The Statistics of Dixonary Scoring

This was originally an extract from the
Coryphæus user manual
by Paul Keating,
seventh edition, May 2003.

Now distributed as an independent document,
fourth edition, December 2008.

Fagotstraat 32, 2287 BD Rijswijk, The Netherlands
Fax: +31 70 336 14 74
Email: coryphaeus@prodigycomputing.com
Website: groups.yahoo.com/group/coryphaeus

The game of Dixonary is played online at
http://groups.google.com/group/Dixonary

Document Revision 138
Change Control $Revision:: 1218 $

The Statistics of Dixonary Scoring

The statistical behaviour of the Dixonary scoring system is not immediately obvious. This analysis is offered to players who care to speculate on such topics as the modification of the scoring rules to equitably accommodate different numbers of players.

In what follows, I will assume that players play *randomly* and *rationally*. *Randomly* means that they do not apply their minds to eliminating obviously false definitions or choosing obviously likely ones, but instead choose on a random basis. *Rationally* means that they do not take actions that will diminish their scores, such as voting for their own definitions. Of course, neither of these assumptions holds universally in practice.

There are two components in any score: points gained from guessing correctly, and points earned by fooling other players into voting for a fictitious definition. I neglect the situation where a player votes but does not submit a definition, or submits a definition but does not vote, because neither of these is a *rational* strategy.

Points gained from guessing correctly

If there are only 3 players, of whom one is the dealer, you are assured of guessing correctly. There are 3 definitions, one of which is yours, so you cast your vote for the other two. One of them must be the correct one. So, with 3 players, you will gain 2 points from “guessing” correctly in every round. With more players, the chances of guessing correctly diminish rapidly, because, intuitively, the more definitions there are to choose from, the less chance there is of hitting on the right one.

To take a concrete example, if there are 10 players, one of whom is the dealer, there are 10 definitions. But each player will recognize his own definition and not vote for it. If he chooses randomly among the remaining 9, he has a $\frac{1}{9}$ chance of guessing correctly. Since he has two votes, but cannot vote for the same definition twice, he has a $\frac{1}{9}$ chance of guessing correctly out of the remaining 8 definitions with his second vote.

You might argue that as there are 8 definitions to choose from, the chances of guessing correctly with the second vote are $\frac{1}{8}$. This isn't so. A player can't be right with both votes. So, $\frac{1}{9}$ of the time he will have guessed correctly with the first vote, and in those cases his chances of guessing correctly with the second vote are zero. He has a $\frac{1}{8}$ chance of guessing correctly with the second vote, but only $\frac{8}{9}$ of the time, giving $\frac{8}{9} \times \frac{1}{8} = \frac{1}{9}$.

So in total he has a $\frac{1}{9} + \frac{8}{9} \times \frac{1}{8}$, or $\frac{2}{9}$ chance of earning two points. If there are 20 players, his chances are $\frac{2}{19}$, and in general, with n players, his chances are $\frac{2}{n-1}$.

In numerical terms, because a correct guess earns 2 points, this works out to $2 \times \frac{2}{9}$, or 0.4444 with 10 players, and $2 \times \frac{2}{19}$, or 0.2105 with 20 players, and in general, with n players, $\frac{4}{n-1}$. Call this quantity g for *guess*.

It is easy to see that as the number of players, n, gets larger, the average effect of guessing correctly, g, gets smaller, and when n is very large, it dwindles into insignificance. For example, with n=25, g=0.1667 and with n=30, g=0.1379. Put another way, the limit of g, as n tends to infinity, is zero.

Points gained from votes

Consider again the round with 10 players. There are $9 \times 2 = 18$ votes in circulation, and if players assign votes randomly, then each definition, correct or not, has an equal chance of collecting those votes. Obviously, no fake definition will get a vote from its author, but this affects all definitions equally, other than the real one, and it has no voting author to affect. So each definition can expect to earn $\frac{18}{10}$ points, which in numerical terms will give each fake definition 1.8 points. Call this υ, for *votes*.

If there are 20 players, then there are $19 \times 2 = 38$ votes in circulation, and if players assign votes randomly, then each definition, correct or not, has an equal chance of collecting those votes. So each definition can expect to earn $\frac{38}{20}$ points, and so for 20 players, υ=1.9.

In general, in a round with n players, there are $2(n-1)$ votes in circulation, and each definition can expect to earn $\upsilon = \frac{2(n-1)}{n}$ points.

As you can see, as the number of players, n, gets larger, the points earned from votes, υ, gets larger too, which compensates for the diminishing value of g. For example, with n=25, υ=1.92 and with n=30, υ=1.9333. But no matter how many players, it will never be more than 2, or, put another way, the upper limit of υ, as n tends to infinity, is 2.

