Ethical Human Sciences and Services, Vol. 3, No. 1, 2001

Better Never Than Late: Peer Review and the Preservation of Prejudice

Peter H. Schonemann, PhD
Purdue University, West Lafayette, IN

This article documents some difficulties authors face who challenge faulty research claims published in mainstream literature. Editors of "reputable journals" may react with stonewalling tactics that tend to enshrine these faulty results. A case in point is the mental test literature, which has long been beset with racist myths. In 1985, Arthur Jensen added a new myth, his "Spearman Hypothesis," which asserts that a positive correlation between White/Black means differences in scores on mental tests and the loadings of the first principal component confirms the existence of a general intelligence factor ("g"). It can be shown by mathematical and geometric deduction, by computer simulation, and by reference to "real data," including Jensen's own, that the assertion is unwarranted, and that the relationship Jensen observed is an artifact that has nothing to do with ethnicity or "g." Nevertheless, it proved impossible for more than 12 years to record this challenge to Jensen's claims in any of the leading journals in psychology and statistics. Typically, their editors invoked arguments having nothing to do with the fundamental question of whether Jensen's claims are true or false. It is concluded that, in view of the transparent racist implications of these claims, such editorial policies—regardless of their motivations—contribute to the preservation of ethnic stereotypes and prejudice.

> An understanding of the way science really works is important because the process does not take place in a vacuum. Scientists are part of society. What they do and how they behave affect the general public in a way perhaps more profound than any other profession. (Broad & Wade, 1982, p. 180)

> Another significant impact of fraud has been the unhappy field of measuring human abilities. Fraud and self-deception have played major parts in studies that have influenced public attitudes on matters of class and race, and in shaping public action on issues such as immigration and education. (Broad &Wade, 1982, p. 219)

In 1988, the American Medical Association sponsored the First International Congress on Peer Review in Biomedical Publication. Reasonable people have little trouble appreciating the risks of promoting, say, a noxious drug in the name of science. However, many people seem to be more sanguine about the dangers of promoting a noxious ideology, for example, racism, in the name of science. Though perhaps less obvious, these dangers are no less real than those ensuing from the irresponsible promotion of noxious drugs: indeed, according to Robitscher (1973, p. 122), between 1921 and 1964, 33,374 Americans were sterilized against their will because their IQ test scores were deemed too low.

As Blum (1978) noted, the history of mental testing is rich in racist rhetoric all the way back to Sir Francis Galton: "An index of psychologists' respect for eugenics is the fact that Hall, Cattell, Yerkes, Terman, Thorndike, and Woodworth all became presidents of the American Psychological Association" (p. 57). Since Galton, articles of eugenic faith have resurfaced in various disguises, most recently in *The Bell Curve* by Herrnstein and Murray (1994). These writings put forward the same three points: (a) "intelligence" can be "measured" (though, perhaps, not *defined*), (b) "intelligence" is highly heritable, and (c), most notorious of all, ". . . black people, statistically, will have a greater handicap in those educational, occupational, and military assignments that are most likely correlated with measures of general intelligence" (Jensen, 1985, p. 206).

The evidence supporting these tenets is just as flimsy today as it was in Galton's days more than a century ago. Occasionally during this time, advocates of this eugenic position endured some criticism. Though well intended, not all of it was technically well founded. Moreover, most of it focused on especially salient instances—such as those involving Cyril Burt, Philip Rushton, or *The Bell Curve*—treating them in isolation as if they were pathological mutations. This narrow perspective may have obscured the deeper problem of how faulty claims could have survived for so long in the mainstream literature, given that scientific articles are subject to peer review. In other words, if the peer review system functioned the way it is supposed to, by identifying instances of faulty logic or poor science, faulty eugenic claims would not continue to appear in the scientific literature.

That a potential connection between dysfunctional peer review and the growth of racist thinking should have been overlooked for so long is somewhat surprising since in hindsight it seems rather obvious: if the review system systematically shields racist claims from valid criticisms, then these claims will eventually acquire scientific respectability. Over the long haul, this will not just affect the proverbial man in the street, but also influential decision makers, including members of Congress, whose business it is to issue legislation that potentially affects the welfare of millions. An example are the racist immigration laws Congress passed in the 1920s. Though psychologists understandably try to downplay their role in these developments, it strains common sense to completely absolve the American Psychological Association establishment from blame for its original enthusiastic endorsement of Brigham's (1923) flawed analyses of the World War I army test data.

Criticisms of the peer review system are of course not new (see, for example, Peters & Ceci, 1982). In 1988, the 100th Congress of the United States held hearings on the broader topic of "scientific fraud and misconduct" (Weiss Hearings, 1988). One witness noted: "Our scientific journals have a central role to play in the dissemination of scientific information. It is appealing to think that they might also play an important role in the correction and cleansing of the scientific literature." The witness continued: "Unfortunately, my experiences in this regard do not suggest that conclusion" (Weiss Hearings, 1988, p. 74).

