Simulation & Gaming

http://sag.sagepub.com

Why the Best Person Rarely Wins: Some Embarrassing Facts About Contests

Warren THORNGATE and Barbara Carroll

Simulation Gaming 1987; 18; 299
DOI: 10.1177/104687818701800301

WHY THE BEST PERSON RARELY WINS

Some Embarrassing Facts About Contests

WARREN THORNGATE
BARBARA CARROLL
Carleton University

One of the most popular methods for deciding between competing candidates for a role, office, or other distinction is to hold a contest. In general, a contest is a set of procedures designed to eliminate all but one winning contestant by assessing selected features or performances of each. Contests are ubiquitous. They are held to determine superior athletes or sports teams, to choose political representatives, to select people with the most intelligence, talent, knowledge, beauty, skill, or "potential," to discover exceptional flowers, automobiles, research grants, building designs, or works of art. It is reasonable to assume that almost every member of a modern society will be a contestant in several contests during his or her lifetime. A large proportion of us will engage in hundreds of contests, knowingly and willingly or otherwise. As the number of people competing for coveted positions in a society increases, so too does the number of contests held to find the winners. Contests are a growth industry.

As contests have proliferated, so too have the rules and

Author's Note: *We wish to thank Ruth Currie for her assistance in the preparation of this manuscript. Reprint requests should be addressed to Dr. Warren Thorngate, Psychology Dept., Carleton University, Ontario, Canada K1S 5B6.*

SIMULATION & GAMES, Vol. 18 No. 3, September 1987 299-320

procedures for conducting them. There are now scores of different procedures for holding a contest, procedures with names such as round (compulsory, "free style," consolation, robin), challenge, runoff, playoff, playdown, season (pre, post, and regular), handicapping, seeding, preliminaries, semifinals, and finals. There are at least as many methods for assessing contestants: consensus, voting (for one, two, three, or more contestants), ranking, averaging judgments of experts, and measurement, including time, distance, weight, attendance, hardship, wealth, volume of laughter, or applause. The combinations of contest procedures and assessment methods seem almost endless; hundreds of them are now employed with little apparent consideration of their strengths and weaknesses.

Indeed, it is surprising that almost no research has been undertaken to determine the relative merits of alternative contest procedures. There is, of course, a large body of research on voting procedures (see Arrow, 1951; Barry & Hardin, 1982; Brams & Fishburn, 1983) that has demonstrated the perils of aggregating individual preferences. But it has not addressed other important aspects of contests, aspects such as the sensitivity of contest structures to contest outcomes. Given a particular contest structure and method for assessing contestants, what are the chances that the best contestant will win? The present research was undertaken to answer this question.

Many contests, and most of those with large numbers of contestants, are hierarchically structured. They begin with "local" competitions; winners of these competitions advance to "regionals"; regional winners compete in "provincials"; provincial winners compete in national finals; national winners advance to international matches. Such hierarchies are characteristic of several amateur sports contests as well as music performance competitions. Their structure is usually justified on pragmatic or logistic grounds. It is costly to gather thousands of contestants and judges together for several rounds of open competition; a hierarchical contest assures that large numbers of contestants will be quickly eliminated by a few judges, and thus reduce the time and effort necessary to select a winner.

Yet judges are fallible, and even the best contestant is bound to have "off" days. The unreliability of judges' performance assessments, and the unreliability of contestants' performances, make it possible that the best contestant may lose a round of the contest. If this occurs, the contestant is out of the running for remaining rounds; hierarchical contests are quite unforgiving in this regard. The chances that "bad luck" will eliminate the best contestant may be very small in preliminary rounds. But as the contest proceeds, winners of new rounds should be increasingly similar in their talent, and small factors (one biased or erroneous judgment, one wrong note or awkward movement) would be expected to be increasingly important in determining the contest outcome.

Many hierarchical contests display tacit recognition of the possible contribution of "error" to outcome. The Canadian Music Competition, for example, averages the assessments of two judges to select local contest winners, and five judges to assess provincial winners, presumably to increase the reliability of judgments in the later contests. Gymnastic and other "rated by expert" competitions often cast off the highest and lowest ratings of judges on their panels then average the remainder, presumably to attenuate both forms of punitive bias. Several contests average performance scores of contestants across rounds, presumably to increase the "test-retest reliability" of assessments. Many other contests require contestants to perform increasingly difficult feats as they progress from one round to the next, presumably to avoid the loss of discriminant validity as the average talent of remaining contestants increases and the range of their talent is reduced. Still other contests attempt to "seed" better contestants in preliminary rounds, presumably to reduce the possibility that the best contestants will face each other in a small local competition and "prematurely" eliminate all but one of their kind.

Such attempts to reduce the insidious effects of error are admirable, but the extent of their success is not precisely known. In a contest with 1,000 contestants, we would certainly hope that the contest procedures would select the best contestant with a much higher probability than 1/1000. Ideally, of course, the best person should be selected with probability 1. But what if the

procedures could only guarantee a 50% chance of selecting the best contestant? Or only a 10% chance? We may then wish to search for procedures with higher validity.

