

# Disputing the Effects of Championship Pressures and Home Audiences

Roy F. Baumeister  
Case Western Reserve University

New data led B. R. Schlenker, S. T. Phillips, K. A. Boniecki, and D. R. Schlenker (1995) to question whether home teams perform unusually badly (choke) when on the verge of winning a championship. Despite the new data, their main findings did not differ significantly from previous work that supported the “home-choke” hypothesis; they merely found that the effect dropped below significance. Their new data were confounded by a rule change favoring home teams. Their analysis omitted many games in which home teams apparently choked and lost. Also, their null findings did not justify strong conclusions. On the positive side, their new data on timing of errors did shed new light on and suggest modifications of previous theory. Implications regarding archival research are discussed.

Baumeister and Steinhilber (1984) provided evidence from archival research with baseball and basketball championships that a burden of expectations, in the form of a supportive audience for a high-stakes performance, can cause people to “choke” (perform poorly). Although this conclusion was counterintuitive, it has been supported by subsequent laboratory work (Baumeister, Hamilton, & Tice, 1985; Baumeister, Hutton, & Cairns, 1990; Heaton & Sigall, 1991), and independent archival investigations by other researchers have found similar effects in golf (Wright, Jackson, Christie, McGuire, & Wright, 1991) and for ice hockey (Wright, Voyer, Wright, & Roney, in press). Despite this converging evidence, Schlenker, Phillips, Boniecki, and Schlenker (1995) reached the conclusion that the hypothesis was mostly wrong and that supportive home audiences are consistently helpful to performance. I disagree with their conclusion.

To be sure, Schlenker deserves to be heard. His early work on self-presentation strongly influenced and inspired me when I began my own career, and so when I moved into studies of performance under pressure I drew heavily on his ideas. Baumeister and Steinhilber (1984) had two main bases for the hypothesis that home teams choke when on the verge of winning a championship: First, the prospect of a desired identity change in front of a supportive audience increases self-attention (Schlenker & Leary, 1982), and second, self-attention causes choking (Baumeister, 1984). The latter part was supported by direct empirical studies, but the former was far less well supported (Schlenker and Leary’s original formulation was a theoretical statement) and must be considered to be vulnerable to refinement or disconfirmation. It is appropriate that Schlenker should enter into this debate, given that a central aspect of the theory was based on his work.

The conclusion by Schlenker et al. (1995) was that Steinhilber and I (Baumeister & Steinhilber, 1984) were wrong: More precisely, it was that home teams do not choke when on

the verge of winning a championship. Although their own data indicated that home teams have lost 52% of seventh (final) games in the World Series, Schlenker et al. still concluded that home teams enjoy an advantage in such games. The fact that different researchers can reach such different conclusions from overlapping (although not identical) data sets raises troubling questions about the usefulness of archival data for testing hypotheses.

## Problems in the Schlenker et al. (1995) Article

The conclusions by Schlenker et al. (1995) were based on updated analyses. Despite the impressive amount of work they put into arguing their case (and despite the clear value of their wholly new findings on the timing of errors), several key problems and ambiguities plagued their investigation.

First, the main data reported by Schlenker et al. (1995) did not differ significantly from what Baumeister and Steinhilber (1984) found. I performed chi-square goodness-of-fit tests on Schlenker et al.’s main results as reported in their Table 2. I used the proportions found by Baumeister and Steinhilber as the expected values and Schlenker et al.’s data as the observed values. None of the numbers differed significantly. For the (baseline) home-field advantage in initial games, Baumeister and Steinhilber found that home teams won 60.2% of such games, and Schlenker et al. found 60.0%. In last games, Baumeister and Steinhilber found 41%, and Schlenker et al., with their new data, found 47%. For Game 7, Baumeister and Steinhilber found 39%, and Schlenker et al. found 48%. The last of these showed the largest discrepancy between the results, yet even that difference failed to approach significance,  $\chi^2(1, N = 29) = 1.17, n.s.$  Thus, Schlenker et al.’s results were essentially similar to those found by Baumeister and Steinhilber. All that Schlenker et al. could have claimed was that the addition of new data from the past decade—data that, they rightly pointed out, have not fit the pattern identified by Baumeister and Steinhilber—dropped the findings below significance but in the same direction. A replication that produces a similar direction of findings as the original, but below significance, is not usually considered to be a valid basis to reject the initial conclusion, although it may raise questions about generality.

