The Bet on Bias: a Foregone Conclusion?

Krueger, Joachim (1998) The Bet on Bias: a Foregone Conclusion?, Psycoloquy: 9,#46 Social Bias (1)

Full text available as:
PsycDoc (Rendered)
ASCII
HTML (Printable)

Commentary/Response Threads

The Bet on Bias: a Foregone Conclusion?

Krueger, Joachim (1998) The Bet on Bias: a Foregone Conclusion?, Psycoloquy: 9,#46 Social Bias (1)

Authors

Krueger, Joachim
Department of Psychology
Brown University, Box 1853
Providence, RI 02912

Joachim_Krueger@Brown.edu

http://www.brown.edu/Departments/Psychology/faculty/krueger.hml

Abstract

Social psychology has painted a picture of human misbehavior and irrational thinking. For example, prominent social cognitive biases are said to distort consensus estimation, self perception, and causal attribution. The thesis of this target article is that the roots of this negativistic paradigm lie in the joint application of narrow normative theories and statistical testing methods designed to reject those theories. Suggestions for balancing the prevalent paradigm include (a) modifications to the ruling rituals of Null Hypothesis Significance Testing, (b) revisions of what is considered a normative response, and (c) increased emphasis on individual differences in judgment.

Commentary on:	Hertwig, Ralph and Todd, Peter M. (2000) Biases to the Left, Fallacies to the Right: Stuck in the Middle With Null Hypothesis Significance Testing, Psycoloquy: 11,#28 Social Bias (20)
Keywords:	Bayes' rule, bias, hypothesis testing, individual differences probability, rationality, significance testing, social cognition, statistical inference

I. THE DIM VIEW OF THE SOCIAL ANIMAL

1. Much of the social psychological research effort has concentrated on the shortcomings of everyday behavior and thinking. Although topics such as attraction, altruism, and accuracy in person perception have also been of interest, the bulk of the findings and the interpretation of these findings reflect a negativistic paradigm. Although this paradigm may have been inspired by the need to understand and address the psychological underpinnings of social problems, I suggest that these negative claims have been exaggerated, and that the reason for this lies in the joint operation of unrealistic expectations for acceptable performance and a data analytic philosophy that all but guarantees the demonstration of flawed behavior and thinking. Most theoretical criteria for optimal responding are narrow (and some are false) and most statistical testing procedures are inflexible and biased against the discovery of rational or optimal functioning. Several remedies are available, but no single methodology provides a satisfactory representation and evaluation of social behavior and thought. Instead, reliance on a diversity of theories and methods promises more robust progress.

2. During the classic period of social psychology, research on conformity, obedience, and discrimination

showed that average people can be induced to behave unethically (for a review see Cialdini 1993). The underlying model of good behavior stressed personal qualities such as autonomy and social responsibility. Many real people violated this model. This discrepancy between demanding theoretical standards and their empirical violation has powered, and biased, the social psychological enterprise. The psychological processes enabling people to resist destructive social pressures received little attention. The cognitive revolution shifted interest from social behavior to social perception. Specifically, it was the "heuristic and biases" paradigm in studies on judgment and decision making (Kahneman, Slovic & Tversky 1982) that served as a model for research on social perception (Gigerenzer & Murray 1987). As before, the tacit understanding was that real people would not meet ideal standards of rational reasoning. After being aggregated within groups, people's performance was compared with normative standards (i.e., no difference) derived from Fisher's statistical theory. The differences between actual and optimal performance itself were then tested for significance; and significance equalled bias. Of the many documented biases, three are modern classics: false consensus (for a review see Krueger 1998a), self enhancement (Armor & Taylor 1998), and overattribution (Gilbert 1998). [See also Koehler (1993) -- Ed.] ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1993.volume.4/psyc.93.4.49.base-rate.1.koehler

3. The first bias, FALSE CONSENSUS, is observed when the estimators' own responses predict their estimates of how prevalent these responses are in the population. Ross, Greene and House (1977) found that, on the average, these participants who volunteered for another study believed that volunteering was more common among their peers than did those who declined to volunteer. Estimates would have been considered unbiased only if participants had ignored their own responses. Like scientists, they should have judged the prevalence of the behavior only on the basis of a substantial data base. The second bias, SELF ENHANCEMENT, is observed when most participants think they are better than average. Brown (1986) found that participants viewed desirable traits as being more descriptive of themselves than of most others. This finding violates the demand that people should realize that on the average they are not better than average. The third bias, OVERATTRIBUTION, is observed when participants fail to explain behavior entirely with reference to situational factors when those factors are (ostensibly) sufficient. Jones and Harris (1967) found that essays that commended or condemned Castro's regime were attributed to pro- and anti-Castro authors, respectively, even when the essayists had been told which position to advocate. To the researchers, the power of the situation to shape behavior was obvious and therefore they expected participants to reject dispositional inferences as circular ("He praised Castro because he is a communist").