However, in the past, most discussions critical of the peer review system have been couched in general terms. While it is usually acknowledged that things are not perfect—what is?—it is left to the imagination to wonder just how defective they really are. For example, Sir Cyril Burt was allowed to publish the exact same twin correlations (.771) for decades, although the sample sizes kept constantly changing. It seems reasonable to ask, as Broad and Wade (1982) did, "Why did scholars reading his papers not spot the flaws?" (p. 209). As I have suggested elsewhere, one answer might be that some scholars actually *did* spot the flaws, but were prevented from announcing their discovery in the mainstream journals because the editors deemed such revelations "inappropriate" or "unsuitable."

The main purpose of the present article is, therefore, to document that on some issues of transparent and far-reaching social relevance it has become virtually impossible to correct published errors in mainstream scientific journals, as long as the official gatekeepers, for whatever reasons, deem such corrections inopportune. As will become apparent, the tactics these editors employ to suppress inconvenient truths vary surprisingly little across a broad spectrum of "respectable" journals, and their arguments bear no resemblance to scientific discourse as it is commonly understood.

Specifically, this article chronicles 12 years of editorial rebuffs of attempts to refute a bold but faulty theory that Professor Arthur Jensen has untiringly and successfully promoted for several decades. This theory, which he dubbed "Spearman Hypothesis," posits a connection between a presumed genetically predetermined "general intelligence" (g) and Black/White differences on mental tests. Thus, it clearly has racist implications.

There is no longer any room for doubt that Jensen's claims are quite simply false. However, more to the point here is the fact that this had been firmly established more than 15 years ago. Through all this time, it proved impossible to refute Jensen's claims in any of the American mainstream outlets, not just in psychology but also in statistics. More important, journal editors in both fields rarely bothered with the fundamental question of whether the disputed claims were objectively true or false. Instead of concerning themselves with objective standards of truth, they usually invoked subjective standards of personal taste. Recurrent themes of their arguments were elusive notions such as "appropriateness" and "suitability." In particular, they seemed blissfully unconcerned about the long-range social consequences of their stonewalling tactics, just as if they had never heard of the sterilization or the immigration laws of the 1920s.

Spearman's Hypothesis

> We cannot blame a half-educated crank for believing his own theories and trying to get them accepted: we must ask what was wrong with a whole scientific community that it allowed itself to be captured by such crazy notions. (Broad & Wade, 1982, p. 192)

> The failure of the scientific method to afford protection against prejudice is particularly disconcerting when the failure extends beyond an individual to the colleagues in his own discipline . . . (Broad & Wade, 1982, p. 202f)

In his book, *Bias in Mental Testing*, Arthur Jensen introduced "Spearman's interesting hypothesis that the magnitude of White/Black mean differences on various mental traits are directly related to the test's g-loadings" (Spearman, 1927, p. 379; Jensen, 1980, p. 535). He subsequently published reanalyses of 11 mental test batteries in a target article for *The Behavioral and Brain Sciences*. The results seemed to confirm Spearman's hypothesis:

> Eleven large scale studies . . . show a significant and substantial correlation between the test's g loadings and the mean Black/White difference . . . on the various tests. Hence, in accord with Spearman's hypothesis, the average Black/White difference on diverse mental tests may be interpreted as chiefly a difference in g . . . (Jensen, 1985, p. 193)

These claims were warmly welcomed in the mental test community. To cite but one example: "Arthur Jensen's reanalysis of data from 11 studies provides convincing evidence that the observed differences between average scores of Black and White samples in the United States in a variety of mental tests are directly related to average differences in g" (Jones, 1985, p. 233; for many more such quotes, see Schonemann, 1997a, p. 666f).

Jensen's claims about Spearman's hypothesis closely interface with the equally specious heritability lore of "intelligence" (e.g., Schonemann, 1997c)—a point not lost on prominent spokesmen of the "cognitive super-elite." "Another line of evidence pointing toward a genetic factor in cognitive ethnic differences is that Blacks and Whites differ most on the tests that are the best measures of g, or general intelligence" (Herrnstein & Murray, 1994, p. 270).

Jensen, to his credit, left no doubt about *his* assessment of the significance of his presumed discovery: "An important practical implication of Spearman's hypothesis [is that] Black people, statistically, will have a greater handicap in those educational, occupational, and military assignments that are most likely correlated with measures of general intelligence. . . . The practical implications of g and Spearman's hypothesis for employment, productivity, and the nation's economic welfare have been discussed in more detail elsewhere" (Jensen, 1985, p. 206). Thus, we are clearly dealing with a non-trivial topic.