Clearly it is desirable to determine the relationships between the structure of a contest, the number of contestants, the amount of error present in judgments or performances, and the chances that the best person will win. What little we now know of these relationships is not encouraging. Einhorn (1978) has shown that one simple form of a contest, a one-round competition for welfare, can frequently lead to judgment errors when the number of welfare applicants greatly exceeds the amount of available welfare funds, even if the assessments of "deservingness" are almost infallible. His research suggests that when a large number of contestants are vying for the scarce commodity "winner," very small judgment or performance errors can drastically reduce the chances that the best person will win. In view of the importance of many contests to the careers and welfare of contestants, it seems judicious to determine if this suggestion is correct, and to examine similar relationships for other contest structures.

In order to determine the relationship between various contest parameters and the chances that the best contestant will win, we ran several thousand Monte Carlo simulations of contests that varied in (1) the number of original contestants, (2) the number of elimination rounds, (3) the amount of random judgment error that added to, or subtracted from, each contestant's performances, and (4) the extent of "seeding" of good contestants in preliminary rounds. Changes in the values of these four parameters were then plotted against the chances that the best contestant in each contest would emerge as the winner.

METHODS

We began each contest in our simulations by assigning a *true talent score* (TTS) to each simulated contestant. In the interests of speed and economy, these TTSs were drawn at random from a triangular distribution that had the same "humped" characteristic

as the normal distribution but which ran much faster. Each TTS was computed by adding two random numbers drawn from rectangular distribution. Each random number ranged from 0.0 to 50.0, so each TTS ranged from 0.0 to 100.0. In this way, the triangular distribution of TTSs was formed—very few contestants had a TTS above 90 (2.9%) or below 10 (2.9%); most (67%) had a TTS within 20 points of the average (50.0).

Each time a contestant "performed," the resulting *performance score* (PS) was calculated by adding a random error term to the TTS of the contestant. This random error term was computed by drawing a number at random from another triangular distribution. The range of this distribution varied from simulation to simulation. In one set of simulations, the error term ranged from –2.5 to +2.5 (*error range*, ER = 5), in another set from –5.0 to +5.0 (ER = 10), in a third set –7.5 to +7.5 (ER = 15), and so forth, increasing by units of 5.0 to the maximum ER = 30. Thus a contestant with a TTS = 80.0 who performed in several contests with ER = 10 might accrue PSs of (80.0 + 1.2) = 81.2, (80.03–3.6) = 76.4, (80.0 + 0.9) = 80.9, and so forth.

The "true-score-plus-error" calculations of PSs were in keeping with the traditional additive model of psychological measurement. We assumed that a contestant's TTS was his/her true talent score, and that the "best" contestant was that person with a higher TTS than any other contestant. We further assumed that the error score added to the true score was analogous to transient effects of a contestant's mood, emotion, health, and so forth, as well as similar effects in contest judges. Over many contests, of course, these transient effects would be expected to cancel each other out and reveal an average PS that would be increasingly close to a contestant's TTS. In any single contest, however, the amount of error variance could be instrumental in determining the contest outcome. Almost all contests, and almost all of our simulations, select the person with the highest contest PS as winner. Thus a contestant with an TTS = 90.0 may, in some contest, give a performance scored as PS = 88.7. Another contestant in the same contest may have an TTS of 86.0 but give a performance scored as 89.6. As a result, the second contestant, not the best person, would win.

We suspected that many people familiar with psychological measurement might be confused by an error index reported in terms of ranges. We therefore ran several Monte Carlo simulations to estimate empirically the relations between ERs S.D.(ER), and the resulting test-retest reliability, *r*, of the PSs. The results are shown in Table 1. The *r*s in Table 1 are quite high by psychometric standards; most psychologists would be thrilled to develop a test of any trait or activity that obtained a test-retest reliability above + 0.90. We believe, however, that they are reflective of average PSs given by panels of judges in such contests as gymnastics and music. For example, in a recent Canadian Music Competition performance (provincial round), one of the candidates received the following scores from five judges: 89, 88, 84, 83, 80. Her average score (84.8) would thus be expected to have a sampling distribution with a SD = 1.7, or an ER of about 10.

Contest simulations were run on two different computers (a Radio Shack Color Computer, and an Apple MacIntosh). The simulation programs were written in two versions of BASIC (Microsoft BASIC and BASIC09) and two versions of FORTH (Color FORTH and MACFORTH). This allowed for some control over any quirks of the pseudo-random number generators available on a particular language implementation. No noticeable differences were found in the simulation results obtained from equivalent programs written in the four language versions. The results reported below were obtained from simulations run under BASIC09. A new random number seed was generated for each simulated contest.