4. While the cognitive revolution generated interest in the flaws of social perception, the inference revolution provided the techniques to reveal them (Gigerenzer 1991). Performing the rituals of Null Hypothesis Significance Testing (NHST), investigators stake their substantive claims about bias on appeals to statistical significance. The theoretical notion of rational or unbiased reasoning assumes the feeble status of a point specific null hypothesis, whereas bias lies in any significant departure from this point. Participants have ample room to err, but only one place to be correct. Not surprisingly, NHST reveals that participants "significantly" miss the point of no bias. With this asymmetric testing, there is a growing conviction that people are cognitively limited or miserly. Investigators demonstrate bias by detecting it. They rarely attempt to detect rational judgment. In the typical study, the detectability of bias increases with improved apparatus, the application of robust statistics, and sheer statistical power. These methodological improvements can dignify even tiny effects with the predicate of statistical significance. With stronger studies, p values shrink and ever smaller biases can emerge [1].

5. The case for rational judgment is rather hopeless because the analytical asymmetry is bilateral. Areas of bias lie on both sides of the point of rationality. Consensus estimates are not only false when they are positively correlated with the raters' own responses, but also when the correlation is negative. In that case, the bias is the perception of false uniqueness (Klar 1996). Opposite biases also coexist in the area of self perception (i.e., enhancement vs. effacement, John & Robins 1994) and attribution (i.e.,

overattribution to dispositions vs. overattribution to situations, Quattrone 1982). The commonplace belief that null hypotheses cannot be proven implies that only irrationality, but not rationality, can be established. With the pursuit of rejections of false hypotheses, the dichotomy between rationality and irrationality is specious. It is tacitly accepted that people are irrational even before testing. The question is how sophisticated the equipment needs to be and how many respondents it takes to reject a belief that no one holds anyway. Because rationality is sandwiched between opposite biases, the data indicate irrationality no matter what the direction of the bias is.

II. MAKING INFERENCES ABOUT (IR)RATIONALITY FAIRER

6. Many critics of NHST have warned that the proliferation of confirmatory results stocks archives with irrelevancies [2], but the damage inflicted on the reputation of the social perceiver is greater than the archive metaphor suggests. Some findings are not only irrelevant, but are predetermined to indicate flawed perception and judgment. This need not be so, however, because statistics has many voices (Gigerenzer 1993). When voices other than Sir Ronald's are heard, the psychological record of human folly may look less gloomy. Consider the following modifications to the standard demonstrational testing procedures.

7. As a subtraction problem has one correct solution, all others are wrong no matter how close they are to the solution. Similarly, there are many ways to err in social perception but only one way to be right. It is impossible to invert the common paradigm by identifying bias with a specific null hypothesis while considering all other outcomes unbiased. When errors are bound to occur, it is useful to ask what they reveal about underlying mental processes. Attributing incorrect responses to incompetence and staking the demonstration of errors on their statistical significance is not illuminating. Preferably, a lawful though incorrect function is found that explains what raters are doing. It is more informative, for example, to know that people represent exponential growth as linear growth than it is to know that they significantly underestimate it (Wagenaar & Sagaria 1975).

