Psychological Bulletin
1985, Vol. 97, No. 1, 129–133

0033-2909/85/$00.75

A Variance Explanation Paradox: When a Little is a Lot

Robert P. Abelson
Yale University

Concerning a single major league at bat, the percentage of variance in batting performance attributable to skill differentials among major league baseball players can be calculated statistically. The statistically appropriate calculation is seriously discrepant with intuitions about the influence of skill in batting performance. This paradoxical discrepancy is discussed in terms of habits of thought about the concept of variance explanation. It is argued that percent variance explanation is a misleading index of the influence of systematic factors in cases where there are processes by which individually tiny influences cumulate to produce meaningful outcomes.

It is generally accepted that percentage of variance explained is a good measure of the importance of potential explanatory factors. Correlation coefficients of .30 or less are often poor-mouthed as accounting for less than 10% of the variance, a rather feeble performance for the influence of a putatively systematic factor. In analysis of variance contexts, the percentage of variance explanation is embodied in the omega-squared ratio of the systematic variance component to the total of the systematic and chance variance components. It, too, is often small; when it is, this is a source of discouragement for the thoughtful investigator.

Psychologists sometimes tend to rely too much on statistical significance tests as the basis for making substantive claims, thereby often disguising low levels of variance explanation. It is usually an effective criticism when one can highlight the explanatory weakness of an investigator's pet variables in percentage terms.

Having been trained, like all of us, in the idiom of variance explanation, I have always believed that when levels of variance explanation are extremely small, then the variables involved are really quite unimportant (however much one may lament the fact in a given case). However, I have been led to reexamine this notion.

A colleague and I recently had an argument in which we took opposing views of the role of chance in sports events. I claimed that many games of baseball and football are decided by freaky and unpredictable events such as windblown fly balls, runners slipping in patches of mud, baseballs bouncing oddly off outfield walls, field goal attempts hitting the goalpost, and so on. Even without obvious freakiness, I claimed, the ordinary mechanics of skilled actions such as hitting a baseball are so sensitive that the difference between a home-run swing and a swing producing a pop-up is so tiny as to be unpredictable, thus requiring it to be considered in largely chance terms.

My colleague argued that chance characterizations of sports events ignore the obvious fact that good teams usually win, that even under freaky circumstances (wind, mud, and so on) skilled players will better overcome difficulties than mediocre players, and furthermore that the visual–motor coordination of skilled athletes is subject to causal analysis.

Without trying to resolve in any serious way the deeper issues involved in the meanings of causation and chance in sports events, a straightforward statistical question can be raised: What percentage of the variance in

Willa Dinwoodie Abelson, Fred Sheffield, Allan Wagner, and Rick Wagner provided helpful comments on an earlier draft of this article. I wish also to thank the faculty and graduate students of the Yale University Psychology Department for exposing themselves to potential collective embarrassment by filling out the questionnaire.

Requests for reprints should be sent to Robert P. Abelson, Box 11A Yale Station, New Haven, Connecticut 06520.

athletic outcomes can be attributed to the skill of the players, as indexed by past performance records? This variance explanation question is analogous to those that characterize psychological investigations, but arises in a context where there exist strong intuitions (among sports fans, at least). A comparison of intuition with fact might therefore prove interesting.

To elicit intuitions, the athletic performance in question must be concretized. A simple performance with which most Americans are familiar, and for which copious records exist, is batting in baseball. The simplest event to consider is whether or not the batter gets a hit in a given official time at bat. It is possible to calculate statistically the proportion of the variance of this event (getting or not getting a hit) explained by skill differentials between batters.

Calculation of Variance Explanation

Let the dependent variable be $X = 1$ for a hit and $X = 0$ for no hit, and conceptualize the data matrix as in Table 1. Columns represent different batters. Rows represent different times at bat in, say, 5 years of at bats for each batter, a period long enough to give a reliable indication of the batters' true averages. The number of at bats might as well be taken as equal for all batters: The subsequent calculation is not affected by this factor.