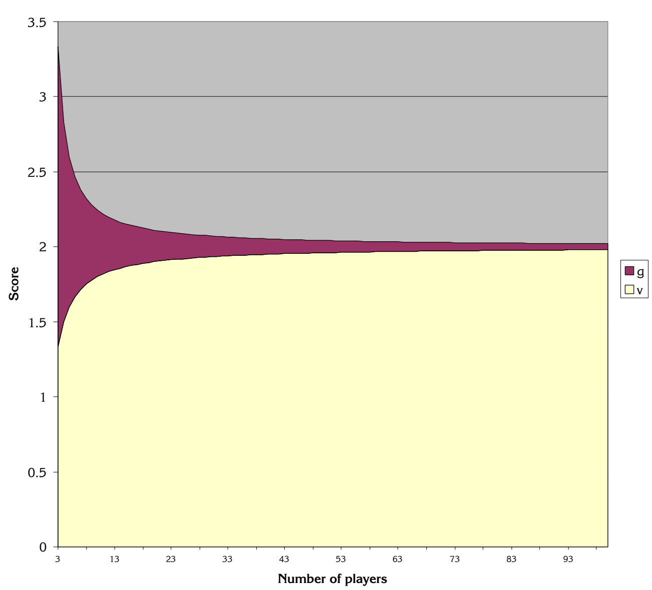
The expected total score

To recap, in general, there are n players in a round. There are n definitions, but each player will recognize his own and not vote for it. So each player, of whom there are $n-1$ not counting the dealer, has one chance in $n-1$ of guessing the correct definition with his first vote, and one chance in $n-1$ of guessing the correctly with the second: total $\frac{2}{n-1}$. If a player does this, he gets two points, and so the resulting average score g is $\frac{4}{n-1}$.

In a round with n players, there are $2(n-1)$ votes in circulation and if players assign votes randomly, then each definition, correct or not, has an equal chance of collecting those votes. So each fake definition can expect to earn $\upsilon = \frac{2(n-1)}{n}$ points.

There is a perception among players that as the number of players decreases, scores increase. This is partly true, but not nearly as true as may be supposed, for two reasons, as is as shown in Figure 1 below.

Figure 1: Average Score



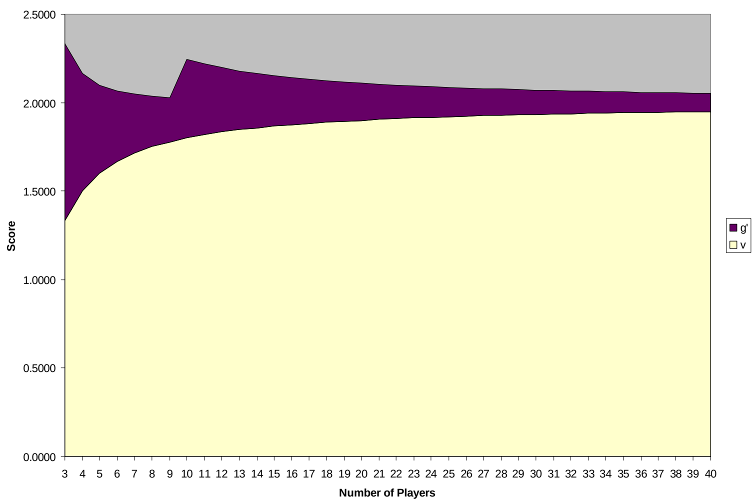
On average, a player can expect to earn $g+\upsilon$ points in any round. So the expected score E in a round with n players, counting the dealer, is given by $E = g + v = \frac{4}{n-1} + \frac{2(n-1)}{n}$. The lower limit of this value is 2, but for realistic values of n, say 15 to 30 players, it ranges from 2.15 down to 2.07.

The 1990 rules impose a moral obligation to vote on players that have submitted a definition, "for otherwise it would be a possible strategy only to submit definitions and never to vote, thereby never giving any other players any votes". The implication is that it is a winning strategy, but in fact, it is a losing strategy. By not voting, you forego the possibility of earning points g from a correct guess, thus reducing your expected score by $\frac{4}{n-1}$. It is true that you deprive other players of your votes, but that only lowers *their* expected scores by $\frac{2}{n-2}$: your two withheld votes, divided among all of the players other than yourself and the dealer. That results in a net reduction of your expected score in a round, relative to the other players, of $\frac{4}{n-1} - \frac{2}{n-2}$. That is why the opening paragraph dismissed this strategy as not *rational*.

Fixing the variability in the expected score

It is tempting to try and "smooth out" the value of $g+v$ so that the expected score is constant for any number of players n. Changing the multiplier for g for a given number of players (that is, awarding one point instead of two for guessing correctly if, say, $n<10$) introduces a discontinuity into the score at the changeover point, which seems so arbitrary that the cure is worse than the disease, as shown in Figure 2 below.

Figure 2: Modified Scoring



To eliminate the effect of the number of players on the score, the only real option is to award, not a constant 2 points, but a fractional value that reflects the reduced chance of guessing correctly as n increases. This value is $1-\frac{1}{n}$, which begins by awarding one point for guessing correctly, but then adjusts it downwards depending on the number of players, with the biggest downward adjustment when n is small. This captures the intuition of the step rule, while avoiding the discontinuity.