8. Ambiguities arise when the null hypothesis is not rejected. The statistical meaning of the null hypothesis hinges on the concept of chance. By negating causation, chance eliminates explanation. If the null hypothesis is true, the data contain only error variance. In contrast, the theoretical meaning of the null hypothesis is rational reasoning. This theoretical view can never be brought to bear, however, because one cannot simultaneously conclude that judgments are rational and random. Unsurprisingly, not many mental mechanisms have been suggested to explain biases of consensus estimation, self perception, and attribution, but few to explain the absence of bias. In other words, the confound between rationality and chance limits the development of theories of rationality, whereas theories of bias prosper. To complicate matters, the null hypothesis seems spuriously true when two biases cancel each other out. Consensus estimates for outgroups are a case in point. Here, the size of the bias hovers around zero. NHST does not us tell whether respondents think rationally about outgroups, whether they have no idea what they are doing, or whether opposite biases of consensus and uniqueness mask each other. To distinguish these possibilities, it is necessary to examine the variability of the responses. Rational responding implies a lower variance than the presence of opposite errors. Another strategy is to force respondents to choose between a rational response and an irrational alternative. According to the null hypothesis of random responding, 50% of the choices will be rational, whereas rational and irrational responding are distinct alternatives (100% and 0% rational responding, respectively). A third strategy is to have respondents estimate the odds of a chance event to occur. When they underestimate these odds regardless of their true value (as they do, for example, in the birthday problem), the evidence for bias is quite strong.

9. NHST is asymmetric because there is no specific alternative hypothesis. In the Neyman and Pearson framework of data analysis, investigators specify beforehand what they consider to be a relevant and reasonable effect size. If a precise research hypothesis is available, there is no self fulfilling promise of finding bias on either side of the null point. Moreover, improvements in apparatus, computation or power now raise the quality of statistical inference. The research hypothesis is now exposed to the danger of being rejected. The asymmetry between making inferences about rationality and irrationality disappears if the probability of rejecting a true null hypothesis (Type I error) and accepting a false one (Type II error) are the same. This equality minimizes the sum of the expected errors. The disadvantage of this method is that it requires (a) a justification for the choice of the alternative hypothesis and (b) consensus among investigators about this justification.

10. When multiple null hypotheses vie for rejection, NHST creates a dilemma. Consider studies designed to eliminate bias. In the debiasing condition, the bias may be significant but smaller than in a control condition. If this happens, the same effect size can support conflicting conclusions. Improvements in apparatus, computation, or power worsen this dilemma. Moreover, the interpretations of significant bias and debiasing are probably asymmetric. Whereas bias signals a categorical difference between two modes of thinking (one good and one bad), debiasing merely signals the reduction of a bad thing. The value of debiasing studies is that it allows comparisons between effect sizes. Consider an example from consensus estimation. Marks and Duval (1991) asked participants to imagine themselves engaging in either their favorite recreational activity or a less desirable alternative. The first group showed greater consensus bias than the second. The latter effect was still significant, but about 1/3 the size of the former. Similar effect ratios are found when self enhancement or attitudinal attributions are debiased (Alicke, Klotz, Breitenbecher, Yurak & Vredenburg 1995; Jones, Riggs & Quattrone 1979).

11. Dawes (1997) noted that when we make statistical inferences, "we often act as if we have solved the problem of induction" (p. 387), and that NHST is "the best available" tool (p. 388). His example, however, is fundamentally different from the "sandwich scenario" in bias research. When the quality of clinical judgment (i.e., social perception) is assessed, actuarial predictions serve as a normative standard that can, in theory, be surpassed. Thus, the null hypothesis of no difference is the upper bound for one directional hypothesis and the lower bound for its opposite. The inferiority of clinical judgment impresses itself upon the reviewers because there are far fewer positive effects than negative effects (Grove & Meehl 1996). But note that NHST is irrelevant for this practical conclusion. A single study does not decide the issue, but dozens of effects of the same direction do, even if some or many of them are not "significant" individually.

III. HANDLING HYPOTHESES: FROM "REJECT" TO "REVISE AND RESUBMIT"

12. These methodological modifications can advance research beyond ritualistic applications of NHST, but they fall short in two respects. Some of them require information (e.g., predicted effect sizes) that may not be available, and none of them reveals the probability that a hypothesis (null or other) is true given the data. Both these problems can be addressed in a Bayesian framework. Here, multiple hypotheses are considered (Fischhoff & Beyth Marom 1983). One of these might as well be the traditional null hypothesis of no effect. This H0 is contrasted with (at least) one research hypothesis, H1, so that the two are mutually exclusive and exhaustive. When investigators fail to agree on the location of H1, they can identify the observed effect with H1. This pragmatic choice is agnostic in that no prediction of an effect size is derived from theory. When the sampling distribution of the observed data is centered around H0 and H1, the p value, p(D|H0), provided by NHST is the portion of the H0 distribution lying beyond the center of the H1 distribution. Because the research hypothesis is located post hoc, the probability of the data given that hypothesis, p(D|H1), is always .5. Thus, p(D|H0) varies

across studies but p(D|H1) does not, and the former is always smaller than the latter. In other words, post hoc research hypotheses are variable and data are fixed so that the data, by definition, lend greatest support to the hypothesis designating the observed effect (Rorer 1991).