CHRONOLOGY OF EFFORTS TO CORRECT JENSEN'S FALSE CLAIMS ABOUT SPEARMAN'S HYPOTHESIS

> Peer review . . . entails deference to specialized, expert knowledge; therefore it is anti-democratic and often plagued . . . by cronyism, elitism, and conflict of interest. (Chubin & Hacket, 1990, p. 188)

> If no one points out errors, then there are no errors. (Bo Yang, 1992, p. 21)

1984

I first became aware of Spearman's Hypothesis when the editor of *The Behavioral and Brain Sciences* (BBS) asked me to review Jensen's (1985) target article. In view of its far-reaching implications, I recommended that the article be published to expose it to broader peer commentary. I suspected his results might be artifactual but was unable to prove this at the time.

1985

After the article had been accepted, the editor of BBS invited me to comment on it. By that time, I had arrived at an argument which explains Jensen's positive correlations between the mean vectors and the first principal component (PC1, see Note) of the *pooled* group as an artifact (Level I argument). I demonstrated this with a simple computer simulation (Schonemann, 1985). Of the 29 commentaries on Jensen's target paper, only two raised doubts about the validity of Jensen's claims. I sent a reprint to Louis Guttman in Israel.

I also reread Jensen's earlier formulations of Spearman's Hypothesis in his *Bias in Mental Testing* (1980) more carefully.

1986

In April, Guttman submitted an article, "The Irrelevance of Factor Analysis for the Study of Group Differences," to BBS as a Continuing Commentary. The editor rejected it. I saw this article for the first time in the fall of 1987 when I met Guttman in Bloomington, Indiana, shortly before he died.

In November 1986, William Shockley called and told me that he thought he had found an error in my (1985) BBS commentary. He asked my permission to tape our conversation, which I agreed to. His main point was that my 1985 BBS argument only addressed the PC1 of the *pooled* group (Level I case), while Jensen's (1985) data, with one exception, showed positive correlations of the mean vector with the PC1s in *both* (Black and White) subgroups (Level II case). I promised to look into this problem and get back to him. After a few hectic days at the office, I succeeded in extending my geometric argument from Level I to Level II by adjoining two additional assumptions: (a) multinormality for the pooled group, and (b) positivity of both within covariance matrices.

The intuitive content of these assumptions is really quite simple: Suppose one cuts a banana near its middle into two halves. Then the main diameters of both elongated halves (the two within group PC1s) will be roughly parallel to the line segment connecting the centers of gravity of both banana halves (the mean difference vector; sliced banana argument; see Schonemann, 1997a, for more details). I convinced myself through computer simulation that this argument took care of Level II. I sent a letter to this effect to Shockley.

This line of reasoning entails a strong testable prediction: If one subgroup is much larger than the other, then its PC1 will correlate more strongly with the mean vector than the PC1 of the smaller subgroup (Unequal Splits Prediction). I verified that this prediction was borne out both for simulated data and also for Jensen's data.

On November 26, 1986, I sent a summary of these results as a Continuing Commentary to BBS, expecting it to be published alongside Shockley's Continuing Commentary. I included results of a computer simulation and a detailed discussion of the Unequal Splits effect, with explicit reference to Jensen's 1985 data. On January 30, 1987, the editor of BBS rejected my note following the advice of an associate editor who wrote:

> [Schonemann] can exhibit a special case (mathematically) where a positive correlation exists. But we already know from Jensen's data that such a positive correlation can exist, and more interestingly, does evidently exist with real data. Of course there is an underlying mathematical description of such special cases, but they are special—that is, the existance (sic) of the positive correlation depends on the structure of the particular data set, and thus is not a tautological result of the method of analysis. (BBS review, 1986)

The reviewer did not explain why anyone would have cared if the artifact had not arisen with *real data.*

1987

In his BBS commentary, Shockley (1987) had claimed, against his better knowledge, that I had conceded an "error" in overlooking that Jensen had "explicitly restricted" his definition of "Spearman Hypothesis" to Level II. However, in fact Jensen had used the simpler Level I version himself in his (1980) book, as well as in his (1985) target article. Since Shockley balked at my request to remove this unfair charge from his commentary, I terminated further dialogue with him. Subsequently, the BBS editor denied me space to rebut Shockley's misrepresentation.

To further corroborate the Level II argument with real data, I analyzed a subset of the Head Start data (see Schonemann, 1997a, for more details). They also bore out the Unequal Splits effect predicted by the Level II Banana Argument. Thus, by early 1987, most of the major points contained in my eventual *Current Psychology of Cognition* target article (Schonemann, 1997a) were in place. All I needed now was a peer-reviewed journal agreeing to publish these results. Among other things, they unequivocally refuted Jensen's outrageous claim that his positive correlations imply ". . . that Black people, statistically, will have a greater handicap in those educational, occupational, and military assignments that are most likely correlated with general intelligence" (Jensen, 1985, p. 206).