SIMULATION I

The first contest simulation contained no hierarchy. It was run to generate results for three sets of control or baseline comparison groups. For each run of the simulation, a fixed number of contestants (1,000 or 100 or 10) was first assigned individual TTSs. Each contestant then "performed" once and obtained an

TABLE 1
Relation Between Error Range (ER), S.D. (ER), and Test-Retest Reliability (r)

ER	5.0	10.0	15.0	20.0	25.0	30.0
S.D.	1.05	2.1	3.15	4.2	5.25	6.3
r	+0.997	+0.991	+0.978	+0.961	+0.941	+0.925

individual PS by the above-mentioned method of adding a random error term to the TTS. The contestant obtaining the highest PS was declared the contest winner. A check was then made to see if this winner was, in fact, the contestant with the highest TTS.

This contest procedure was repeated 1,000 times for each error range (ER = 5, 10, 15, 20, 25, 30), and each contest size (number of contestants, N = 1000, 100, 10). In total 18,000 contests were run in total, with a new set of TTSs and PSs generated for each. A count was made of the number of contests in which the winner was the contestant having the highest TTS. Eighteen such counts resulted, six for each of the three levels of N, and these were interpreted as estimates of the probabilities that the best person would win. The results are shown in Figure 1.

As can be seen in Figure 1, increases in judgment error lead to reductions in the chances of the best person (highest TTS) winning. When the contest has relatively few participants (e.g., N = 10), the effects of error are relatively small. Even so, there is a 5% chance that the best person will not win when the error is miniscule (ER = 5), and almost a 25% chance when the error is somewhat more noticeable (ER = 30). When the contest has many participants (e.g., N = 1,000) the effects of error are quite dramatic. An ER as small as 10 appears to give the best person less than a 50% chance of winning. An ER of 30 gives the best person less than one chance in 6. Recall that these ERs reflect rather high test-retest reliabilities. Our hunches about the relevance of Einhorn's (1978) results thus appear to be confirmed: At least in "free-for-all" contests with large numbers of contestants,

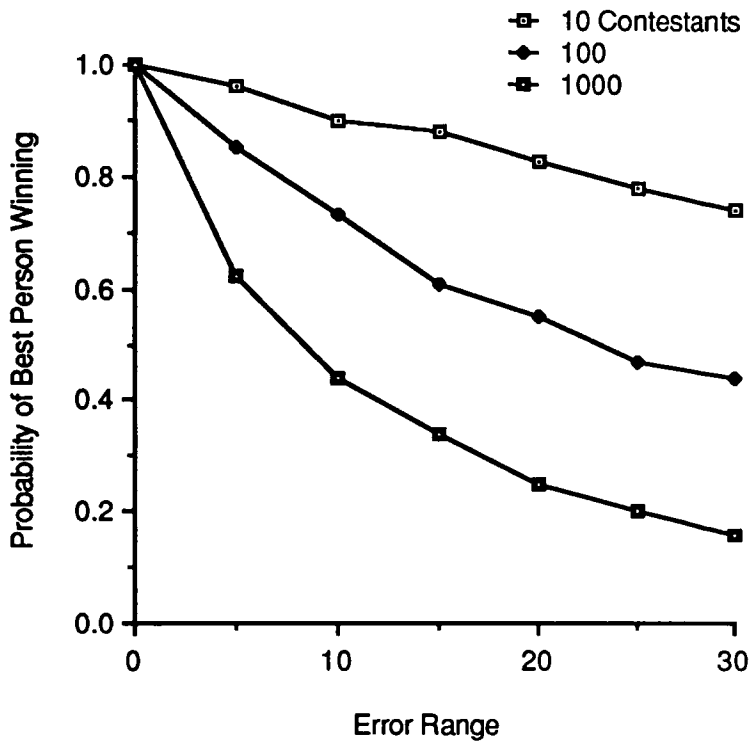

Figure 1: Probability of Best Person Winning in Nonhierarchical Contests

even slightly fallible judgments can greatly increase the chances of omission errors.

SIMULATION II

It is often argued that one-shot contests, such as those run in Simulation I, do not give contestants enough opportunity to show their talents, and that a single shot of bad luck for the best person may combine with a single shot of good luck for some other contestants to produce an "unfair" contest outcome. To reduce the chances of this possibility, it is often suggested that

contestants engage in two or more performances to attenuate the chances that one shot of good or bad luck will determine the outcome. Yet if a contest has many contestants, the prospect of judging all of them several times can become daunting. This possibility can be avoided if only the best contestants in one contest are allowed to proceed to the next contest. Thus hierarchical elimination contests are born.

Do contest hierarchies increase the chances that the best contestant will win? In order to find out, we ran the second contest simulation in hierarchical fashion. Two sets of simulation runs were performed, one as a two-round hierarchy, and one as a three-round hierarchy. In the first, 100 original contestants were randomly divided into groups of 10; a "local" contest was held for each of these groups, and the resulting 10 winners were then run in an eleventh and final contest to determine the final winner. In the second, 1,000 contestants were first randomly divided into groups of 10 to engage in local contests, the 100 local winners were then randomly divided into groups of 10 to engage in "regional" contests, and the 10 regional winners engaged in a single final contest.