13. The key advantage of Bayes' Rule is to provide what NHST cannot, namely the probability that the null hypothesis is true given the data, p(H0|D), and its complement, the probability that the research hypothesis is true given the data, p(H1|D). These posterior probabilities depend not only on the obtained data but also on the prior probabilities of the hypotheses. Specifically, the ratio of the two posterior probabilities is equal to the ratio of their inverse conditional probabilities multiplied by the ratio of the prior probabilities of the two hypotheses, i.e.,

$$p(H0|D)/p(H1|D) = (p(D|H0)/p(D|H1))X(p(H0)/p(H1)).$$

The posterior probability of the null hypothesis given the data is

$$p(H0|D) = (p(D|H0)p(H0))/(p(D|H0)p(H0)+p(D|H1)p(H1)).$$

Bayesian analyses have not been popular in social psychology because the assignment of priors smacks of subjectivity. If investigators set priors whimsically, they can stack the deck against any hypothesis they oppose. Inasmuch as investigators disagree on the priors, they cannot agree on what the data mean. To put Bayes to work for a community of investigators, consensus on the priors is essential. A modest method of specifying priors is to profess ignorance. An ignorant investigator considers contending hypotheses equally likely. Uniform priors make sense if there is no pertinent research on the question of interest and if there are no theoretical reasons to favor one hypothesis. Once a phenomenon has been recorded, however, its effect size can become the research hypothesis for replication studies (Greenwald, Gonzalez, Harris & Guthrie 1996). After the first study, the selection of H1 is no longer agnostic, but driven by knowledge. When a particular bias has been replicated many times (e.g., consensus), isolated reports of the opposite (i.e., uniqueness) do not make that opposite likely, although p may be at .05 in an individual study. In other words, the Bayesian approach reduces the bilateral asymmetry of bias research [3].

IV. REVISITING THE POINT OF RATIONALITY

14. The foregoing recommendations are intended to give hypotheses of rationality and irrationality a roughly equal shot of being demonstrated. Changing habits in statistical analysis alone, however, are not sufficient to remove exaggerated claims of mental malfunctioning. It is necessary to take a closer look at the interpretation of obtained effect sizes. Does the lack of a difference indeed represent rational reasoning? To address these questions, the three social cognitive biases are reexamined.

15. In consensus estimation, the criterion of no difference is the wrong benchmark for rational predictions. People's own responses are by definition more likely to be the responses of the majority than the responses of the minority. Therefore, one's own responses should be used for consensus estimation to minimize errors. This strategy leads to results that look like consensus bias, although it is based on normative inductive reasoning. People may ignore their own responses only when they are well informed about the actual consensus in the group. In this case, attaining accuracy is equivalent to avoiding bias. But what is the optimal or normative difference when people have little relevant knowledge? A radical solution is to assume that people are ignorant about actual consensus. Like researchers breaking into a new area, social perceivers can invoke the principle of indifference. Then, when they consider the first empirical observation (i.e., their own response), their estimates for the prevalence of that response should differ by 33% depending on their own response (i.e., perceivers

expect that 2/3 of all others agree with them; see Dawes 1989; or Krueger & Clement 1996 for derivations) [4]. This difference does not represent the only acceptable norm for consensus estimation, but its upper bound. As people acquire information about how others respond, they should also treat that information as sample observations. The more information they have, the more their estimates should reduce bias and increase accuracy. Unfortunately, it is difficult to know how much participants know. One response to this difficulty is to accept differences between 0% and 33% as good enough. Another question is whether the responses of other individuals receive as much weight as one's own. It seems that they do not, and that consensus estimates are indeed egocentrically biased (Krueger & Clement 1994).