Much as in the usual analysis of variance fashion, Equation 1 decomposes the entire set of X_{ji} in Table 1 into a true mean B_i for the ith batter and an error component e_{ji} for the jth occasion for the ith batter:

$$X_{ji} = B_i + e_{ji}. \tag{1}$$

The variance components σ_B^2 and σ_e^2 attaching to the two terms give the ingredients necessary to answer our variance explanation question. The former represents the variability of true batting averages, the latter the variability of performance given the batting average.

Both components depend on the distribution of true batting averages. Let the mean of the distribution of B_i be μ_B and the standard deviation σ_B. To compute the within-batter variance, σ_e^2, consider a batter with true

Table 1
Hypothetical Data Matrix for Batting Outcomes

	Batters								
At bats	1	2	3	·	·	*i*	·	·	·
1	0	0	1	—	—	—	—	—	—
2	1	0	0	—	—	—	—	—	—
3	0	0	0	—	—	—	—	—	—
·	—	—	—	—	—	—	—	—	—
·	—	—	—	—	—	—	—	—	—
j	—	—	—	—	—	X_{ji}	—	—	—
·	—	—	—	—	—	—	—	—	—
Batting average	.282	.214	.301	—	—	—	—	—	—

Note: 0 = no hit; 1 = hit.

average, B_i. On occasions when this batter gets a hit, $X_{ji} = 1$, and from Equation 1, $e_{ji} = 1 - B_i$. When the batter fails, $X_{ji} = 0$, and $e_{ji} = -B_i$. The first type of event happens on the proportion B_i of all occasions, the second type of event on the proportion $(1 - B_i)$. Weighting the squares of the e_{ji} by these proportions, the result is

$$\begin{aligned} \sigma_{e(i)}^2 &= B_i(1 - B_i)^2 + (1 - B_i)(-B_i)^2 \\ &= B_i(1 - B_i)[(1 - B_i) + B_i] \\ &= B_i(1 - B_i). \end{aligned} \tag{2}$$

(This is simply the formula for the variance associated with a binomial event around a true proportion B_i; I have rederived it in order to be explicit.)

Now consider the fact that because batting averages differ, the error variance is not the same for all batters. To obtain a summary value for σ_e^2, Equation 2 must be averaged over all values of B_i, weighted by the probability $p(B_i)$ of their occurrence.

$$\begin{aligned} \sigma_e^2 &= \sum_i B_i(1 - B_i)p(B_i) \\ &= \sum_i B_i p(B_i) - \sum_i B_i^2 p(B_i). \end{aligned} \tag{3}$$

The respective terms on the right are by definition the raw first and second moments of the distribution of B_i. That is,

$$\begin{aligned} \sigma_e^2 &= \mu_B - (\mu_B^2 + \sigma_B^2) \\ &= \mu_B(1 - \mu_B) - \sigma_B^2. \end{aligned} \tag{4}$$

Hence, the omega-squared ratio for proportion of variance attributable to skill is:

$$\omega^2 = \frac{\sigma_B^2}{\sigma_B^2 + \sigma_e^2} = \frac{\sigma_B^2}{\sigma_B^2 + \mu_B(1 - \mu_B) - \sigma_B^2}$$
$$= \frac{\sigma_B^2}{\mu_B(1 - \mu_B)}. \quad (5)$$

Finally, realistic values are needed for σ_B and μ_B to substitute in Equation 5. These parameters of the distribution of true batting averages of course differ somewhat from year to year and league to league. However, the bulk of the distribution of observed batting averages of major league regulars in a given year typically lies between the low .200s and the low .300s. This suggests parameters such as μ_B = .270 and σ_B = .025. These values yield

$$\omega^2 = \frac{(.025)^2}{(.270)(.730)} = .00317.$$

In other words, the percentage of variance in any single batting performance explained by batting skill is about one third of 1%.

What's Going on Here?