As in Simulation I, 1,000 contest sets were simulated for each of the 12 combinations of hierarchy (two vs. three) and error range (ER = 5, 10, 15, 20, 25, 30). A record was kept of the probability that the person with the highest TTS won the final. The results are shown in Figure 2.

A comparison of Figures 1 and 2 reveals that, contrary to popular belief, the establishment of a hierarchical elimination procedure produces no noticeable improvement in the chances that the best person will win. The probabilities of the best person winning either one-shot or elimination contests with 100 and 1,000 contestants are virtually identical for each error range.

SIMULATION III

It is sometimes argued that local contests with few contestants have an unfair advantage over local contests with many contes-

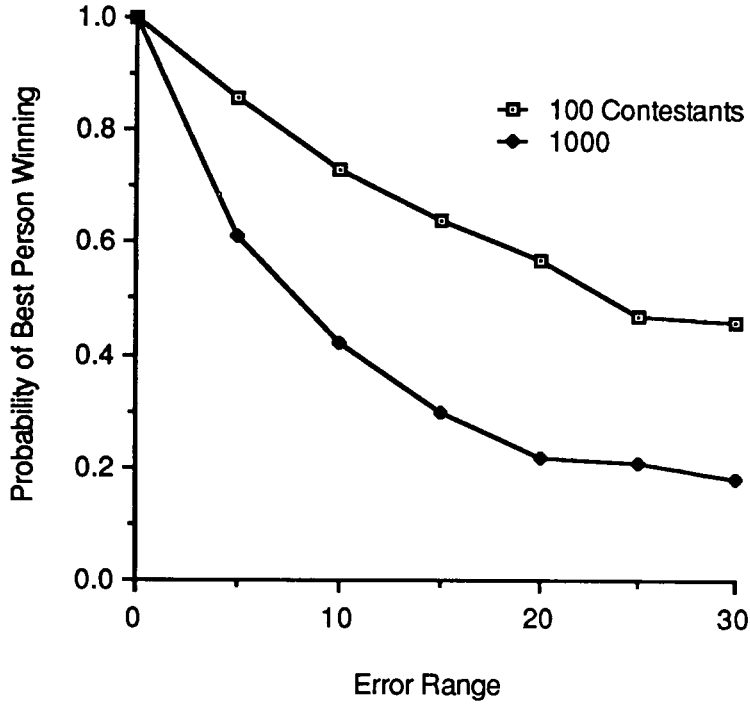

Figure 2: Probability of Best Person Winning in Hierarchical Contests

tants: The competition in the former may be less severe than in the latter; the former may produce "big fish in small ponds," while the latter may eliminate many "bigger fish" from their "bigger ponds." As a result, relatively mediocre people from small local contests may be allowed to proceed to subsequent rounds, and relatively good people from large local contests may be eliminated because they are competing against more contestants. We designed two variants of the two-round ($N = 100$) contests run in Simulation II to test whether or not variations in the sizes of contests would affect the chances of the best person winning. In the first variant, four contests were held in the first round: a 40-person contest, a 30-person contest, a 20-person contest, and a 10-person contest. The four winners were then pitted against each

other in a final competition. In the second variant, only two contests were held in the first round: One had 90 contestants, the other had 10. The two winners competed in the final round. As before, 1,000 contest sets were run for each of these two variants, and at each of the six error ranges.

Neither variant had any noticeable effect upon the probability of the best person winning; both produced outcome probabilities that were indistinguishable from those shown for N = 100 in Figures 1 and 2. It is true that local contests with relatively large numbers of contestants generally produce winners with higher TTSs than locals with relatively small numbers of contestants. As a result, the "big pond" winners generally win subsequent contests against "small pond" winners. However, it does not seem to affect the chances that the best person will win.

SIMULATION IV

It is still possible that contests with equal numbers of contestants may have widely different talent scores. For example, one local music contest may be held in a large city and attract 10 contestants enrolled in its world-renowned music conservatory; another local contest may be held in a small village and attract 10 contestants who wandered into the neighborhood music store. In this case, most of the better contestants might be concentrated in one local competition, and the best one—by a very small stroke of bad luck—may never proceed to subsequent rounds. It is this possibility that makes many a contest organizer consider "seeding" the better contestants in preliminary contests in order to increase the number of better people who make it to the finals.

But does seeding influence the chances that the best person will win all rounds? In order to find out, we ran another two sets of contests (10 contestants in each of 10 contests for the first round; one contest with the 10 winners for the final) with vastly different seeding arrangements. In the first set, the 10 contestants with the highest TTSs were seeded one to each local contest; this insured that each local would have one "good" contestant. In the second

set, the 10 contestants with the highest TTSs were all placed in one local; this insured that 9 of the 10 best would be eliminated in the first round. Once again, we ran 1,000 simulated contests for each of the 12 combinations of seeding (distributed vs. lumped) and error range.