15. In self perception, it is NOT "logically impossible for most people to be better than the average person" (Taylor & Brown 1988, p. 195). If it were, the self/other difference would indicate both enhancement bias and inaccuracy. But most people are indeed better than average when the distributions of their trait or performance scores are negatively skewed. Among college students, for example, high self esteem and good grades are more common than low scores and poor grades. If respondents are aware of the skewness in these distributions and if they have little valid knowledge about their own scores, a positive bias is justified. Like consensus bias, self enhancement guarantees some errors, but it increases accuracy. The self enhancer who consistently expects to be better than average will be right more often than the agnostic who predicts by the flip of a coin. Bias and accuracy can be analytically separated in a signal detection framework. This approach construes self enhancement as a low threshold for rating the self favorably. Inaccuracy, on the other hand, is the inability to discriminate between those features that describe the self and those that do not.

17. In attribution research, it is customary to assume that situational causes (e.g., the experimenter's instructions to write a "pro" essay) are fully sufficient to explain the observed behavior (i.e., the writing of a "pro" essay). After proper discounting, the inferred attitude should revert to the prior probability of the attitude. The inferred attitude should be whatever it seems to be before individuating information is available. Morris and Larrick (1995) suspected that social perceivers do not share the view that the situational causes are always fully sufficient explanations of behavior. They may doubt that all target persons comply with the experimenter. Such skepticism would be healthy given the imperfections of compliance techniques, which social psychologists have so richly documented. The question is whether participants discount dispositional causes too little given their own probabilistic understanding of the sufficiency of the candidate causes. Morris and Larrick replicated the Castro study but elicited judgments as probability estimates. Instructions insured that the target person's attitudes were perceived as independent of the writing requests the person received. Neither of these potential causes of writing a certain type of essay (e.g., pro) was perceived to be fully sufficient. Therefore, the optimal posterior probability of a pro attitude, given that the target person had agreed to write a pro essay, was higher than the prior probability of a pro attitude. This difference was compared with the difference between estimated prior and posterior probabilities. The assessment of bias thus became a question of a difference between differences.

18. In all three areas, the revised criterion for no bias moves toward the effect that appears most often (consensus, enhancement, dispositionism). Thus, effect sizes shrink (or even reverse) upon reanalysis. On the one hand, this revision accentuates the problems of NHST having to do with the detection of small effects. Just how much more can one's own responses, for example, be weighted in consensus estimation relative to the responses of others before investigators can infer an egocentric bias? On the other hand, the null hypothesis is no longer a "nil" hypothesis, and therefore the confound between chance variation and rational reasoning is gone.

V. AN IDIOGRAPHIC APPROACH TO ACCURACY AND

BIAS

19. However small a significant difference between normative criteria and average judgments might be, the standard interpretation is that "people" are biased. Individual differences in judgment are considered statistical error. NHST encourages the reduction of this error through an increase in the number of perceivers. The more perceivers there are, the more it will seem that they are biased. But if all are indeed biased, then it should be of interest to demonstrate this bias within individuals. To do this, it is necessary to sample not only individuals but also judgment items. When they use multiple items, investigators can ask how well perceptions predict external criteria (accuracy), and how consistent perceptions are internally (bias). In consensus estimation, correlations between endorsements and actual consensus indicate accuracy, whereas correlations between item endorsements and estimated consensus indicate bias (Krueger & Zeiger 1993). In self perception, correlations with objective criteria or aggregated peer ratings indicate accuracy, whereas correlations between self ratings and desirability ratings indicate bias (Krueger 1998b). In attitude attribution, correlations between estimates and external criteria indicate accuracy (Wright & Drinkwater 1997), whereas correlations between probability estimates and their Bayesian counterparts indicate bias (Krueger 1996). The idiographic separation of accuracy and bias then makes it possible to study how they are related to each other (e.g., does bias increase accuracy?) and to other variables of theoretical interest (e.g., criteria with adaptive significance).

20. The idiographic approach controls the prejudice against the detection of rationality. Whereas in the study of bias, nonzero correlations indicate departures from rationality, the opposite is true in the study of accuracy. Here, (positive) correlations indicate detectable levels of rationality. This is not to say that all is well if significant levels of both accuracy and bias are detected. Instead, it can be hoped that meaningful (coherent or functional) patterns of associations will replace isolated significances as the stories that need to be told (Sinha & Krueger 1998).