Mainstream Psychology Journals 1

> In the present case I feel almost no responsibility with regard to publishing your reply since, firstly, the article to which you object was published by my predecessor . . . (*Psychological Review,* 1976)

1988

Starting at the top, I sent a non-technical summary of my empirical results to *American Psychologist,* the flagship of the American Psychological Association. The article contained the Head Start toy data analyses, the simulation results under Level II, the Split Banana argument, and, most important, empirical verifications of the Unequal Splits prediction covering both Jensen's and the Head Start toy data. The editor rejected the paper on January 19:

> As you know, the AP is read by each member of the APA. The limited space we have available constrains us further, and we can print only articles that are of interest to a broad range of psychologists and that have broad consequences for the science and practice of psychology. . . . In your case, I would urge you to send the paper to *Behavioral and Brain Sciences,* where I feel your audience will be found. (*American Psychologist,* ms. review, 1988)

On January 22, 1988, I sent this manuscript to *Psychological Review*, another leading APA journal. The reviews were mixed and the editor seemed to hold out hope for a suitable revision:

> I am in agreement with reviewers B, C, and D about your demonstration that the resulting Spearman correlations are inevitable given the assumptions you make. . . . However, . . . I do not feel that the present argument takes us far enough . . . On the other hand, should you address these broader issues and examine how one would go about evaluating Spearman's hypothesis in a way that *is not* artifactual, then I would think the contribution would be substantial in a way that is not true for the present paper. (*Psychological Review,* ms. review, 1988a, emphasis added)

Though I was unable to think of a way of revising an article intended to show that Spearman's Hypothesis is an artifact so that it shows that it is not an artifact, I did spend considerable time responding to all other points the reviewers had raised. The editor of *Psychological Review* rejected the revision on June 23:

> I have read your revised manuscript . . . and, although the manuscript reads well and is quite convincing in its demonstration of the nature of the artifact underlying Jensen's empirical support for what he calls Spearman's hypothesis, I continue to believe that the paper represents only a very small theoretical contribution over what you have previously published on this issue. . . . It is true that you have effectively responded to various criticisms that the reviewers raised the last time around . . . nevertheless I do not feel that these additions . . . make this work publishable here. (*Psychological Review,* ms. review, 1988b)

On July 12, 1988, I submitted my article to the *Harvard Educational Review.* This journal had previously solicited and published Jensen's (1969) epochal essay "How Much Can We Boost IQ and Scholastic Achievement?" that had ushered in the eugenics renaissance of the 1970s. The editor rejected my paper out of hand, apparently without having it reviewed: "It is not our usual policy to critique all the manuscripts we don't accept" (*Harvard Educational Review,* ms. review, 1988).

Statistical Interlude—Mainstream Statistics Journals

> I have to express my view . . . that to a considerable extent the review processes which supposedly scientific journals appear to have followed are strongly defective, and the journals involved are among the most prestigious of the world. (Kempthorne, 1978, p. 4)

In the Spring of 1988, I succeeded in proving the following Perfect Collinearity Theorem: If the pooled group is exactly multinormal, and the distribution is bisected into an HI and an LO group by a plane orthogonal to the PC1 of the pooled group, then the *cosines* of all three PC1s with the mean difference vector will not just be positive, but exactly unity, so that all four vectors will be perfectly collinear. This algebraic proof provided a rigorous justification for the intuitive geometric banana argument presented previously.

On June 20, 1988, I sent a note containing a proof of this theorem to *Biometrika*. There was only one reviewer, who counseled rejection, as the editor stated: "I am very sorry but, in the light of the enclosed report, I cannot accept the paper. The referee's points 1 and 2 seem very cogent" *(Biometrika,* ms. review, 1988a). However, the review contained a third, less cogent point: "As a minor point, why does the author claim that [a certain matrix] is not positive (presumably positive definite) in figure 1d?"

Citing from his own textbook the definition of a "positive matrix," I reminded this editor that there is a basic difference between "positive" and "positive definite." A positive matrix need not be positive definite, and a positive definite matrix need not be positive. Reviewers suffering from this type of misapprehension are hardly in a position to assess the stringency of my proof. The *Biometrika* editor agreed, but refused to reconsider, now arguing lack of space. He suggested I try elsewhere: "Have you considered, say, a short note for *The American Statistician* or *Probability and Statistical Letters*?" (*Biometrika* ms. review, 1988b).

Following his advice, I sent the note to *The American Statistician* on July 22, together with the *Biometrika* review. The editor rejected it with very similar arguments: "I don't think the paper is suitable for *The American Statistician*. It's a good paper, but its contents are more suited for a journal like *Multivariate Analysis, Biometrika, Psychometrika,* etc." (*The American Statistician,* ms. review, 1988).