---

Correspondence concerning this article should be addressed to Roy F. Baumeister, Department of Psychology, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7123.

Should the most recent decade's (1983–1993) games have been included on the same basis with the data Baumeister and Steinhilber (1984) analyzed? This brings up the second central problem with Schlenker et al.'s (1995) findings, which is that recent rule changes have given a new advantage to home teams. Schlenker et al. noted that different eras in baseball history have produced different patterns of home versus visitor success in final games, but they contended that these differences must be due simply to chance variation, "identical to flipping an inordinate number of 'tails' in a row when flipping a coin." (p. 637). They neglected to mention important rule changes that may have contributed to the differences. The new data they added were confounded.

The most important rule change has to do with what in baseball terms is called the *designated hitter*. The meaning of this rule has to do with specialization. In most professional sports, specialization of players has progressed as far as the rules allow, with the extreme being football, in which almost no player plays both offense and defense, and substitutions occur on almost every play. Baseball rules have long resisted this trend and have required all players to participate in both offense and defense. As a result, the most important defensive player was typically an exceptionally poor offensive player: Pitchers routinely had the lowest batting averages among major league players.

To remedy this, the designated hitter rule was introduced. This rule allows the pitcher to play only defense. When his turn at bat comes up, another player, chosen solely for his offensive ability, is allowed to take the pitcher's place. All leagues except one adopted this designated hitter rule. The exception, however, was the National League, which furnishes one of the contestants in every World Series. As a result, the World Series found itself with two teams that were designated differently and were accustomed to different rules and opportunities. American League pitchers had not batted all season (or even in previous years in the minor leagues) before the World Series; National League teams did not include designated hitters on the roster.

The initial solution to this problem was to alternate rules each year, but after 1985 this was changed to alternate by games so that the home team's preferred rules were the ones used in every game. Clearly this new pattern produced a confound for the theoretical question of how home teams perform in the final game. In the 1970s, the rules governing designated hitters in the championship series merely produced random error variance, insofar as the rules were independent of game location. The new pattern, however, brought a consistent and systematic advantage to every home team in every World Series game. It is therefore hardly surprising that home teams have fared better in the 1980s and 1990s—indeed, winning 75% of early games and 100% of seventh games since 1986.

Third, the analyses presented by Schlenker et al. (1995) did not correspond to their hypothesis. Their hypothesis was that when the home team has the opportunity to win the championship, their skilled performance will deteriorate. This would logically be tested by examining all games in which the home team had the chance to win the championship. Instead of looking at such games, however, Schlenker et al. merely looked at the final game in every series. Sometimes the final game was relevant, sometimes not. Games in which the home team had the chance to win the title but choked and lost, thereby forcing another

game (and thereby providing data favorable to our hypothesis), were systematically excluded from their analysis.

For example, in the 1992 World Series, the Toronto team was playing to win their first baseball championship ever—in fact, the first championship for any Canadian team, which made that series especially relevant to the hypothesis of identity change. The only game in that series in which the home team had a chance to win the championship was Game 5, held in Toronto with the home team leading 3–1 in games. Consistent with Baumeister and Steinhilber's (1984) hypothesis, the home team blew the lead and lost that game. That game was not included in Schlenker et al.'s analyses, however. Instead, they included Game 6 from that series, in which the home team did not have the possibility of winning the championship, although its outcome should have been irrelevant. Thus, they omitted games relevant to the hypothesis but included irrelevant ones.