VI. CONCLUSION: DIVERSIFY AND MERGE

21. The study of social perceptual biases is important and interesting, but the proliferation of demonstrations within NHST hardly improves our understanding of how people manage to attain serviceable perceptions of their social world. With increased sensitivity to alternative statistical approaches, a more realistic portrait of the social perceiver may emerge. The use of diverse statistical methods may at last gain ground since it has been endorsed by the "Task Force on Statistical Inference" (American Psychological Association 1996). Methodological improvements might work particularly well in conjunction with interest in alternative theoretical perspectives, such as ecological or evolutionary approaches to social perception. Some investigators believe that social perception needs to be adaptive rather than rational (Cosmides & Tooby 1996; Gigerenzer & Goldstein 1996). From this perspective, the question concerns how social perception works so well rather than whether it fails significantly. Much like the differences in method, differences in theory need not be mutually exclusive (Hammond 1990). I do not advocate the wholesale replacement of mathematical standards of rationality with ecological standards of adaptiveness. Preliminary efforts to integrate these two perspectives have been made, although their ultimate success remains to be seen (see Krueger 1998c for an integrative approach to the study of consensus estimation).

FOOTNOTES:

[1] This is not to say that all reported biases are small in size. As one Psycoloquy referee pointed out, many of the biases reported by Kahneman and Tversky in the 1970s had enormous effect sizes. When effects are that large, NHST is unnecessary; when they are small (as they are in many social

psychological studies), the present concerns apply.

[2] For more pros and cons of NHST see Hagen's (1997) target article and the commentaries on it (1998); or the forum edited by Harris (1997). [See also Chow (1998). -- Ed.] ftp://ftp.princeton.edu/pub/harnad/BBS/.WWW/bbs.chow.html

[3] This approach favors the accommodation of theory to data, but it does not permit the categorical rejection of predictions (Kerr 1998).

[4] The size of this effect varies when the priors are not uniform, but posterior estimates will always covary with the data (i.e., one's own response).

REFERENCES

Alicke, M. D., Klotz, M. L., Breitenbecher, D. L., Yurak, T. J. & Vredenburg, D. (1995) Personal contact, individuation, and the better- than-average-effect. Journal of Personality and Social Psychology 68: 213-226.

American Psychological Association (1996) Task force on statistical inference initial report. www.apa.org/science/tfsi.html

Armor, D. A. & Taylor, S. E. (1998) Situated optimism: Specific outcome expectancies and self regulation. Advances in Experimental Social Psychology 30: 309-379.

Brown, J. D. (1986) Evaluations of self and others: Self enhancement biases in social judgments. Social Cognition 4: 353- 376. Cialdini, R. B. (1993) Influence. New York: Quill.

Chow, S. L. (1998) Multiple Book Review of "Statistical Significance: Rationale, Validity and Utility." Behavioral and Brain Sciences 21: 169-240. ftp://ftp.princeton.edu/pub/harnad/BBS/.WWW/bbs.chow.html

Cosmides, L. & Tooby, J. (1996) Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment and uncertainty. Cognition 58: 1-73.

Dawes, R. M. (1989) Statistical criteria for a truly false consensus effect. Journal of Experimental Social Psychology 25: 1-17.

Dawes, R. M. (1997) Qualitative consistency masquerading as quantitative fit. In M. L. Dalla Chiara et al. (Eds.), Structures and norms in science (pp. 387-394). Kluwer Academic Publishers.

Fischhoff, B. & Beyth-Marom, R. (1983) Hypothesis evaluation from a Bayesian perspective. Psychological Bulletin 90: 239-260.

Gigerenzer, G. (1991) From tools to theories: A heuristic of discovery in cognitive psychology. Psychological Review 98: 254-267.

Gigerenzer, G. (1993) The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), A handbook for data analysis in the behavioral sciences: Methodological issues (pp. 311- 339). Hillsdale, NJ: Erlbaum.

Gigerenzer, G. & Goldstein, D. G. (1996) Reasoning the fast and frugal way: Models of bounded rationality. Psychological Review 103: 650-669.

Gigerenzer, G. & Murray, D. J. (1987) Cognition as intuitive statistics. Hillsdale, NJ: Erlbaum.

Gilbert, D. T. (1998) Ordinary personology. In D. T. Gilbert, S. T. Fiske & G. Lindzey (Eds.) Handbook of social psychology (4th ed., pp. 89-150). Boston: McGraw Hill.

Greenwald, A. G., Gonzalez, R., Harris, R. J. & Guthrie, D. (1996) Effect sizes and p values: What should be reported and what should be replicated? Psychophysiology 33: 175-183.

Grove, W. M. & Meehl, P. E. (1996) Comparative efficiency of formal (mechanical, algorithmic) and informal (subjective, impressionistic) prediction procedures: The clinical-statistical controversy. Psychology, Public Policy, and Law 2: 293-323.