Here too, we found no significant effect of seeding on the chances that the best person would win. The relationship of error range to the probability of the best person winning was identical to that shown in Figures 1 and 2 (N = 100). With distributed seeding, the 10 most talented people did tend to proceed to the final; with lumped seeding, the final was almost always won by the winner of the one talented local. But the chances that the most talented person would walk away with first place in the final was affected only by the error range of the judged performances, and not by the seeding arrangement.

SIMULATION V

Not all contests are one-shot affairs or hierarchical elimination matches. Another form of contest, popular in boxing matches and daytime quiz shows, selects winners by sequential challenges. In simplest form, two contestants engage in a match. The winner is then challenged to another match by a third contestant. The winner of this match is challenged by a fourth contestant. The contest proceeds in this way until all contestants have engaged in at least one match. The contest winner is the person who remains when the challenges end.

How reliable is this winner-challenger procedure in selecting the best contestant? Once more we ran 1,000 100-person contests for the six ER levels, and counted the number of times the best person won. Each of 6,000 contests was run using the winner-challenger procedure described above; each match was decided by adding a new random error term to the current winner's TTS, adding a new error term to the current challenger's TTS, then noting which of the two resulting PSs was higher.

Once again, we obtained a now-familiar outcome: The probabilities that the person with the highest TTS would win were the same as to those shown in the $N = 100$ graphs of Figures 1 and 2. Here too it was the size of the error term, and not the structure of the contest, that affected outcome probabilities.

SIMULATION VI

Because the amount of error seemed so important in determining the probability of the best person winning, we decided to see just how much this probability could be raised by adopting one prototypical method for reducing the error: averaging contestants' PSs across contests. We ran 12,000 hierarchical contests using the same experimental design described in Simulation II. This time, however, second round (for $N = 1{,}000$) and final round scores of a contestant were calculated by averaging the current round PS with her or his previous PSs. Thus a contestant with a TTS of 90 who won the first round with a PS = 89.2 and attained a PS = 93.6 on the second round would be given an average contest PS of $(89.2 + 93.6)/2 = 91.4$. If this contestant's average PS = 91.4 exceeded the nine average PSs of the remaining contestants, she or he would be declared the winner for the round. If a third round were held, and the same contestant attained a PS = 88.8, his or her contest score would be $(89.2 + 93.6 + 88.4)/3 = 90.4$. This contest score would again be compared to those of the other nine finalists, and the highest of the 10 would be declared the contest winner.

The results of these "running average" simulations are shown in Figure 3, along with the comparable results obtained in Simulation II. As expected, averaging did increase the chances of the best person winning; when $N = 100$, the probability increased by about 0.08, when $N = 1000$ it increased by about 0.12. It is heartening to know that such increases are possible. But for $N = 1000$ and an ER ≥ 15 the best person still has a greater than even chance of losing.

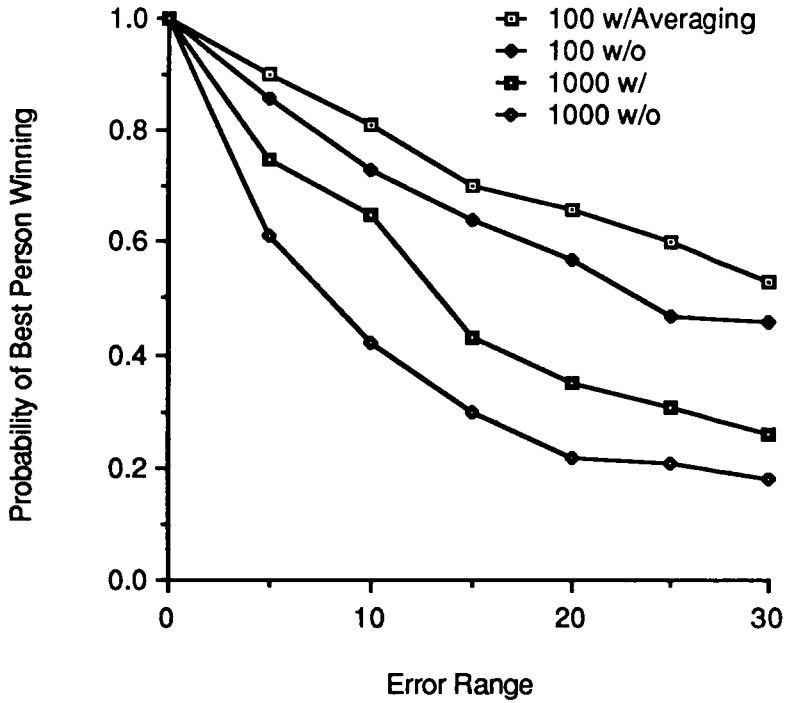

Figure 3: Probability of Best Person Winning in Hierarchical Contests with Averaged Performance Scores