Indeed, the most famous such game of the past decade was likewise omitted. In the 1986 American League championship series, the California Angels were about to clinch the pennant at home. They led in the score, and their opponents were down to their final out in the last inning. With two strikes against the batter, the pitch that could have clinched the pennant was instead hit for a home run, allowing Boston to win the game and, later, the series. This game was noteworthy because observers said that the losing pitcher was never the same after this traumatic failure; his career quickly deteriorated, and he committed suicide less than 3 years later.

In their article, Schlenker et al. (1995) made a big point of saying that Baumeister and Steinhilber's (1984) effect, the home choke, had not happened during the past decade (1983–1993). Yet the Toronto and California examples show that it did happen; Schlenker et al. unfortunately omitted the relevant supportive evidence.

In defense of Schlenker et al. (1995), I wish to point out that this error in designing the analysis was originally made in Baumeister and Steinhilber's (1984) article. In fact I only realized it about a year after the article was published. It apparently escaped the notice of all the reviewers of both manuscripts. Still, that mistake has different implications for the two arguments. Baumeister and Steinhilber found significant effects despite using this analysis strategy, which provided a weak test, and, given that significance, the weakness of the analysis strategy was not all that important. In contrast, Schlenker et al. tried to prove the null hypothesis, and to do that (if it can be done at all), they should have done everything possible to avoid using a weak test or introducing random error variance into the analysis. Before concluding that an effect is not there, one seemingly ought to conduct the most sensitive and proper analyses.

A recent effort by Butler (1994) was unfortunately inconclusive. Butler tallied all World Series games (including sweeps, repeat champions, etc.) from 1924 to 1993. Home teams have won 50% of the games in which there was the possibility of winning the championship, as compared with 57% of all other games. These percentages fit the view that the usual home-team advantage disappears when one is on the brink of the championship, but the difference was not significant.

One additional point regarding sensitivity of analysis techniques is relevant. The use of simple game outcome—won or lost—is dichotomous and thus could well fail to detect substan-

tial patterns of choking, just as most laboratory studies might find nonsignificant results if they relied on dichotomous measures alone. If a team wins all its World Series games by six or seven runs but then wins only by one or two runs in the final game (as happened in the 1987 World Series, for example), this might reflect a pattern of choking, but it would not show up in an analysis that counted only the game outcome.

Lastly, Baumeister and Steinhilber (1984) analyzed the results of the sixth games in championship series because these games unconfound the pressure arising from imminent, possible victory from the pressure of facing elimination. Schlenker et al. (1995) found precisely the same results, namely that home teams win such games when on the brink of elimination but lose them when on the brink of becoming champions. Ironically, however, Schlenker et al. discussed this replication of results as if they had found something different and even contradictory. Thus, when they found evidence that fit Baumeister and Steinhilber's hypothesis, they unfortunately obscured this fact by the way they described it.

To summarize, despite using an imprecise, weak analysis strategy and including data that were confounded by rule changes (which they did not mention), Schlenker et al. (1995) still found results that did not differ significantly from those found by Baumeister and Steinhilber (1984). They merely found that the effect dropped below significance. This does not seem to justify a sweeping conclusion that the home-field disadvantage hypothesis is wrong.

#### Implications for Future Research With Archival Methods

Of broader interest are the implications about the role of archival research in social psychology. I suspect that Schlenker et al. (1995) had the same experience as Baumeister and Steinhilber did, namely that the initial excitement of working with nonlaboratory, supposedly more real phenomena was soon replaced by a chagrined realization of how many ambiguous circumstances and arbitrary decisions accompanied the use of such data. Thus, regarding the present research: Should defending champions be excluded, because there is no identity change (cf. Wright et al., in press)? Should semifinal series be included? Should four-game sweeps be included? How far back in time should one go? Many of these issues have no clear theoretical relationships to the hypotheses. Although manipulation of such factors might move a finding back and forth across the magical .05 criterion, even the best intentioned researcher might find no solid *a priori* basis for making these decisions, and often it is difficult to make or even anticipate them all before knowing how they will affect the hypothesis. In contrast, the similar decisions regarding a laboratory experiment (e.g., whether to exclude a participant who seems intoxicated) are much rarer and can be made before data collection.

