A Lockstep Scoring System for Dixonary

A modest proposal

First edition, April 2010.

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 149
Change Control $Revision:: 1376 $

Why a lockstep scoring system is desirable

The scoring system of Dixonary is unsatisfactory in several respects.

1 Scores decline as more players play in a round. If there are only a few players it is "too easy".

2 Scores do not depend solely on the skill of the players, but also on the skill of the dealer: if a dealer chooses an easily-guessed word, players do well irrespective of the skill they display.

A companion paper to this, *The Statistics of Dixonary Scoring*, shows that these problems cannot be fixed simply by tinkering with the number of points awarded to a correct guess. Here we explain how to deal with these problems properly.

What we mean by *lockstep*

A handicapped game is one where a contest of unequal resources is turned, as nearly as possible, into a game with equally likely outcomes. This may be done for benign purposes, such as in a yacht race, the idea being to reward the sailors' skill, not the skill of their boatbuilders, or the depth of their sponsor's pockets. It may be done for social reasons, as in golf, to enable mismatched players to give each other an interesting game. And it may be done, as in horseracing, in an attempt to turn a game of skill (picking the fastest horse and best jockey) into a game of chance.

The lockstep system proposed here is, if you like, the opposite of a handicap system, because it tries to eliminate chance as far as possible from the scoring, and so offers a system where scores reflect players' abilities, and are not affected by uncontrollable factors such as the number of players in a round or the dealer's ability to choose a good word.

Basic principles

The basic principles of the system are as follows. Agree with them, and you will probably like the proposal. If you disagree with any of them, then you probably don't need to read much further.

1 *Skill* means the ability to write convincing definitions and to detect fakes. Scores should reflect skill, not just cumulatively, but accurately in every game outcome. A player's expected score, for

totally random play, should be constant in every round. It should be unaffected by the number of players or the dealer's skill.

2 Guessing correctly is not luck. It is affected both by the dealer's skill in choosing a word, and by the number of skillful fakes that compete for a voter's credence. Guessing correctly when it is easy should earn fewer points than when it is difficult.

3 You can't tell whether a word was easy or difficult to guess by counting the number of players. But you can do it by counting the number of correct guesses.

To agree with Principles 1 and 2 is to evince dissatisfaction with some aspect of the existing scoring system. Principle 3 is a criticism of the system used in *that other game**.

Keeping expected scores constant from round to round

There are several ways that Principle 1 could be maintained, but this system chooses a simple one: every player has two points to award, and so in a round with n players, there are $2n$ points in circulation. The average score must obviously be the total points divided by the number of players, $2n \div n$. So the average score is always 2, irrespective of the number of players, or the number of correct guesses.

You can score well only by taking points away from other players, and so the game score always reflects the skill that a player displays in a particular round, *vis-à-vis* his or her opponents' play in that round.

A player who scores higher than 2 in a round is "above average" and a player who scores lower is "below average". The average is *always* exactly 2.

How it works

Every player, *including the dealer,* has two points to award.

The non-dealer players award their points, as at present, by voting for definitions they like.

---

* *That other game* is a game similar to Dixonary with a slightly different scoring system that was active for many years on the MPGAMES forum on CompuServe and is still active on Yahoo! Groups. The most important differences are: only one point is awarded for a correct guess when the number of players is small; the dealer scores for a D0; and getting combined with the dictionary definition counts as a correct guess.

The dealer, on the other hand, must award his or her points by rule, which is that the two points are divided equally among the players who guess correctly. If no player guesses correctly then the dealer gets the points.

A player who votes for the correct definition gets one point (in other words, the vote is reflected back to the voter instead of to the author of the definition), plus a share of the dealer's two points. A sole correct guesser will get 3 points this way: one reflected vote and two from the dealer; but if there are 4 correct guessers then each one earns only $1\frac{1}{2}$: one reflected vote and a quarter-share of the dealer's two points. This lower score reflects the fact that the word was easier to guess (Principle 2).

Of course, more often than not, this means that the dealer will have to award fractional points.

There are also rules that keep the game points constant at $2n$ in special situations, such as when players vote for their own definitions, or the dealer combines definitions, which would otherwise push the total score for the round upwards or downwards. These rules also involve fractional points.

The justification for fractional points

Fractional points are inconvenient, but insisting on integral points for a correct guess is in conflict with Principle 2. To have this remark spelt out in detail, read *The Statistics of Dixonary Scoring*.

The chief objection to fractional points is that they are hard to compute by hand. But you can't actually play Dixonary without a computer, and dealers who don't want to use a dealing program can always score the round using a spreadsheet.

If you think that fractional points are messy and impractical, then you probably don't need to read any further. But as it is the foundation of the proposal, we will assume in what follows that you are at least prepared to accept it as a basis for discussion.

The correct-guess rule

When players vote correctly, they get their own vote reflected back, and in addition, get $\frac{2}{g}$ from the dealer, where g is the number of correct guesses. Dividing a fixed number of points among the players who guess correctly meets principles 2 and 3 by reducing the points earned by guessing correctly according to the number of correct guesses.