One's first reaction to this result is incredulity. My personal intuition was jarred by this result, which seems much too small. To check my own intuition against those of others, I circulated a one-item questionnaire to all graduate students and faculty in the Department of Psychology at Yale University. This group was chosen not simply for convenience, but because they would be familiar with the concept of variance explanation. Respondents were asked to refrain from answering if they knew nothing about baseball or the concept of variance explanation. Participants were asked to imagine a time at bat by an arbitrarily chosen major league baseball player, and to estimate what percentage of the variance in whether or not the batter gets a hit is attributable to skill differentials between batters.

The median of the 61 estimates of the variance attributable to skill was 25%, an overestimate of the calculated estimate by a factor of 75. The estimates of over 90% of the sample were too high by a factor of at least 15. Only 1 person gave an underestimate. I also posed the skill variance question to colleagues outside of Yale (some of whom are well known for their statistical acumen) and commonly received answers around 20% or 30%. The outcome of the statistical calculation, .3%, is indeed surprising.

Another attack on the paradox is to look for flaws in the statistical calculation. One thing to consider is the sensitivity of Equation 5 to variations in the parameters σ_B and μ_B. The term $\mu_B(1 - \mu_B)$ does not change appreciably with small variations in μ_B; the value for ω^2, in other words, would be nearly the same if I took μ_B = .265 or .260 or .275 rather than .270. The ratio is more sensitive, though, to variations in σ_B. If σ_B were more than .025, then ω^2 would of course be bigger. However, .025 is, if anything, a generous estimate. If lifetime batting averages are taken as more indicative of true ability than season-by-season averages their standard deviation would be used for σ_B. Calculated from data in James's (1983) baseball abstract, the mean lifetime average was .268 and the standard deviation of lifetime averages for all major league regulars active in 1983 was .021. Even if I generously inflated this estimate to include nonregular players—even if I, say, doubled it to .042—the omega-square for skill variance would still be below 1%.

Could Equation 5 ever give a large value for ω^2? Yes, if every batter batted either 1.000 or .000 (i.e., either perfect or perfectly awful), then ${\sigma_B}^2 = \mu_B(1 - \mu_B)$, and $\omega^2 = 1$, as one would expect. This extreme situation contrasts sharply with reality. (Indeed, a way to understand the paradox is to realize that in the major leagues, skills are much greater than in the general population. However, even the best batters make outs most of the time.)

So the paradox remains. When I told my colleague the result of the calculation, he said, "You mean to tell me that the difference between George Brett and Len Sakata doesn't amount to anything?" This comment places the burden of the skill variance on extreme exemplars. The statistical calculation, of course, includes players of all levels of ability, most of them nearly average. Also, the comment appeals to the long-run differences in ability, whereas the calculation refers to the single at bat, a much chancier proposition. Thus, the paradox may arise in part because the intuitive way of conceptualizing the question is intrinsically different from the appropriate statistical formulation, as in the phenomena discussed by Kahneman and Tversky (1982) and by Nisbett, Krantz, Jepson, and Fong (1982).

Is the statistical formulation therefore somehow unfair or irrelevant (Cohen, 1981)? Hardly. The single at bat is a perfectly meaningful context. I might have put the question this way: As the team's manager, needing a hit in a crucial situation, scans his bench for a pinch hitter, how much of the outcome variance is under his control? Answer: one third of 1%. Qualification: This assumes that the standard deviation of batting averages against a given pitcher is the same as the standard deviation of batting averages in general.

One might also argue that, in this framework, the manager may be able to choose someone two standard deviations above average and definitely avoid someone two standard deviations below average. By so doing, he would effectively double the standard deviation, and thus quadruple the skill component of variance. Even at that, the percentage of variance explanation would be only about 1.3%. In variance explanation terms, the difference between, say, George Brett and Len Sakata really is of small consequence. To appreciate why this is so and perhaps alleviate one's sense of paradox, it may be helpful to picture this comparison as in Table 2.

In Table 2 the rows represent batters with widely different skill levels, and the columns represent the outcome variable of getting a hit or not. The entries represent projected frequency of each outcome per 1,000 at bats. Even though hits are almost 50% more frequent for the .320 than for the .220 batter, the correlation between skill and outcome is not very sizable. The phi coefficient calculated from Table 2, for example, is .113. Taking the square of this as an estimate of variance explanation yields 1.3%.