One conclusion, therefore, is that archival data in general may be less than optimal for testing and building psychological theories. I certainly agree with Schlenker et al. (1995) to the effect that Baumeister and Steinhilber's (1984) archival study has received undue attention in the literature and textbooks. The laboratory and field research I published the same year (Baumeister, 1984) was far more solid methodologically, yet it

has not received as much attention. The laboratory remains the best place to test and build theories.

On the other hand, it does seem desirable that social psychologists venture out of the laboratory periodically to look for converging evidence. If nothing else, this can help respond to critics who airily assert that the effects social psychologists find in the laboratory will be found only in the laboratory. Archival studies can play an important role there. Undoubtedly the attention that has been paid to Baumeister and Steinhilber's (1984) research is partly attributable to the fact that it showed effects on salient, familiar, nonlaboratory behaviors. It seems essential that such studies are run and published. Perhaps the best compromise is that these should be regarded as extending, illustrating, and confirming laboratory studies rather than as primary, direct tests of theory.

One advantage of laboratory research is that competing factors and processes can be excluded. A field or archival study might fail utterly to find evidence of a causal hypothesis even if the hypothesis is correct.<sup>1</sup> Of course, such random events can work either to the advantage or the disadvantage of any given hypothesis, and so archival studies may be especially susceptible to spurious findings; then again, randomness and multiple causality will generally increase error variance, which will reduce significance levels, thereby making archival studies especially unlikely to yield significant results.

The championship choke by home teams is a good example. Even if the hypothesis is fully correct, it is hardly strong enough to overcome all other factors, such as injuries or umpire errors. On the basis of Baumeister and Steinhilber (1984), one may conclude that only about one out of every five championship series has its outcome altered by this pattern of choking. Schlenker et al.'s (1995) data suggested a figure around one out of every eight or nine. In hockey, Wright et al. (in press) suggested about one of every three eligible series is affected, after invoking a series of statistical controls (e.g., deleting defending champions). In golf, Wright et al. (1991) found an effect on the order of about one-half stroke per 18 holes. Clearly, the effect is not so strong as to overwhelm all the many other variables that can enter into determining the outcome. Moreover, I (Baumeister, 1995) have suggested that a simple rule change may have given home teams an advantage in the World Series that was sufficient to eliminate any sign of a home choke, at least if one stays with the relatively insensitive (dichotomous) measure of winning versus losing the game.

Given causal complexity of real-world phenomena, it should be relatively easy to fail to find any given effect. I have faulted Schlenker et al. (1995) for ignoring relevant rule changes, for

<sup>1</sup> There may be other boundary conditions too. As players orient more toward the national television audience, the effect of the immediate presence of the home audience might be muted; visitors may choke too. Also, the theory about choking under pressure pertains only to skilled performances (Baumeister, 1984). When effort or stamina is the main factor, one would expect entirely different patterns of results (see Baumeister et al., 1990), so I would not expect a home choke in weightlifting or marathon running. It is even plausible that no effect would be found in soccer; although there is plenty of skill involved in soccer, it is arguable that the decisive factor (especially in a grueling tournament) is effort versus fatigue.

incorporating irrelevant games that added error variance to the analysis, and for relying on an insensitive measure (game outcome). Yet none of these was an outrageous or theoretically incongruous procedure. I would not be criticizing their procedures if their results had confirmed ours (and indeed much of what they did paralleled that of Baumeister & Steinhilber, 1984). The difference may have been that Baumeister and Steinhilber were looking to support the hypothesis, whereas Schlenker et al. were looking to reject it. It is unfortunately true that researchers can design their empirical investigations—laboratory studies as well as archival studies—to maximize or to minimize their chances of finding a significant effect.

