Modeling Unconscious Gender Bias in Fame Judgments:
Finding the Proper Branch of the Correct (Multinomial) Tree

Sean C. Draine, Anthony G. Greenwald,
University of Washington

and

Mahzarin R. Banaji
Yale University

Abstract. In the preceding article, Buchner and Wippich used a guessing-corrected, multinomial process-dissociation analysis to test whether a gender bias in fame judgments reported by Banaji and Greenwald (*Journal of Personality and Social Psychology*, 1995, *68*, 181-198) was unconscious. Buchner and Wippich concluded that the gender-bias effect was not unconscious on the basis of finding no difference in model-estimated familiarity between previously presented nonfamous male and female names. This conclusion is questioned by noting that (a) the gender difference in familiarity that Buchner and Wippich modeled was different from the critical gender difference in criterion for fame judgments reported by Banaji and Greenwald, (b) the assumptions of Buchner and Wippich■s multinomial model exclude processes that are plausibly involved in the fame judgment task, and (c) constructs of Buchner and Wippich■s that correspond most closely to Banaji and Greenwald■s gender-bias interpretation are formulated so as to preclude modeling that interpretation.

Banaji and Greenwald (1995; BG, hereafter) used the false fame effect (Jacoby, Kelley, Brown, and Jasechko, 1989) to examine an implicit, and possibly unconscious, stereotype that associates male (more than female) gender with fame-deserving achievement. In BG■s four experiments subjects were asked, in the first of two sessions, to judge the pronounceability of each of a list of male and female names, half famous and half not. One or two days later, subjects were asked to judge the fame of names on a larger similarly composed list, including both the old names (i.e., those seen in the previous session), and new ones. BG performed a signal detection analysis on the fame judgments, examing whether name gender affected sensitivity to fame (measured by dN) or the criterion for fame judgments (measured by log $\$$). Consistently in all four of BG■s experiments, subjects used a lower (more liberal) criterion of fame for judging old male (than female) names. The findings showed that subjects were more likely to attribute their sense of familiarity with old names to fame when the name was male rather than female.

Because BG■s gender difference in fame judgments occurred only when names had presumably been given a boost in familiarity by an unremembered prior presentation, BG suggested that it reflected an unconsciously operating, or implicit, stereotype. The implicit-cognition interpretation was also supported by subjects■ self-reported post-experimental unawareness of a relation between name gender and their judgments, and by the lack of correlation of individual differences in criterion difference between male and female familiarized

names and explicit measures of gender stereotyping. The possibility of implicit operation of discriminatory stereotypes is significant because such stereotypes may be difficult to suppress, even by well-intentioned persons. (Greenwald & Banaji, 1995, review previous findings that similarly indicate unconscious operation of stereotypes.)

Buchner and Wippich (1995; BW, hereafter) set out to test whether or not the gender stereotyping observed by BG can be considered unconscious. BW used an ☻extended measurement◪ multinomial process dissociation model (Buchner, Erdfelder, & Vaterodt-Plönnecke, 1995) as the methodological tool for distinguishing conscious from unconscious components of fame judgments. The relevant portion of BW◪s findings and conclusions was summarized as follows:

> In both Experiment 1 and 2, we found that the criterion for calling a name famous was more liberal for male names than for female names. While these results ... replicated findings reported by Banaji and Greenwald (1995), they also presented problems for assessing whether the biases in the fame judgments were due to unconscious, automatic memory processes. . . . Unfortunately, in both of our experiments a supposedly ☻unconscious◪ effect on fame judgments disappeared as soon as response bias effects were taken into account explicitly by applying the extended measurement model for the process dissociation procedure. (Buchner & Wippich, in press, p. 34 of draft ms.)

In the following paragraphs, we comment on BW by noting that (a) the gender difference that they observed was not the same one that was critical to BG◪s conclusion about unconscious or implicit operation of a gender stereotype, (b) BW◪s extended measurement model omits representation of some processes that were likely involved in fame judgments, and (c) the portion of BW◪s extended measurement model that corresponds to BG◪s critical gender-bias finding is formulated so as to define that gender-bias effect out of existence. Individually, and certainly also in combination, these three points imply that BW◪s methods and findings are focused on issues other than interpreting the gender bias observed by BG.

What gender difference did BW observe? In Session 1 of BW◪s procedure, subjects studied a list of 10 famous and 60 nonfamous male and female names. In Session 2, a day later, the same subjects classified, as famous or not, names on a longer list that included Session 1's 60 nonfamous names along with 60 new nonfamous and 60 famous names. For BW◪s *inclusion* condition subjects were (mis)informed that, if they could recall the name as one that had been presented in Session 1, they could be sure that it was famous. For their *exclusion* condition, subjects were instead informed that if they could recall the name as one presented in Session 1, they could be sure that it was *not* famous. BW modeled the probability of classifying names as ✍famous✑ under these inclusion and exclusion instructions using a multinomial binary tree model of the process of making the judgments in each condition (see Figure 1 of BW). According to their model, subjects◪ judgments of fame are determined by three cognitive processes that they identified as *conscious* influences of memory, *unconscious* influences of memory, and *guessing*.

BW identified unconscious influences of memory on fame judgments with a model parameter (u_c) that increases to the extent that exclusion-condition subjects, contrary to their instructions, classify as "famous" names that had been presented in the first session. By contrast, the conscious contribution to fame judgments is estimated by a model parameter (*c*) that increases to the extent that, consistent with instructions, inclusion-condition subjects classify as "famous" names that had been presented in the first session, or exclusion-condition subjects classify those same names as "nonfamous." In both of their experiments, BW observed a gender difference in that the probability of classifying as "famous" all names (and especially new nonfamous names) was greater for male than female names (see their Tables 1 and 2).