SIMULATION VII

In our introduction we speculated that the effects of error would be more pronounced in later rounds of a hierarchical elimination contest than in preliminary rounds. Each time a round of contests is held, the winners should have a higher average TTS than do the losers. More importantly, however, the winners should have less variability between their TTS than do the losers. Because the variability of their TTSs is reduced, any fixed amount of error should magnify its effects. For example, an ER of 5 may have minimal effects if the TTSs of first-round contestants range from 0 to 100, but if the TTSs of finalists range

only from 90 to 100, an ER of 5 may have a large effect on who the winner will be.

In order to verify and measure the reduction in variability that might occur in large hierarchical contests, we conducted a small and final simulation. Ten more three-stage, $N = 1000$, contests were run in the same manner of Simulation II for each of three ERs: ER = 0, ER = 15, and ER = 30. This time we kept track of the distributions of TTSs prior to each contest round, and computed the means and standard deviations of these distributions. Histograms of these distributions, their means and S.D.s, are shown in Figures 4a, 4b, and 4c.

As expected, the well-spread "prior" TTS distributions at the left of Figures 4 are transformed into highly skewed TTS distributions at the right as the result of the contests held. However, the addition of error attenuates this transformation. Scores of the finalists shown in Figure 4c, for example, don't have the kurtosis of those in Figure 4a: when the contest is error-free (ER = 0, Fig. 4a), 88% of the finalists have scores of 90 and above; with errors present, only 62% (ER = 15, Fig. 4b) to 28% (ER = 30, Fig. 4c) of the finalists demonstrate this amount of true talent. And what happened to all the other contestants with TTSs of 90+? All of them were victims of bad luck.

The extent to which luck determines contest outcomes does, indeed, seem to increase with number of previous contests. To illustrate, consider again the distribution shown in Figure 4b. Prior to the first round, the ratio of true talent to luck—of signal to noise—was [S.D.(TTS/S.D.(ER)] = [20.1/3.15] = 6.4. The 100 winners of the first round (i.e., the semifinalists) had a smaller range of talent, so in the semifinal contests the ratio of talent to luck was [10.8/3.15] = 3.4. The 10 winners of the semifinals had an extremely restricted talent range, so here the ratio of talent to luck was [5.2/3.15] = 1.6. A little extrapolation should show that once the error range reaches about 50, luck will have equal importance to talent in determining the final winner for contests of this size. Performance scores with this error range are not implausible. They correspond to a test-retest r of about + 0.80—still a passable reliability, and a good validity, by psychometric standards.

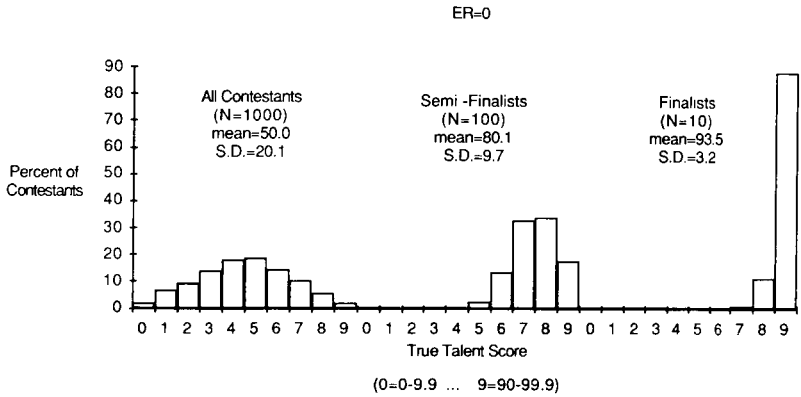

Figure 4a: Distribution of True Talent Scores in Three-Round Hierarchical Contests with ER = 0

DISCUSSION

The results of our assorted contest simulations suggest three general conclusions. First, as the number of contestants increases, so too does the importance of luck (random errors of contestants or their judges) in determining a contest winner. Second, the structure of a contest—whether hierarchical, winner-challenger, seeded, or free for all—makes no noticeable difference in the overall chances that the best person will win, but in hierarchical elimination contests, the effects of luck increase with each elimination round. Third, methods such as averaging individual performance scores over rounds in a hierarchical contest can attenuate the effects of luck, but cannot eliminate these effects.

These conclusions may be viewed from several different perspectives. From a contestant's perspective, the conclusions provide both good news and bad. It may be comforting for a losing finalist of a large hierarchical contest to know that luck really does play a large role in the selection of a winner. It may be humbling for the winner of the same contest to know there is a substantial chance he or she was not the most talented contestant.