Hagen, R. L. (1997) In praise of the null hypothesis statistical test. American Psychologist 97: 15-24. Commentaries in American Psychologist 98: 796-803.

Hammond, K. R. (1990) Functionalism and illusionism: Can integration be usefully achieved? In R. Hogarth (Ed.), Insights in decision making: A tribute to Hillel J. Einhorn (pp. 227-261). Chicago, IL: University of Chicago Press.

Harris, R. L. (1997) Ban the significance test? Psychological Science 8: 1-20.

John, O. P. & Robins R. W. (1994) Accuracy and bias in self- perception: Individual differences in self-enhancement and the role of narcissism. Journal of Personality and Social Psychology 66: 206-219.

Jones, E. E. & Harris, V. A. (1967) The attribution of attitudes. Journal of Experimental Social Psychology 3: 1-24.

Jones, E. E., Riggs, J. M. & Quattrone, G. A. (1979) Observer bias in the attitude attribution paradigm: Effect of time and information order. Journal of Personality and Social Psychology 37: 1230-1238.

Kahneman, D., Slovic, P. & Tversky, A. (1982) Judgment under uncertainty: Heuristics and biases. Cambridge: Cambridge University Press.

Kerr, N. L. (1998) HARKing: Hypothesizing After the Results are Known. Personality and Social Psychology Review 2: 196-217.

Klar, Y. (1996) Dysphoria and consensus estimates for behavioral choices: Equally inaccurate but in opposite directions. Journal of Research in Personality 30: 278-289.

Koehler, J. J. (1993) The Base Rate Fallacy Myth. PSYCOLOQUY 4(49)
ftp://ftp.princeton.edu/pub/harnad/Psycoloquy/1993.volume.4/psyc.93.4.49.base-rate.1.koehler
http://www.cogsci.soton.ac.uk/bbs/Archive/bbs.koehler.html

Krueger, J. (1996) Probabilistic national stereotypes. European Journal of Social Psychology 26: 961-980.

Krueger, J. (1998a) On the perception of social consensus. Advances in Experimental Social Psychology

30: 163-240.

Krueger, J. (1998b) Enhancement bias in descriptions of self and others. Personality and Social Psychology Bulletin 24: 505-516.

Krueger, J. (1998c) The projective perception of the social world: Adaptive but egocentric. In J. Suls & L. Wheeler (Eds.), Handbook of social comparison. New York: Plenum.

Krueger, J. & Clement, R. W. (1994) The truly false consensus effect: An ineradicable and egocentric bias in social perception. Journal of Personality and Social Psychology 67: 596-610.

Krueger, J. & Zeiger, J. S. (1993) Social categorization and the truly false consensus effect. Journal of Personality and Social Psychology 65: 670-680.

Marks, G. & Duval, S. (1991) Availability of alternative positions and estimates of consensus. British Journal of Social Psychology 30: 179- 183.

Morris, M. W. & Larrick, R. P. (1995) When one cause casts doubt on another: A normative analysis of discounting in causal attribution. Psychological Review 102: 331-355.

Quattrone, G. (1982) Overattribution and unit formation: When behavior engulfs the person. Journal of Personality and Social Psychology 42: 593-607.

Rorer, L. G. (1991) Some myths of science in psychology. In D. Cicchetti & W. M. Grove (Eds.), Thinking clearly about psychology (Vol. 1, pp. 61-87). Minneapolis: University of Minnesota Press.

Ross, L., Greene, D. & House, P. (1977) The false consensus effect: An egocentric bias in social perception and attribution processes. Journal of Experimental Social Psychology 13: 279-301.

Sinha, R. R. & Krueger, J. (1998) Idiographic self-evaluation and bias. Journal of Research in Personality 32: 131-155.

Taylor, S. E. & Brown, J. D. (1988) Illusion and well-being: A social-psychological perspective on mental health. Psychological Bulletin 103: 193-210.

Wagenaar, W. A. & Sagaria, S. D. (1975) Misperception of exponential growth. Perception and Psychophysics 18: 416-422.

Wright, J. C. & Drinkwater, M. (1997) Rationality vs. accuracy of social judgment. Social Cognition 15: 245-273.

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

Contact site administrator at: moj199@ecs.soton.ac.uk