Because of this problem, precisely focused comparisons should be used whenever possible. If Schlenker et al. (1995) wanted to conclude that home teams enjoy an advantage in the last game of the World Series, they should have shown that the percentage of such games won by home teams was significantly higher than 50%—a logical impossibility in this case, because 47% (their observed percentage) cannot be significantly higher than 50%. Instead, they merely lumped all games together and said that home teams did better overall, which is what Baumeister and Steinhilber (1984) found too. The existence of a general home-field advantage is not in dispute; only the effect on final games is at issue. By mixing relevant and less relevant data, one can come up with an overall impression that conceals important exceptions.

A general rule in scientific work is that positive findings are regarded as more meaningful than null findings, and this rule should probably be especially strongly applied to archival and other nonlaboratory studies. I think this would dovetail well with a policy of using such studies as converging evidence rather than as primary tests of hypotheses. Findings from archival work, even more so than laboratory work, should perhaps be understood as raising possibilities more often than producing definite conclusions about theoretical relationships.

Another intriguing possibility is that both Baumeister and Steinhilber (1984) and Schlenker et al. (1995) were correct in an important sense: Maybe home teams really did choke during the 1950s, 1960s, and 1970s but ceased to do so during the 1980s. Schlenker et al. were appropriately skeptical of their own findings about differences in performance by eras, but it is hard to deny that the American professional sports scene has changed dramatically in recent decades, and it would be surprising if these changes had no effect on performance patterns. It is clear (from laboratory work; Baumeister, 1984; Heaton & Sigall, 1991) that there are strong individual differences in the propensity to choke under pressure, and it is quite plausible that the top professional athletes of the 1980s and 1990s—expecting regular national television audiences, earning multimillion-dollar salaries, accepting roles as mass media superstars and celebrities, and being much more disproportionately drawn from ethnic and racial minorities—may be of a more highly self-aware group than their predecessors, which would reduce their vulnerability to choking (Baumeister, 1984). Such a change would be reflected in a lesser pattern of choking across all skill sports and all settings, not just championship games, but it too is a possibility that deserves to be considered.

In this regard, I was especially disappointed by the suggestion by Schlenker et al. (1995) that Baumeister and Steinhilber's

(1984) results were due to a fluke of randomness akin to getting a few dozen consecutive coin flips to come up tails. To be sure, the significant findings of any article in this journal could be attributed to such a deviation from chance, but I think that such interpretations should be a last resort, especially because they are so improbable.<sup>2</sup> (After all, one could just as plausibly suggest that the success of home teams over the past decade, which is crucial to Schlenker et al.'s argument, is the random fluke.) With archival work, the probability of cultural and historical change seems a much more likely and defensible explanation than some (by definition) highly improbable deviation from statistical normalcy. In sports, there are rule changes (as I already pointed out with baseball) as well as changes in self-selection processes and in the socialization of athletes. These seem the most likely places to look.

My point is that archival research will find it difficult to establish or reject any of these possibilities conclusively. However, archival data can suggest possible patterns and indicate parallels to causal processes from laboratory work, and they can also point directions for further experimentation. Given the conflicting conclusions regarding championship choking, it appears that this particular line of work will need further laboratory work before firm theoretical answers are available.

In my view, the most valuable contribution made by Schlenker et al. (1995) was their new data on the timing of fielding errors. Their findings, coupled with prior laboratory studies (Baumeister et al., 1985; Heaton & Sigall, 1991) and other archival work (Benjafield, Liddell, & Benjafield, 1989) suggest a refinement of choking theory to emphasize the concern about failing to live up to inflated expectations. When one reaches the verge of a major championship, positive identity change, or other success, one's supporters may come to want and expect it to happen, and concern over the possibility of letting them down at the last minute may prove detrimental to skilled performance. This does not, however, contradict the championship choke; indeed, it may explain it.

<sup>2</sup> Thus, Schlenker et al. (1995) suggested that a 19-year-long random fluke yielded a .182 average despite a .600 true probability, which seems so unlikely that it would occur only once in many centuries.