A side-effect of this is that when there are no correct guesses, that is, when $g = 0$, the number of points to award, $\frac{2}{g}$, is undefined.

We could simply say that in that case no points are awarded, as in the present system, but that would reduce the total score for the round below $2n$, the round constant. If the dealer awards no points, then in that round the round score is not $2n$ but $2n-2$.

So, the dealer's two points have to go somewhere, and we shall award them to the dealer. This is not just a mathematical convenience. It also addresses another problem. It is widely recognized that a D0 is generally the result of unusual skill on the part of the dealer, and simultaneous lapses on the part of the other players. But the only thing the dealer currently gets out of it is congratulations. Rewarding the dealer in this case accords with Principles 2 and 3. It also concedes that the system used in *that other game* is not entirely without merit.

If we accept that the dealer's two points have to go somewhere, but we don't accept that they should go to the dealer, the only other alternative would be to split the dealer's points among all of the other players, even though none of them guessed correctly: that is clearly absurd.

Special situations

Even if you are convinced of the merits of arranging the total scores so that the average per player comes out at 2 in every round, you may still find the working out of the special cases wildly overcomplicated. If the mathematical going gets too heavy, you can skip to the Computational Shortcut on page 10.

Players who submit a definition, but subsequently fail to vote

If in a round there are d players who submit a definition, *and subsequently do not vote,* the votes these delinquent players would have contributed to the game are lost. This would reduce the total score for the round. Instead of being the round constant, $2n$, it would be $2(n-d)$.

We cannot compel a player to submit a vote by the deadline. We could conceivably disqualify the definition. But we could not reasonably do that if anybody voted for it, because if we did, those votes would presumably also be disqualified. If we did that, yet *more* votes would "go missing", making the problem worse not better.

To maintain the round score at the round constant of $2n$ when a submitter of a definition does not vote, one option would be for the dealer to cast the missing votes. He could do this by dividing, not the usual 2 votes among the players who guessed correctly, but $2d+2$ votes. But boosting the votes for guessing correctly in this way would clearly be in conflict with Principle 2, because it would increase the value of guessing correctly, not because the word was harder to guess, but simply because some player neglected to vote.

Alternatively, dealers could take the "missing" votes for themselves, or award them to the delinquent player. But both of these options are clearly absurd and in conflict with Principle 2.

The alternative is to increase the value of a vote. So, instead of a vote counting 1, it will count $1+\frac{2d}{n}$. In this way, the votes not cast by the delinquent player are in effect cast by the remaining players, by increasing the value of their votes to make up for the two he or she failed to cast; and the fake definition that gets the most votes benefits the most. All of the players stand to benefit equally. Even the delinquent players benefit, because their definitions will collect a share of the increased value of every vote they catch.

Disqualified from voting

Players who DQ are bound by the rules not to vote. From a scoring point of view, this is identical to the previous case, and we increase the value of a vote to $1+\frac{2d}{n}$, where d is the number of DQs.

Players who vote for their own definitions

Voting for one's own definition has long been regarded as a legitimate tactic (though it is arguably against the 1990 rules). Players who do this diminish their score, because in doing this they forgo one of two chances to guess correctly. From a scoring point of view, it amounts to the same thing as failing to vote, except that it is a partial failure rather than a total failure. So again, this is identical to the previous case, and we increase the value of a vote to $1+\frac{d}{n}$, where d is the number of players who voted for their own definition.

Though discouraged by the 1990 rules, it is legal and does sometimes happen that a player, especially a new player, casts only one vote. Since this is also a partial failure to vote, we add such a player to the value of d, as if he had voted for his own definition, if he submitted one; and subtract him from d if he did not submit a definition.

Points and values

Up to now we have carefully chosen to talk about "increasing the value of a vote" without specifying the units in which this value is denominated. Traditionally, a *point* awarded by the dealer has the same value in the game as a *vote*. If we increase the value of a vote, we have a terminological problem if we say that a vote is worth, say, 1.026 *points*, because when the dealer awards a *point*, the value of that *point* will also be increased—to 1.026 points. It at the very least confusing to reckon the value of a *point* to be 1.026 points, so in what follows we will talk about *dealer votes*. Players have two *votes* to cast; dealers divide two *dealer votes* among correct guessers, and they may sometimes award compensatory *dealer votes*. Usually the value of a vote (player vote or dealer vote) is one point, but sometimes (when the number of votes in the game deviates from $2n$) the value of a vote is adjusted upwards or downwards, to keep the total number of points awarded in the round equal to $2n$.

Combined definitions

It is the practice for dealers to combine definitions only with reluctance, and this reluctance is greatest when the one of the definitions to be combined is the real one. But it still happens, and when it does, it introduces perturbations into the score.