Using this approach, with 10 players we would award 0.9 points for guessing correctly, and with 20 players we would award 0.95. Since a player's chance of guessing correctly in these situations is $\frac{2}{9}$ and $\frac{2}{19}$ respectively, the modified value of g, let's call it g', would be $g' = 0.9 \times \frac{2}{9} = 0.2$ for 10 players, and $g' = 0.95 \times \frac{2}{19} = 0.1$ for 20 players. Recall that with 10 players, $\upsilon = 1.8$ and with 20 players, $\upsilon = 1.9$. You can see easily that with these values, the modified expected average score $E' = g' + \upsilon$ for a round, irrespective of the number of players, is 2, because for 10 players, $E' = 0.2 + 1.8 = 2.0$, and for 20 players, $E' = 0.1 + 1.9 = 2.0$.

You may think that awarding fractional points is impractical. So we could simply stop at this point. But let us pursue the issue to its logical conclusion. Even this approach carries within it the seeds of its own destruction, as we shall see.

Calculating the fractional points to award

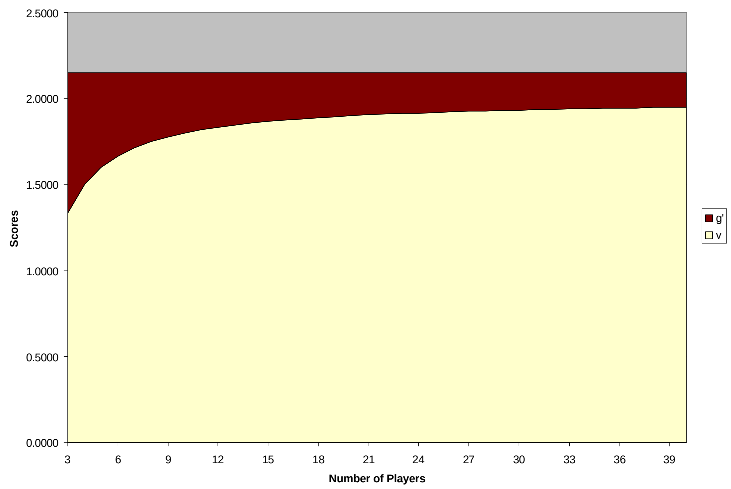
The first difficulty with awarding $1-\frac{1}{n}$ points for a correct guess is that it is too low a value. It results in a constant expected score of 2.0 for every round, and we would like to arrange matters so that the expected score would come out somewhere in the range 2.07–2.15, as it does at present. Let's set F (for finagling factor) equal to, say, 0.1, so that the expected scores will come out at 2.1 instead of 2. Then, for n players, the adjusted award for guessing correctly is given by $1-\frac{1}{n}+F$.

In practice, this means awarding 1 point for a correct guess when there are 10 players, 1.05 when there are 20 players, and 1.06 points when there are 25 players.

Then, irrespective of how many players submit definitions, the expected score will be roughly what it would be with 22 players under the current scoring system. The scores would look like Figure 3 on the next page.

The problem is that there's little justification in choosing $F = 0.1$. We have got rid of the variability in the expected scores in Figure 1, and we have got rid of the ugly discontinuity in Figure 2, but we still have the problem that there is no rational basis for deciding that $E' = 2.1$, just as there is no rational basis for changing the points for guessing correctly when there are 10 players. $F = 0.1$ is just as arbitrary as 10.

Figure 3: Variable Fractional Award Scoring



Fractional points based on history

We could look to the history of the game for guidance on a good value for *F*. The historical average number of players in a round, for the period (to round 1797) for which we have published statistics, is 25.06. Under the current scoring system, that gives $E = g + v = 2.08$. For 21 players, which is closer to the average number of players for recent rounds (since 1551), $E = g + v = 2.10$.

But consider this: the *actual* average score per nondealing player over that period, as opposed to the *expected* value *E*, was a disappointing 1.96. (For recent rounds it is better, at 2.04, but still pedestrian.) So it appears that players do substantially worse in practice than the assumption of *random* and *rational* behaviour would suggest. Put another way, they do not play randomly, but instead apply their minds to choosing the best definitions; but in general, any player whose cumulative average is less than 2.1 (which is about half of them) would score higher with random play.

So which value do we choose for *F*? The value 0.083 would bring the expected scores for new games E' into line with the expected long-term historical value of *E*. The value –0.04, on the other hand, would bring the expected scores for new games E' into line with the actual long-term historical record, and the value +0.04 would approximate the recent historical record (where *recent* is also entirely arbitrary).

This shows that even appealing to history does not help us with the problem of an arbitrary value for F.

Conclusion

We conclude by concurring with the commonsense argument that awarding fractional points is messy, and that the present scoring system, despite the defects illustrated in Figure 1, has simplicity to commend it, and that tinkering with the number of points awarded for a correct definition either makes matters worse, or is arbitrary, or both.

For a (not entirely serious) proposal for an alternative scoring system that avoids these and other defects, see the companion paper *A Lockstep Scoring System for Dixonary*.