Larger Implications

I have given an example from a nonpsychological context in which the percentage of variance actually explained by an independent variable (skill) is pitifully small, whereas "everyone knows" that the variable in question has substantial explanatory power. The paradox probably does not depend on some peculiarity of the intuitions of psychologists. The public cannot reasonably be asked the exact question about variance explanation, but it is a safe guess that skill is considered relatively important by the typical baseball fan.

Table 2
Correlation Between Skill and Outcome

	Outcome	
Skill of batter	Hit	No hit
Well above average	320	680
Well below average	220	780

Note. 1,000 at bats per batter.

What does the baseball paradox suggest for the usual standards for conceptualizing variance explanation? If one-third percent indicates such a trivial degree of explanation as to be virtually meaningless, should differential batting skill then be dismissed as an explanatory variable in baseball? Or should one instead be more suspicious of variance explanation as an index of systematic influence, and revise the notions surrounding less than 1% of variance explanation?

The answer lies in the type of example under consideration. The baseball example, as it turns out, exaggerates the paradox. The baseball case may take advantage of the "illusion of control" (Langer, 1975), by which skill influences are exaggerated at the expense of chance influences. Beyond that, however, there is a sound basis for the belief that systematic differences in batting averages are nontrivially predictive of success in baseball, in ways not captured by the statistical calculation. First, the individual batter's success is appropriately measured over a long season, not by the individual at bat. Second, a team scores runs by conjunctions of hits, so a team with many high-average batters is more likely to stage rallies than a team with many low-average batters. Thus, team success over a long season is influenced by average batting skill far more than is individual success in the single at bat because the effects of skill cumulate, both within individuals and for the team as a whole.

The statistical effects of cumulation are well known, although they are usually discussed in methodological contexts, such as the psychometrics of reliability of measurement or the prediction of behavior from attitude measures (Epstein, 1979). The message here is that it is the *process* through which variables operate in the real world that is important. In the present context, the attitude toward explained variance ought to be conditional on the degree to which the effects of the explanatory factor cumulate in practice. Some examples of potentially cumulative processes are educational interventions, the persuasive effects of advertising, and repeated decisions by ideologically similar policy makers. In such cases, it is quite possible that small variance contributions of independent variables in single-shot studies grossly understate the variance contribution in the long run.

Thus, one should not necessarily be scornful of miniscule values for percentage variance explanation, provided there is statistical assurance that these values are significantly above zero, and that the degree of potential cumulation is substantial. On the other hand, in cases where the variables are by nature nonepisodic and therefore noncumulative (e.g., summary measures of personality traits), no improvement in variance explanation can be expected.

In sum, the large intuitive overestimation of the variance in batting outcome explained by skill is not simply an error in the appreciation of statistics. It reflects an intuition that skill does matter. Indeed it does, in the long run, albeit not very consequentially in the single episode. The baseball paradox is thus a model for similar paradoxes that may arise in psychological contexts.

References

Cohen, L. J. (1981). Can human irrationality be experimentally demonstrated? *The Behavioral and Brain Sciences, 4,* 317–331.

Epstein, S. (1979). The stability of behavior: I. On predicting most of the people most of the time. *Journal of Personality and Social Psychology, 37,* 1097–1126.

James, B. (1983). *The Bill James baseball abstract.* New York: Ballantine Books.

Kahneman, D., & Tversky, A. (1982). Variants of uncertainty. *Cognition, 11,* 143–157.

Langer, E. J. (1975). The illusion of control. *Journal of Personality and Social Psychology, 32,* 311–328.

Nisbett, R. E., Krantz, D. H., Jepson, C., & Fong, G. T. (1982). Improving inductive inference. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), *Judgment under uncertainty: Heuristics and biases* (pp. 445–459). New York: Cambridge University Press.

Received February 14, 1984
Revision received June 19, 1984 ■