At the same time, the increasing importance of luck in

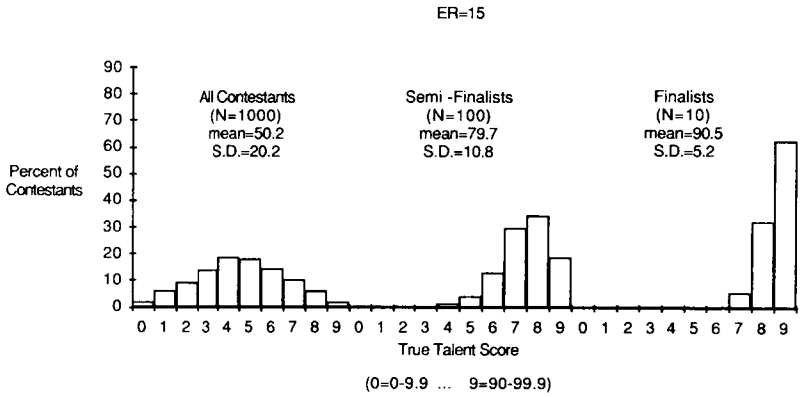

Figure 4b: Distribution of True Talent Scores in Three-Round Hierarchical Contests with ER = 15

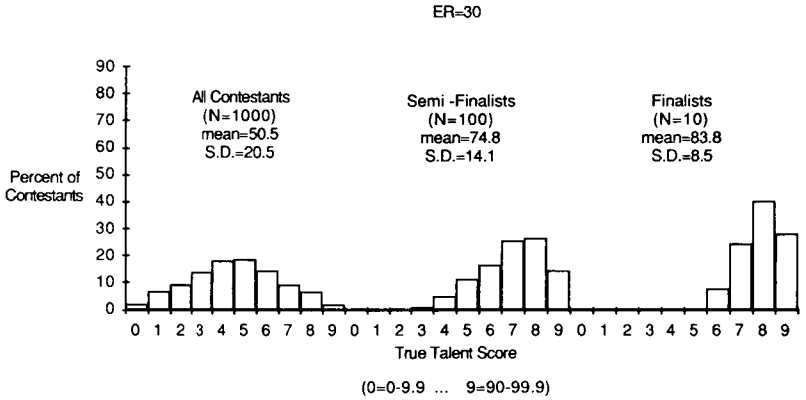

Figure 4c: Distribution of True Talent Scores in Three-Round Hierarchical Contests with ER = 30

successive contest rounds can create a paradox for many talented contestants. Consider a child who, after a moderate amount of practice, wins a local swim meet, or gymnastic or piano competition. It would be quite reasonable for the child (and the child's parents) to conclude that the win reflected talent, dedication, and a little luck. Bolstered by these conclusions, the child may be

induced to enter a regional contest, then a provincial, then a national, and beyond. With each success the competition will become more severe, and will likely cause the child to devote more time, attention, and money to pursue the winning goal. Such devotion is regularly sustained by a strong belief in the personal control of one's fate. Each success, however, reduces the validity of this belief. The more a contestant relies on talent and dedication to win, the less important these factors become in winning again. In very large contests, a contestant's talent and dedication can do little more than boost the chances for an opportunity to be lucky.

Examples of the importance of luck in determining the winners of large contests appear to be numerous. Olympic finalists commonly win or lose by fractions of a second, inch, or point. These small amounts are generally well within the range of day-to-day variability in the performances of each finalist. The finals of professional sports teams often "go to the wire." Division champions are sometimes so well matched that a single fumbled pass, bad call, or injury, one lucky interception, gust of wind, or foul shot can decide the final outcome. "Aesthetic" champions may well be decided by even more whimsical factors. In contests of artistic performance, the nationalities and philosophies of judges, the elegance of a single motion or phrase, one grimace, one smile, one high note that is flat by 2 cycles per second, can—and often do—determine the difference between losers and winners.

The importance of such random minutia in determining the outcomes of large contests might be of little concern if the lives of contestants were unaffected by them. But in most large contests the lives of many contestants are very much affected by them. It is therefore disturbing that as contests increase in size they are increasingly likely to eliminate rather than reward talent and dedication, and are increasingly likely to distinguish winners and losers by factors that have nothing to do with talent or dedication.

If talent and dedication, rather than luck, are to be rewarded more often, then it would seem judicious to discourage the growth of large contests and encourage the proliferation of smaller ones.

As our simulations have shown, the chances of the best person winning increases as the number of contestants decreases. If we extrapolate the findings of Simulations I-V, we can rest assured that when a contest is held to determine the best of a small lot of contestants who vary widely in their true talent, then, regardless of the contest procedure, it will quite likely accomplish this goal. Because luck plays a smaller role under these conditions, the contests would be arguably more fair. And as they proliferated, so too would the number of talented winners.

In many contests, however, contestants do not vary widely in their true talent. In a national final or in a contest with stringent eligibility criteria, contestants may be relatively evenly matched. Under these conditions, the results of Simulation VII show that the ratio of talent to luck decreases. To ensure that talent is rewarded more often in these "close" contests, we can extrapolate the results of Simulation VI and suggest that the winner be chosen on the basis of some averaged score for a series of performances.