On August 24, I sent the note to the *Journal of the American Statistical Association*. The editor rejected it on October 11: "I don't see that anyone is questioning the *correctness* of your result. However, I am afraid to report that I find the paper not suitable for *JASA,* so that we will not be able to publish your paper" (*Journal of the American Statistical Association,* ms. review, 1988, emphasis in the original).

Thus, none of the reviewers and editors of these illustrious statistics journals found anything wrong with the proof of my Perfect Collinearity Theorem. However, they did not regard the question of its truth or falsity relevant. What exclusively mattered to them were space, appropriateness, suitability, and publishability.

Mainstream Psychology Journals 2

> In general, a group should never be condemned for the errors of one of its members. But a scientific community has a special problem when it persistently fails to detect gross and manifest error in a finding central to the discipline. (Broad & Wade, 1982, p. 202f)

1990

The editor of *Multivariate Behavioral Research* (MBR) contacted me to discuss the possibility of having Guttman's BBS Continuing Commentary published posthumously as a target article in MBR. I seconded the idea.

1992

Multivariate Behavioral Research published Guttman's "Last paper" as a target article (Guttman, 1992), together with 10 odd commentaries. One of them contained the mathematical proof of my Perfect Collinearity Theorem which the *Biometrika* editor deemed more suitable for *The American Statistician,* while *The American Statistician* editor had deemed it more suitable for *Biometrika.*

1994

In their widely acclaimed book, *The Bell Curve,* Herrnstein and Murray (1994) devoted considerable space to Spearman's Hypothesis. Among other things, they alleged:

> Another commentator suggested that Jensen had inadvertently built into his own analysis the very correlation between g loading and Black/White differences that he purported to discover (Schonemann, 1985). In the next round . . . after being apprised of a response by physicist William Shockley (Shockley, 1987), he withdrew his argument. (p.726)

The authors selectively cite various commentators on Guttman's (1992) MBR target article, but meticulously avoid mention of my commentary in the same issue (Schonemann, 1992). Far from withdrawing my original argument, I had actually extended it from Level I to the stronger Level II, and I had generalized Guttman's refutation from factors to principal components, which was what Jensen had used in his analyses.

1995/1996

In May, 1995, the editor of *Multivariate Behavioral Research* sent me an article for review. The author claimed my Perfect Collinearity Theorem was false. (He later recanted the claim in private correspondence, and in a subsequent article published by MBR, he reversed himself, calling it "trivial"). In view of the obvious conflict of interest, I excused myself as a reviewer. I made it clear that I had no objection to seeing the article published, provided I was offered space to reply to it, if it were judged publishable by other reviewers.

Along with my review, I submitted my empirical results to MBR. Ten months later, the editor apologized for the long delay: "Although I sent your paper to four individuals to review, I have only been able to get two of them to respond. One, as will become obvious when you read it, was Professor Jensen." (*Multivariate Behavioral Research,* ms. review, 1996). The enclosed review reads, in part:

> Every other year or so for the past several years some editor has sent me some version of his "Toy Factor" article by Schonemann for review. (I think I last reviewed it for *Psychological Review,* where all four of the reviewers recommended rejection [This statement is untrue, see above]). A shorter form of the article was earlier rejected by *Behavioral and Brain Sciences,* for which one of the referees (a statistician at Harvard and a past president of the Psychometric Society) pointed out its central fallacies and recommended rejection. (*Multivariate Behavioral Research,* ms. review, 1995)

1997

On July 1, I submitted the manuscript that had been rejected by *American Psychologist, Psychological Review, Harvard Educational Review,* and *Multivariate Behavioral Research* to *Cahiers de Psychologie Cognitive-Current Psychology of Cognition,* a French-English language journal. All reviews were constructive. The editor accepted the article as a target article (Schonemann, 1997a) on May 3, subject to minor revisions. Twelve of 14 commentators accepted my Level II argument. Some transposed it into a factor analytic setting (for which it was not intended since Jensen had used principal components, not factors).

The two dissenting commentators produced scenarios apparently intended as counter-examples to my claim that the positive Spearman correlations in Jensen's data are artifacts. Neither met the constraints of Jensen's data. One author (Maraun, 1997) constructed a ficticious example which conflicts with "one of the most striking and solidly established phenomena in all of psychology . . . the fact of ubiquitous positive correlations of all mental ability" (Jensen, 1980, p. 249). The other author (Steiger, 1997) marshaled three major objections:

1. Following Loehlin (1992), he questioned the multinormality assumption underlying my Level II argument. Loehlin (1992) had previously argued that, if two normal distributions with different means are pooled, "the combined population . . . will be bimodal, with each subpopulation normally distributed around its own mean" (p. 261). As I showed in my response, this claim is clearly false (Schonemann, 1997b, p. 790). Steiger argued, more generally, that the multinormality assumption that undergirds my Level II argument was unrealistic for Jensen's data. However, he produced no empirical evidence to support this bold claim.
2. He also raised the specter that my simulations may have unwittingly introduced a general factor in accordance with Thomson's sampling theory—but again without adducing any supporting evidence. On checking, I verified that this concern was unwarranted (Schonemann, 1997b, p. 791).
3. Finally, he produced a numerical "counterexample," presumably intended to back up his claim that (my multinormal) "Model 2 need not fit the data" (Steiger, 1997, p.765). "If we adopt [his] Model 1, it is difficult to see why the 'Spearman hypothesis correlations' would be inevitable" (p. 765). More generally, Steiger believed ". . . the process Model 2 represents is not the process being studied, and the data characteristics it presupposes do not match the characteristics of the observed data" (p. 764).

Now while it is quite true that a model—any model—"need not fit the data," the point at issue was whether it fit Jensen's data. The only way to find out which model fits them better is by comparing the fit of both models to these data. This Steiger failed to do—as it turned out—for very good reasons. Nor did he explain how his Model 1 is supposed to account for the robust Unequal Splits effect—perhaps for the same reasons.

1998

Partly in response to Steiger's challenge, and partly in order to close the gap I had left in my Level II argument between correlations and cosines that had escaped notice of all the reviewers, I submitted a short paper to MBR entitled "Some New Results on Spearman's Hypothesis" (Schonemann, 1998). After a lengthy delay, the editor sent me three negative reviews.

In the first part of this paper, I showed that Steiger's (1997) claim—that the multinormality assumption underlying my Level II argument "does not match the characteristics of the observed data" (p. 764)—is not only unsubstantiated but false. In the one case where the fit can actually be compared (viz. the parameterization of Model 1 provided by Steiger, 1997, p. 766), the multinormality assumption fits Jensen's (1985) data much better than Steiger's Model 1.

In the second part, I stated and proved a theorem that relates cosines to correlations. It closes a conceptual gap in my Level II argument that all previous reviewers and commentators had overlooked: This argument and the Perfect Collinearity Theorem are both couched in terms of cosines as collinearity measures of pairs of positive vectors. This is reasonable because correlations are undefined as collinearity measures if one of the vectors is constant (as is, e.g., a vector of ones). The Perfect Collinearity Theorem states that, in the case of positivity and perfect multinormality, the cosines between the mean difference vector and all three PC1s will be perfect. This implies that the correlations will be high but does not answer the question: What if the assumptions of the theorem are met only approximately? The new theorem gives a precise answer: the Spearman Hypothesis correlation will be positive if, and only if, the cosine between the mean difference vector and the first eigenvector exceeds the product of the cosine between the mean difference vector and a vector of one times the cosine between the first eigenvector and a vector of ones.

Readers need not worry unduly about the technical jargon. The main reason for stating the theorem in such detail here is to set the stage for a fuller appreciation of the remarks of a particular Reviewer (C, below) of the *New Results* paper (Schonemann, 1998).

EFFECTIVE STRATEGIES FOR THE PRESERVATION OF PREJUDICE

> . . . in the study of human affairs evasion and deception are as a rule much more profitable than telling the truth. (Andreski, 1972, p. 12)

> By and large, those features in the social organization of science that encourage and reward careerism also create the incentive for fraud. (Broad & Wade, 1982, p. 220)

Of the three negative MBR reviews of my *New Results* paper, one in particular, submitted by an anonymous "Reviewer C," deserves to be quoted extensively. Reviewer C (1998) started out by recapitulating the essence of my *Current*

Psychology of Cognition target article. Without further ado, he acknowledged that my Level I and Level II arguments are both perfectly valid:

> . . . namely, that if n variables are positively correlated and the means on these variables (M1, M2 . . . Mn) for one group of subjects are larger than the means for another group (m1, m2, . . . mn), then the correlation between a "d" vector of differences between the means (M1-m1, M2-m2, Mn-mn) and a "PC1" vector of the first principal component coefficients (C1, C2 . . . Cn) for those variables (whether calculated separately within the two groups or over the sample of the two groups combined) is, as Schonemann worded it, "an artefact" (more commonly spelled artifact) and that Jensen (1980; 1985; 1987; 1992), in analyses of eleven data sets, and several scientists who had accepted and praised Jensen's analysis, seem not to have recognized this artifactual condition and incorrectly regarded the correlation between the d vector and the PC1 vector as entirely an "empirical discovery" supporting an hypothesis of a g factor among measures of cognitive abilities. (MBR reviewer C, 1998)

In short, reviewer C agreed with me on all important points. This put him squarely at odds with MBR reviewers A and B. For example, for MBR reviewer A (1998), "most of the present article is a giant smoke screen designed to deflect attention from the fact that (a) there really is no 'artefact.' . . ."