In judging whether the gender-of-name difference in judgments of fame could be given an interpretation as reflecting unconscious process BW concluded no, because their multinomial model's u_c parameter was similar in value for male and female names. As they expressed it, "if the gender bias was an unconscious effect, then we would expect u_c to be larger for male than for female names" (page 20 of draft). From BW's discussion and their multinomial model, it can be seen that their u_c parameter represents the unlabeled sense of familiarity that can occur when, in Session 2, a previously presented nonfamous name is not recalled as having been seen in Session 1. In contrast to BW's method of testing for unconscious contributions to gender stereotyping, BG assumed that familiarity-without-recall should be *equal* for male and female names. Consequently, BW's finding of no difference in the u_c parameter between male and female names was fully consistent (and not, as BW suggested, at odds) with the BG interpretation.

What happens when male and female nonfamous names seem familiar? As already described, BW interpreted their u_c parameter as the probability of being in the state of familiarity in response to an old nonfamous name when the name was not recalled as having been seen in the prior session. It can be seen in BW's Figure 1 that this state is assumed *always* to produce a judgment that the name is "famous." However, self-reports of subjects who have been in false fame experiments indicate that this state can also lead to judgments of "nonfamous," either (a) when subjects attribute the familiarity to extra-experimental sources (they might have nonfamous acquaintances with the same first or last names, or they might judge that the names seem rather common – e.g., Jane Smith may seem familiar even if you don't know anyone, famous or otherwise, with that name), or (b) when subjects attribute the familiarity (correctly) to forgotten Session-1 exposure.[1] BW's model appears to be limited in its ability to model false fame experiments because, counter to a reasonable interpretation of that task, it includes no representation of paths that can lead from familiarity-without-recall to any judgment other than "famous."

---

[1] Occurrences of condition (a) of the preceding sentence would produce violations of BW's assumptions for both inclusion and exclusion conditions, and occurrences of condition (b) would produce violations of their assumptions for the exclusion condition.

How might the multinomial model demonstrate a criterion difference in assigning fame to familiarized male and female names? As explained in the preceding two paragraphs, BG supposed that familiarized nonfamous male and female names should have equal familiarity when presented in Session 2. What BG presumed to differ between male and female names was their likelihood of being judged famous once that state was achieved. In BW's model, familiarity-without-recall always leads to judgment of "famous" for old nonfamous names and therefore does not map onto the idea of a variable criterion for assigning fame to unrecalled-but-familiar-seeming male and female names. At the same time, the multinomial model's guessing parameters (g_i for guessing in the inclusion condition and g_e for the exclusion condition) may provide analogs to signal detection theory's concept of a variable response criterion. BW were able to test for differences in the g parameters associated with male and female names. The critical BG finding (their gender difference in criterion for familiarized male and female names) might be modeled as a gender difference in g for old *but not new* nonfamous names. Unfortunately, the structure of BW's multinomial model is such that g is obliged to be equal for old and new nonfamous names. Consequently, the intrinsic structure of the BW binary tree model precludes its providing a model of the critical BG finding.

Conclusion. The three points made in this comment indicate that Buchner and Wippich's (in press) methods and findings were focused on issues other than interpreting the gender bias observed by BG as being possibly unconscious in nature. We do not consider that Banaji and Greenwald (1995) provided ultimately conclusive evidence on the conscious versus unconscious nature of the gender stereotyping that they observed. Conceivably, more complex extensions of the measurement model developed by Buchner et al. (1995) will yet shed light on this interesting issue.

Postscript. In their following rejoinder, Buchner and Wippich suggest that subjects' opportunity to attribute Session-2 name familiarity to extra-experimental exposures can be safely ignored in the multinomial model of a false fame experiment (Buchner & Wippich, 1996b, p. 7 of preprint). However, in the typical word-list experiment for which the Buchner et al. (1995) model was developed, subjects know that they can attribute Session-2 familiarity of words only to (1) extra-experimental exposures and/or (2) Session-1 exposure. In false fame experiments, there is a third attribution opportunity — to (3) actual fame. BW modeled only attributions (2) and (3). However, even the relatively uncommon first and last names that they used must have had many unpaired extra-experimental exposures, much like low or moderate frequency words in the language. The multinomial model of a false fame experiment therefore needs to accommodate three types of attributions for Session-2 name familiarity, rather than only two.

References

Banaji, M. R., & Greenwald, A. G. (1995). Implicit gender stereotyping in judgments of fame. *Journal of Personality and Social Psychology, 68,* 181-198.

Buchner, A, Erdfelder, E., & Vaterrodt-Plᨆnnecke, B. (1995). Toward unbiased measurement of conscious and unconscious memory processes within the process dissociation framework. *Journal of Experimental Psychology: General, 124*, 137-160.

Buchner, A & Wippich, W. (1996).   Unconscious gender bias in fame judgments? *Consciousness and Cognition*, 5, XXX-XXX.

Buchner A. & Wippich, W. (1996b). Investigating fame judgments: On the generality of hypotheses, conclusions, and measurement models. *Consciousness and Cognition*, 5, XXX-XXX,

Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102*, 4-27.

Jacoby, L. L., Kelley, C., Brown, J., & Jasechko, J. (1989). Becoming famous overnight: Limits on the ability to avoid unconscious influences of the past. *Journal of Personality & Social Psychology, 56*, 326-338.