

---

Joachim Krueger (2000) Three Ways to get two Biases by Rejecting one Null. Psychology: 11(051) Social Bias (21)

**Volume:** 11 ([next](#), [prev](#)) **Issue:** 051 ([next](#), [prev](#)) **Article:** 21 ([next](#) [prev](#) [first](#)) **Alternate versions:** [ASCII Summary](#)

**Topic:** Social bias

[View Topic](#)

**Article:** 21) Krueger 11(051) Three Ways to get two Biases by

[View Article](#)

---

PSYCOLOQUY (ISSN 1055-0143) is sponsored by the American Psychological Association (APA).

---

# THREE WAYS TO GET TWO BIASES BY REJECTING ONE NULL

## Reply to Hertwig & Todd on Krueger on Social- Bias

*Joachim Krueger*

*Department of Psychology*

*Brown University, Box 1853*

*Providence, RI 02912*

*<http://www.brown.edu/Departments/Psychology/faculty/krueger.html>*

*[Joachim\\_Krueger@Brown.edu](mailto:Joachim_Krueger@Brown.edu)*

## Abstract

When the null hypothesis of rational responding is sandwiched between "biases to the left [and] fallacies to the right" (Hertwig & Todd, 2000), its cause is lost. Opposite biases can be "detected" in three different scenarios: (1) multiple studies performed in different research paradigms, (2) multiple studies performed in a single paradigm, and (3) single studies performed in a single paradigm. In the first two scenarios, moderator variables offer useful information that can take research beyond significance testing. The third scenario, however, harbors the deepest prejudice against demonstrations of human rationality. Here, regression artifacts are easily misinterpreted as evidence for the co-existence of opposite biases.

## Keywords

*Bayes' rule, bias, hypothesis testing, individual differences probability, rationality, significance testing, social cognition, statistical inference*

---

1. One scenario that produces apparently opposite biases involves different researchers working in

different paradigms. Hertwig & Todd (2000) observe that this happens quite frequently, and they offer Bayesian decision making (conservatism vs. base-rate fallacy) and perceptions of randomness (gambler's fallacy vs. the belief in the hot hand) as examples. In this scenario, not all claims of co-existing biases are incoherent because they can appeal to plausible moderator variables. In the case of Bayesian decision making, many studies have examined the degree to which people generalize the properties of observed instances to the population from which these instances were drawn. Subjects are asked, for example, how much a sample of colored marbles tells them about the contents of the urn from which the marbles were drawn. Typically, subjects respond conservatively, revising their estimates about the population less than Bayes's Theorem demands. In studies of base-rate neglect, subjects encounter vivid examples of particular instances, and they make judgments about the category membership of these instances rather than about the population to which these instances belong. Typically, estimates are 'representative' rather than 'conservative,' which means that they depend too much, rather than too little, on the unique properties of the observed instance.

2. In short, judgments are 'conservative' when they refer to the population and when sample information is pallid, and they are 'representative' when they refer to the instance itself and when that instance is vivid. In the case of perceptions of randomness, the gambler's fallacy occurs when people hold the false belief that random processes correct themselves in the short run. The hot hand fallacy occurs when people misperceive a process of chance as a process of skill. When there are multiple studies conducted in different research paradigms, as in these two examples, differences between study outcomes may point to the dependence of human judgment on important moderator variables. It is unwise to generalize the results of urn-and-marble studies to all forms of Bayesian decision making. Instead, it is desirable to examine the role of these moderators in the same research project.

3. The focus of the target article of the social-bias thread (Krueger 1998) was another scenario. Often, research conducted in a single paradigm is designed to yield evidence for either one of two contradictory biases. Research on self-perception is a case in point. There is no bias if self- and other-evaluations are the same. Any significant departure from this null hypothesis of no difference indicates either self-enhancement or self-deprecation. While the rejection of the null hypothesis alone does not say much, the identification of moderator variables can add important insights. Moderator variables predict the direction and the magnitude of the effect, and thus de-emphasize the value-laden contrast between rational and irrational thinking. Moderator variables may capture differences between tasks or between people. Task difficulty, for example, is a powerful predictor of the 'better-than-average' effect. Getting along with others is perceived to be easy, and most people believe they have more of this ability than others do. Acting is perceived to be difficult, however, and most people think their acting is worse than average (Krueger 1999). Narcissism is a person characteristic that predicts differences between self- and peer-ratings. Narcissists rate themselves more favorably than their peers rate them (Robins & John 1997). Hertwig & Todd note that some forms of probabilistic reasoning (e.g., the conjunction fallacy) are moderated by task and person characteristics. Taken together, these examples suggest that the search for moderator variables is a fruitful strategy. It promises to improve understanding of how differences between humans and tasks interact in producing judgmental outcomes. Narrow normative models make unrealistic assumptions about the homogeneity of people and tasks.