Consider now the nature of contests from two other perspectives: the perspective of contest organizers, and the perspective of contest audiences. It is safe to say that contests are organized for a variety of reasons. One of these reasons may be the necessity of finding the most talented person in some given category. It seems equally likely, however, that other reasons dominate the motives of many contest organizers. For example, many contests are organized for the benefit of the organizer, not for the benefit of the contestants. In such cases, the organizer is likely to be less concerned with the validity or fairness of contest procedures than with the potential for these procedures to attract large numbers of enthusiastic, paying observers.

Thus it would not be surprising if the National Hockey League tradition of holding best-of-seven-game playoff rounds is based in part on considerations of fan tolerance for boredom. A series of, say, 21 games would be a truer test of talent. Yet such an extended playoff would surely put off those who would otherwise pay dearly to view a shorter series in which chance would play a far more important role.

Still other contests may be organized for the benefit of judges rather than contestants. The resume and vitae "contests" that are commonly held to select new employees, students, or other recruits often evolve procedures designed to minimize judgmental burden or conflict. In such contests, the goal of selecting the best contestant may be sacrificed for the goal of selecting an acceptable or satisfactory candidate with minimal judgmental effort. "Prescreening" of numerous applicants may well be based on (fallible) indicators of talent. But the result is often to produce a short list of talented candidates who cannot easily be further distinguished on the basis of these criteria. In response to this problem, judges will often resort to a frantic search for some other criterion that will easily differentiate the remaining contestants. It could be looks, or charm, or personality. More often it could be the extent of social contacts. These extra criteria may have little relation to talent, but they may alleviate the agony of indecision, and allow the judges—if not the losers—to get on with their lives. (For further analyses of this "elimination by aspects" choice heuristic, see Thorngate, 1980; Tversky, 1972.)

Judging from their frequent lack of interest in talented losers, and their frequently fickle reactions to aging winners, contest audiences often seem less concerned with the fate of contestants than with the excitement and rewards of the contests themselves. Fair contests are not necessarily exciting ones. A contest procedure that may painstakingly select the most talented contestant will often bore an audience to tears. It is probably far more exciting for an audience to witness a small group of very talented people give their best, suffer greatly, and leave their fate to chance than it is to see a single very talented contestant coast to a well-deserved victory. It is probably more exciting to place a bet on a less talented "long shot" and reap the large rewards of good luck than it is to bet on a more talented "sure thing" and collect the small rewards of a predictable outcome. Seeding contestants may not effect the final outcome. But is does increase the chances that a contest will end with a bang and not a whimper.

Contests can also become ritualistic, and ceremonious. The parades and souvenirs that so often accompany the finals of large

contests obviously do nothing to increase the chances that the best contestant will win. But they do seem to enhance the feeling of celebration among audience members, and add to the joy of witnessing victory and defeat. They may also boost the motivation and anxiety of contestants, causing some to excel and others to crack under pressure. A contest may thus select only for steady nerves, sangfroid, or psychopathy. But many audience members may be having such a good time, they may not really care.

In order to undertake our simulations we have had to assume that the true talent scores of contestants could be known in advance. Of course, these scores are normally not known in advance; if they were, there would be no reason to run contests. A case can be made that talent is what talent contests test and thus, by definition, losers don't have as much of it as do winners, at least at the time of the competition. Yet this attitude tends to reify contests, and excuse them from error. Those who await contest outcomes in hopes of discovering what or who is worthy or good at once relinquish responsibility for their own tastes and preferences.

In sum, we believe we have demonstrated that luck or chance plays a highly significant role in determining the outcomes of contests, particularly large contests. Chances are generally good that the best persons will not win, and thus that very talented individuals will suffer defeat for reasons entirely beyond their control. If a society can afford such errors, there is little reason to worry about this result. But if a society wishes to cultivate talent and reward it fairly, then large contests—though often exciting and profitable to their organizers and audiences—should not be encouraged.

REFERENCES

ARROW, K. (1951). Social Choice and Individual Values. New Haven, CT: Yale Univ. Press.

BARRY, B., and R. HARDIN (Eds.) (1982). Rational Man and Irrational Society? Newbury Park, CA: Sage.

BRAMS, S., and P. FISHBURN (1983). "America's unfair elections." Sciences, 23 (6): 28-34.

EINHORN, H. (1978). "Decision errors and fallible judgment: Implications for social policy," pp. 142-169 in K. R. Hammond (Ed.), Judgment and decision in public policy formation. Boulder, CO: Westview Press.

THORNGATE, W. (1980). "Efficient decision heuristics." Behavorial Sci., 23: 219-225.

TVERSKY, A. (1972). "Elimination by aspects: A theory of choice." Psych. Rev., 79: 281-299.