evidence for Sensory Prediction Deficits” among this group of participants. Which group? The blue group.

One last example. Using the Deese-Roediger-McDermott “false memory” paradigm, two groups of participants were presented auditorily with lists of semantically related words (e.g., *bed*, *rest*, *awake*, *tired*, and *dream*), and later asked to discriminate between words they’d heard and words they hadn’t heard, including words that were semantically associated to words they’d heard (e.g., *sleep*). As shown in Figure 2, the green group demonstrated significantly better memory discrimination than the purple group; the green group was less likely to falsely recognize words they hadn’t heard, despite the false words’ semantic association with words they’d heard.

The green group's better memory discrimination was attributed to their mentally representing words "in an aberrant manner," even though a concurrent — and direct — test of semantic clustering found no differences between the green and purple groups. The green group's aberrant semantic mental representations was hypothesized to stem from "anatomic abnormalities ... or as a result of an as-yet unknown pathology."

When another research team reported no difference between green- and purple-type participants in either false recall or false recognition, the authors of the study that had observed the green group's better discrimination interpreted the other study's lack of a between-group difference to the green group also having "frontal-executive impairment."

So, we have a group of individuals whose more factual descriptions of a meaningless picture were interpreted as insignificant and talkative. We have a group whose more accurate tactile matching was interpreted as sensory prediction deficits. And we have a group whose heightened memory discrimination in one study was interpreted as the result of an as-yet unknown pathology, and whose equivalent performance in another study was interpreted as frontal executive impairment.

Confused? If I told you that the group interpreted as providing insignificant and talkative descriptions comprised "normal females," the group interpreted as unable to predict the sensory consequences of their actions comprised persons diagnosed with schizophrenia, and the group interpreted as having aberrant mental representations and frontal executive impairment comprised persons diagnosed with autism, would it help? It shouldn't.

Maggio (1991) recommends that we test our writing for bias by substituting our own group for the group we are discussing. If we feel offended, then our writing is biased. I recommend that we test our interpretations for bias by peeling off the labels, as I've done here. If our interpretations make little sense, then our science is biased.

## **References**

Maggio, R. (1991). *The bias-free word finder: A dictionary of nondiscriminatory language*. Boston: Beacon Press.