4. The third scenario is the most troubling. Although many times pronounced dead, the practice of extracting opposite biases from a single data set continues. Moderator variables offer little hope to those who seek to establish these biases. Instead, moderators help lay bare research artifacts. The general strategy in this scenario is to plot some social judgment against some personal characteristic. The judgment variable may be consensus estimates (in projection research), ratings of confidence (in calibration research), or one's perceived performance relative to others (in self-perception research). The personal characteristic may be its actual prevalence in the population (projection), the actual probability of being correct (calibration), or one's actual performance relative to others (self-perception). The data in

each of these three areas share two pervasive characteristics. First, there is a directional main effect. People tend to think that they are in the majority (projection), that they perform better than they actually do (overconfidence), and that they outperform most others (self-enhancement). Second, the performance variable never predicts the judgment variable perfectly. Thus, much of the 'evidence' for asymmetric biases reduces to regression artifacts.

5. Together, the main effect of a single directional bias and the imperfect correlation between judgment and reality guarantee the deceptive appearance of two opposite biases co-existing within the same sample. The subjects with the lowest actual scores produce large "overestimation errors," whereas the subjects with the highest actual scores produce comparatively small "underestimation errors." In the area of projection, this means that people who actually hold minority positions greatly overestimate the prevalence of their own positions, whereas people with majority positions slightly underestimate the prevalence of theirs. In the area of judgment calibration, this means that people who are often wrong are grossly overconfident, whereas people who are often right are slightly underconfident. The spurious nature of these apparent asymmetries has been thoroughly exposed (see Krueger & Clement 1997 on projection; Erev, Wallsten & Budescu 1994 on calibration; and Dawes & Mulford 1996 on both).

6. Still, the sufficiency of a simple model (with one source of systematic bias and one source of random error) is often overlooked. Instead, it is assumed that low scorers somehow reason differently (and more poorly) about themselves than high scorers do. Then differentiating psychological factors (i.e., moderators) are invoked to explain these apparent asymmetries. This strategy is not parsimonious; what is worse, it is misleading about the nature of social perception. Consider a recent study of self-perception. Kruger & Dunning (1999) gave subjects various tests (e.g., of grammar) and asked them to estimate their performance percentile. When estimated percentiles were plotted against actual percentiles, the results showed the typical pattern of (a) an overall self-enhancement (i.e., better-than-average) effect and (b) an imperfect correlation between estimated and actual percentile. What makes this article interesting, aside from the extensive media attention it received, [NOTE 1] is that the authors in one breath acknowledge and reject the regression phenomenon. "At first blush, the reader may point to the regression effect as an alternative interpretation of the results [but] despite the inevitability of the regression effect, we believe that the overestimation we observed was more psychological than artifactual" (p. 1124).

7. The regression effect is indeed inevitable (unless, of course, we discover perfectly correlated variables). The predictor variable is negatively correlated with the difference between itself and the criterion variable. Hence overestimation errors decrease as actual performance percentiles increase. "No [difference] scores ever need to be calculated to obtain this correlation" (McNemar 1969, p. 177). Consider a situation analogous to the psychology of over- and underestimation. We have known since Galton that the height of fathers (imperfectly) predicts the height of sons. Moreover, successive generations of sons have been taller. It follows that very short fathers beget (on average) sons that are much taller than they themselves are, and that very tall fathers beget (on average) sons that are a little shorter than they themselves are. Should we assume that very short fathers do anything in particular to sire sons that are much taller than themselves? [NOTE 2]

8. Kruger & Dunning do not consider this a rhetorical question. Low performers, they argue, grossly overestimate their standing relative to others not because of a regression effect but because they lack the "meta-cognitive skills" that enable people to comprehend their own failures. High performers are thought to possess such skills, although one wonders what they use them for. The first presumed skill is projection. High performers, but not low performers, are thought to assume that others do as well as they themselves do. No direct evidence is presented for the idea that projection is related to estimation errors. An argument could be made that projection affects the raw estimated percentiles, but that would mean that low rather than high performers project the most. The estimates of the low performers are closer to

the 50% mark, which is the point of greatest assumed similarity with others.