In contrast to MBR reviewers A and B, MBR Reviewer C further agreed with me that the Level II scenario is a realistic account of the actual empirical situation reflected in Jensen's data: "The point was well made: Jensen was wrong in interpreting the results of his calculations as support for an hypothesis of a g factor among cognitive tests." He further agrees with me that Steiger's numerical example does little to undermine this conclusion, because, as I had shown in the paper he was asked to review, all available evidence confirms that multinormality fits Jensen's IQ data much better than Steiger's Model 1. However, reviewer C did not view this as a worthwhile contribution to the Spearman Hypothesis debate:

> The only thing new in this manuscript is an extended rejoinder to the CPC article of Steiger, who made up numbers that did not represent the conditions and assumptions of the Jensen data and analyses, and then demonstrated that, yes, under these conditions it is possible to have a "d" vector that correlates zero with the PC1 vector. MBR #810 is nothing more than a statement that the conditions Steiger invented are not conditions that well represent conditions under which Jensen made his incorrect "discovery."

MBR reviewer C then addressed the significance of the topic, that is, the mystery why all this has not been sorted out long ago:

> It is true that there is widespread belief that the g hypothesis among cognitive measures is well supported by data and that differences between racial/ethnic groupings are particular manifestations of this support. It should be well communicated that such beliefs are not well founded on the results from scientific studies. For the most part, the articles of CPC are such communications. But the journal does not have a large readership. It might be deemed desirable to get that message out through other scientific journals—for example, MBR.

This thought had also occurred to me. After all, it had been the American psychological establishment, not the French one, that had kept the bad news bottled up for more than 12 years.

So far, all points reviewer C made seemed to bode well: He agreed with my basic results that Jensen's correlations are artifacts under both the Level I and the Level II interpretation. He dismissed Steiger's "made-up numbers" as irrelevant. He agreed with me on the significance of the Spearman Hypothesis for g and the race debate in the U.S. And, finally, he also deemed it desirable to "get that message out through other scientific journals, for example MBR." What, then, might be his recommendation to the editor? "This article should not be published in MBR. There is little reason to publish it anywhere" (MBR reviewer C, 1998). Why not?

> . . . for the simple reason that it does not communicate very well. It is a statement from a geometric analogy, usually involving cosines. It is not well developed in terms of the mathematics of component analysis or factor analysis, which if it were, would better indicate the inaccuracy of Jensen's conclusions.

CONCLUSION

> The upshot of these operating principles is that the peer review printed word is subject to secrecy, caprice, delay, rejection, and premature disclosure. It is not difficult for any of these to occur; in many instances reviewers need only act in their rational self-interest . . . Thus motives other than the extension of certified knowledge may influence editorial decisions, although such actions betray the community's trust in the gatekeepers. (Chubin & Hackett, 1990, p. 94)

The preceding account documents concretely and in detail how the peer review system currently operates in the social sciences across a representative sample of leading journals. The salient point is not, of course, that we may occasionally run into a sub-optimal editor. Rather, it is the uncanny uniformity of editorial responses across time and disciplines that warrants concern. The common theme—surely one of the most solidly replicated findings in the social sciences—boils down to this: on a purely technical issue at the heart of the current IQ debate with potentially far-reaching implications for society—some of which Jensen himself had spelled out with admirable lucidity—none of the psychology editors cared one whit whether the claims at issue were true or false. All they seemed to care about was the specious appearance of infallibility of previous editorial decisions.

Contrary to what one might have expected, the same was true for the editors of the statistics journals. In particular, none of the editors, whether from statistics or from psychology, showed the slightest concern about the potentially harmful consequences that might ensue from embalming faulty research claims in our mainstream journals.

Some may wonder what could possibly have motivated these editors to participate in this mockery of science. It is probably pointless to speculate about motives. Some of the editors may indeed sympathize with Jensen's implied eugenic message, but others may not. Most of them probably do not care one way or another. Numerous other explanations, more banal than that of outright racism, spring to mind, not all of them mutually exclusive: arrogance, technical incompetence, careerism, false loyalties to the "peer group," worries about continued funding of "research" as usual, and so on. The essential point, in any case, is not motives, but the potential consequences of such editorial decisions, whatever the motives. Recent history shows all too clearly that, under the right set of circumstances, banal motives can have disastrous consequences, including mass murder in the name of science:

> —There is a difference between Poles and Jews?
> —Oh yes!
> —What difference?
> —The Poles weren't exterminated and the Jews were. That's the difference. An external difference.
> —And the internal difference?
> —I can't assess that. I don't know enough about psychology and anthropology.
> (Lanzman, 1985, p. 83)

NOTE

A *First Principal Component* (PC1) is a weighted average of a number of given variables, for example, of subtests of a test battery. If all given variables correlate positively with each other (as subtests tend to do by design), then the weights defining the PC1 are all positive and the PC1 scores will not differ much from the average test scores.