## References

- Baumeister, R. F. (1984). Choking under pressure: Self-consciousness and paradoxical effects of incentives on skillful performance. *Journal of Personality and Social Psychology*, 46, 610–620.
- Baumeister, R. F., Hamilton, J. C., & Tice, D. M. (1985). Public versus private expectancy of success: Confidence booster or performance pressure? *Journal of Personality and Social Psychology*, 48, 1447–1457.
- Baumeister, R. F., Hutton, D. G., & Cairns, K. J. (1990). Negative effects of praise on skilled performance. *Basic and Applied Social Psychology*, 11, 131–148.
- Baumeister, R. F., & Steinhilber, A. (1984). Paradoxical effects of supportive audiences on performance under pressure: The home field disadvantage in sports championships. *Journal of Personality and Social Psychology*, 47, 85–93.
- Benjafield, J., Liddell, W. W., & Benjafield, I. (1989). Is there a home field disadvantage in professional sports championships? *Social Behavior and Personality*, 17, 45–50.

Butler, J. (1994). [World Series game outcomes]. Unpublished calculations.

Heaton, A. W., & Sigall, H. (1991). Self-consciousness, self-presentation, and performance under pressure: Who chokes, and when? *Journal of Applied Social Psychology, 21*, 175-188.

Schlenker, B. R., & Leary, M. R. (1982). Social anxiety and self-presentation: A conceptualization and model. *Psychological Bulletin, 92*, 641-669.

Schlenker, B. R., Phillips, S. T., Boniecki, K. A., & Schlenker, D. R. (1995). Championship pressures: Choking or triumphing in one's own territory? *Journal of Personality and Social Psychology, 68*, 632-643.

Wright, E. F., Jackson, W., Christie, S. D., McGuire, G. R., & Wright, R. D. (1991). The home-course disadvantage in golf championships: Further evidence for the undermining effect of supportive audiences on performance under pressure. *Journal of Sport Behavior, 14*, 51-60.

Wright, E. F., Voyer, D., Wright, R. D., & Roney, C. (in press). Supporting audiences and performance under pressure: The home-ice disadvantage in hockey championships. *Journal of Sport Behavior*.

Received July 7, 1994  
 Revision received August 29, 1994  
 Accepted November 3, 1994 ■



**AMERICAN PSYCHOLOGICAL ASSOCIATION  
 SUBSCRIPTION CLAIMS INFORMATION**

Today's Date: \_\_\_\_\_

We provide this form to assist members, institutions, and nonmember individuals with any subscription problems. With the appropriate information we can begin a resolution. If you use the services of an agent, please do **NOT** duplicate claims through them and directly to us. **PLEASE PRINT CLEARLY AND IN INK IF POSSIBLE.**

PRINT FULL NAME OR KEY NAME OF INSTITUTION	MEMBER OR CUSTOMER NUMBER (MAY BE FOUND ON ANY PAST ISSUE LABEL)	
ADDRESS	DATE YOUR ORDER WAS MAILED (OR PHONED)	
CITY STATE/COUNTRY ZIP	<input type="checkbox"/> PREPAID <input type="checkbox"/> CHECK <input type="checkbox"/> CHARGE CHECK/CARD CLEARED DATE: _____	
YOUR NAME AND PHONE NUMBER	(If possible, send a copy, front and back, of your cancelled check to help us in our research of your claim.) ISSUES: <input type="checkbox"/> MISSING <input type="checkbox"/> DAMAGED	
TITLE	VOLUME OR YEAR	NUMBER OR MONTH

*Thank you. Once a claim is received and resolved, delivery of replacement issues routinely takes 4-6 weeks.*

(TO BE FILLED OUT BY APA STAFF)	
DATE RECEIVED: _____	DATE OF ACTION: _____
ACTION TAKEN: _____	INV. NO. & DATE: _____
STAFF NAME: _____	LABEL NO. & DATE: _____

Send this form to APA Subscription Claims, 750 First Street, NE, Washington, DC 20002-4242

**PLEASE DO NOT REMOVE. A PHOTOCOPY MAY BE USED.**