9. The second presumed meta-cognitive skill is the accuracy of self-knowledge on an item-by-item basis. Kruger & Dunning counted the test items for which a subject correctly predicted his or her own outcome (success or failure). This "skill" variable was related to actual performance, but this is hardly surprising. If people take the test in good faith, they must assume that their individual answers are most likely to be the correct ones (or they would not have given them). Thus, most subjects may be optimistic regarding their success on most (e.g., 8 out of 10) items regardless of their actual number of successes. In contrast, success on most test items (e.g., on 8) is the defining feature of high performers, while failure on most items is the defining feature of low performers. Even if subjects are completely unable to discriminate those test items on which they succeeded from those on which they failed, high performers predict more of their own outcomes correctly ( $M = 68\%$  given the present assumptions) than low performers do ( $M = 32\%$ ). These data cannot be used to charge poor test takers with social-perceptual ineptitude. Even at second blush then, regression effects explain overestimation (and underestimation) errors quite well.

10. Hertwig & Todd suggest that sometimes different heuristics explain the judgments of different people working on the same task. This view is constructive because it motivates the study of moderator variables (Greenwald, Pratkanis, Leippe & Baumgardner 1986). To be sure, moderator variables need to be carefully identified and cleansed of statistical artifacts. As the foregoing example showed, this does not always happen. I discussed the dual-bias hypothesis in self-perception in some detail because I consider it symptomatic for the generalized rush towards the discovery of new irrationalities. The burden of proof ought to be heavier, and judgments about human rationality should itself be more judicious and rational.

## NOTES

[1] On February 20, 2000, Garry Trudeau poked fun at a public figure by having a "Doonesbury" character say that "Now there's evidence that [...] incompetent people don't grasp their deficiencies [...] I've noticed that the hallmark of the self-absorbed and boorish is that they haven't a clue that they present as such. You following this, caller?" The reply is "No. Can we get back to me?"

[2] Perhaps very short men feed their sons particularly well to ensure growth, and very tall men dilute their sons' diet to stunt it. One would need two psychodynamic hypotheses to make sense of this (e.g., perhaps short fathers seek to compensate for perceived own "short" comings and tall fathers try to protect themselves against filial challenges).

## REFERENCES

Erev, I., Wallsten, T. S. & Budescu, D. V. (1994). Simultaneous over- and underconfidence: The role of error in judgment processes. *Psychological Review* 101: 519-527.

Dawes, R. M. & Mulford, M. (1996). The false consensus effect and overconfidence: Flaws in judgment, or flaws in how we study judgment? *Organizational Behavior and Human Decision Processes* 65: 201-211.

Greenwald, A. G., Pratkanis, A. R., Leippe, M. R. & Baumgardner, M. H. (1986). Under what conditions does theory obstruct research progress? *Psychological Review* 93: 216-229.

Hertwig, R. & Todd P. M. (2000). Biases to the left, fallacies to the right: stuck in the middle with null

hypothesis significance testing. PSYCOLOQUY 11(28).

<ftp://ftp.princeton.edu/pub/harnad/Psycology/2000.volume.11/psyc.00.11.028.social-bias.20.hertwig>  
<http://www.cogsci.soton.ac.uk/psyc-bin/newpsy?11.028>

Krueger, J. (1998). The bet on bias: A foregone conclusion? PSYCOLOQUY 9(46).

<ftp://ftp.princeton.edu/pub/harnad/Psycology/1998.volume.9/psyc.98.9.46.social-bias.1.krueger>  
<http://www.cogsci.soton.ac.uk/cgi/psyc/newpsy?9.46>

Krueger, J. & Clement, R. W. (1997). Consensus estimates by majorities and minorities: The case for social projection. *Personality and Social Psychology Review*, 1: 299-319.

Kruger, J. (1999). Lake Wobegon be gone! The "below-average effect" and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology* 77: 221-232.

Kruger, J. & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessment. *Journal of Personality and Social Psychology* 77: 1121-1134.

McNemar, Q. (1969). *Psychological statistics*. New York: Wiley.

Robins, Richard W. & John, O. P. (1997). Effects of visual perspective and narcissism on self-perception: Is seeing believing? *Psychological Science* 8: 37-42.

---

**Volume:** 11 ([next](#), [prev](#)) **Issue:** 051 ([next](#), [prev](#)) **Article:** 21 ([next](#) [prev](#) [first](#)) **Alternate versions:** [ASCII Summary](#)

**Topic:** Social bias

[View Topic](#)

**Article:** 21) Krueger 11(051) Three Ways to get two Biases by

[View Article](#)