Spearman's Hypothesis asserts a monotone relationship between the weights that define the PC1 and the mean differences between two groups of subjects (e.g., Whites and Blacks, rich and poor): Subtests most closely related to a presumed underlying "general ability factor" (g) exhibit the largest mean differences between the two groups.

General ability (g) is a hypothetical variable that fundamentally differs from the PC1: According to Spearman (who coined the term), g is supposed to account for the positive intercorrelations among subtests or items of intelligence tests. Once it is removed ("partialled out"), all intercorrelations are supposed to reduce to zero. This virtually never happens in real life. Hence, "g does not exist." By confusing g with PC1 (which always exists but does not leave zero residual correlations, and hence proves nothing about the existence of g), Jensen and his followers have created the mirage of progress in mental testing where, in fact, standards have been steadily declining over the last few decades.

REFERENCES

Andreski, S. (1972). *Social sciences as sorcery.* New York: St. Martin's Press.

Blum, J. F. (1978). *Pseudoscience and mental ability.* New York: Monthly Review Press.

Bo Yang. (1992). *Choulou de Zhongguo Ren.* Taipei: Xingguang Chuban.

Brigham, C. C. (1923). *A study of American intelligence.* Foreword by Robert M. Yerkes. Princeton: Princeton University Press.

Broad, W., & Wade, N. (1982). *Betrayers of the truth: Fraud and deceit in the halls of science.* New York: Simon and Schuster.

Chubin, D. E., & Hackett, E. J. (1990). *Peerless science: Peer review and U.S. science policy.* Albany, NY: State University of New York Press.

Guttman, L. (1992). The irrelevance of factor analysis for the study of group differences. *Multivariate Behavioral Research, 27,* 175-204.

Hearnshaw, L. S. (1981). *Cyril Burt, psychologist.* New York: Vintage Books.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve.* New York: Free Press.

Hirsch, J. (1981). To "unfrock the charlatans." *Sage Relations Abstracts, 16,* 1-67.

Jensen, A. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review, 39,* 1-123.

Jensen, A. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. (1985). The nature of the Black-White difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8,* 193-263.

Jones, L. V. (1985). Golly g: Interpreting Spearman's general factor. *Behavioral and Brain Sciences, 8,* 233.

Kempthorne, O. (1978). Logical, epistemological, and statistical aspects of nature-nurture data interpretation. *Biometrics, 34,* 1-22.

Lanzman, C. (1985). *Shoah. An oral history of the Holocaust.* New York: Pantheon Books.

Loehlin, J. C. (1992). On Schonemann on Guttman on Jensen, via Lewontin. *Multivariate Behavioral Research, 27,* 261-263.

Peters, D. P. & Ceci, S. J. (1982). Peer review practices of psychological journals: The fate of articles, submitted again. *Behavioral and Brain Sciences, 5,* 187-257.

Robitscher, J. (1973). *Eugenic sterilization.* Springfield, IL: Charles C Thomas.

Schonemann, P. H. (1992). Extension of Guttman's result from g to PC1. *Multivariate Behavioral Research, 27,* 219-224.

Schonemann, P. H. (1998). *Some new results on Spearman's hypothesis.* Unpublished manuscript.

Schonemann, P. H. (1997a). Famous artefacts: Spearman's Hypothesis. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition, 16,* 665-694.

Schonemann, P. H. (1997b). Rise and fall of Spearman's Hypothesis. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition, 16,* 788-812.

Schonemann, P. H. (1997c). On models and muddles of heritability. *Genetica, 99,* 97-108.

Schonemann, P. H. (1985). On artificial intelligence. *Behavior and Brain Sciences, 8,* 241-242.

Shockley, W. (1987). Jensen's data on Spearman's hypothesis: No artifact. *Behavioral and Brain Sciences, 10,* 512.

Spearman, C. (1927). *The abilities of man.* New York: MacMillan.

Steiger, J. H. (1997). Alternate models and the evaluation of social issues. *Cahiers de Psychologie Cognitive / Current Psychology of Cognition, 16,* 762-768.

Weiss Hearings. (1988). *Scientific fraud and misconduct and the federal response.* Washington, DC: U.S. Government Printing Office.

Acknowledgments. This article is dedicated to Professor Jerry Hirsch, University of Illinois, a courageous and untiring critic of charlatanism in the social sciences (e.g., Hirsch, 1981). I am also deeply indebted to Professor Barry Mehler (Ferris State University, founder of the Institute of the Study of Academic Racism) for his encouragement and support through many years, without whom this story might never have been told.

Offprints. Requests for offprints may be directed to Peter H. Schonemann, PhD, Dept. of Psychological Sciences, Purdue University, West Lafayette IN 47907.