Submitted definitions combined

When two definitions are combined into one, each of the authors gets a vote when the fake definition fools another player. That means that this player's vote is double-counted.

Double-counting of votes increases the game total above the round constant $2n$. To avoid this, we *could* award $\frac{1}{2}$ to the authors of combined definitions: in other words, they have to share the votes that the combined definition catches.

But that is in conflict with Principle 1. It is not a reflection on an author's skill that someone else hit on the same idea for a definition: if anyone is "to blame", it is the dealer, for choosing an obvious word.

So, when two definitions are combined, to eliminate the effect of double-counting, we adjust the value of a vote downwards, so that instead of being worth 1, it is worth $1-\frac{\upsilon}{n}$ where υ is the number of votes for the pair of combined definitions. Where three definitions are combined, the downward adjustment will need to be bigger because votes are triple-counted: $1-\frac{2\upsilon}{n}$.

On rare occasions, there may be more than one combined definition in a round. In the general case, we shall say there are m fake definitions D_1, D_2, ... D_m, and these were compiled from submissions numbering s_1, s_2, ... s_m (in other words, if definition D_1 was combined from 3 submissions then $s_1=3$; and if definition D_2 was the work of one author then $s_2=1$). During the voting, these definitions receive votes totalling υ_1, υ_2, ... υ_m. The value of a vote will be $1-\frac{\sum_{i=1}^{m}\upsilon_i(s_i-1)}{n}$. In usual situations, most definitions will not be combinations. For those definitions, s_i will be 1, which means that $s_i-1=0$ and so those definitions will not contribute to the summation. In the usual case there will be only one combined definition, and this expression will reduce to the simpler $1-\frac{\upsilon}{n}$ or $1-\frac{2\upsilon}{n}$ mentioned above.

Submitted definition combined with the real one

The 1990 rules award no points to the player whose definition is combined with the real one, other than the votes the combined definition catches from other players. Some players think this is unreasonable, but that is not the concern of this paper.

In lockstep scoring, a correct guess normally gets the vote reflected back, plus a share of the dealer's 2 dealer votes. But if a player can't score for voting for the correct definition, his or her vote "goes missing". On top of that, his or her share of the 2 dealer votes goes missing.

Let us say that Sally's definition was combined with the correct one, and she voted for the combination, and Theo, Ursula and Vernon also guessed correctly.

We treat Sally as having voted for her own definition: no points. Sally also doesn't get points for guessing correctly, as in the current system. But we can't simply share the two dealer votes among Theo, Ursula and Vernon. That would be in conflict with Principle 2. The word didn't become harder to guess, and so worth more dealer votes, just because

Sally was deprived of points by the operation of an unrelated rule. So in this case we count Sally as one of the correct guessers g. The dealer awards a $\frac{1}{2}$ point to Theo, Ursula and Vernon. The other $\frac{1}{2}$ point can't go to Sally, so it "goes missing".

In the normal case, a vote for the correct definition reflects the vote back at the voter. But when a definition is combined with the real one, correct guessers have their votes double-counted: so Theo, Ursula and Vernon get one vote reflected back, but Sally also gets a vote from each of them. But Sally's vote for her own definition is not reflected back; nor does it go to another player; and so that vote is undercounted.

If g is the number contributors to the combined correct definition that subsequently voted for it, then we need to correct the adjustment for combined definitions by reducing v_i for that definition thus: $1 - \frac{\sum_{i=1}^{m} v_i(s_i - 1)}{n} + \frac{g}{n}$.

Players who do not submit a definition, but subsequently vote

If in a round there are d players who vote *without having submitted a definition,* this means that there are $2n$ points available, but only $n - d$ definitions. These d delinquent players can only earn points from guessing correctly, and the points they would have earned as their share of the players' votes are divided among the remaining $n - d - 1$ players. But the total score for the round remains $2n$, the round constant, and so this situation needs no further examination. It may be irrational for players to choose to play in a way that increases others' score at their own expense, but the irrationality is theirs and not the scoring system's.

Compensatory dealer votes (dealer points)

It is entrenched in the game to award compensatory dealer votes (almost invariably 2) to players who have been disadvantaged in some way, despite it being nowhere explicitly authorized in the rules. Mostly, dealer compensatory dealer votes are awarded for a lost or mangled definition, and are restitution for not being able to collect votes.

Awarding compensatory votes adds to the number of votes in circulation, because the player who receives them can still vote. To deal with this situation, we need to adjust the value of every vote downwards. Instead of being worth 1, a vote is worth $1 - \frac{d}{n}$, where d is the number of dealer votes awarded in the round. This downward adjustment applies to the dealer votes too, of course, so that a player awarded 2 dealer votes will actually only score $d = 2(1 - \frac{2}{n})$ for them. In a round with 20 players, this amounts to $2(1 - \frac{2}{20}) = 1.8$ points. Of course, players who get dealer votes may guess correctly, and so score more than 1.8.

The present system of awarding compensatory dealer votes for a lost definition has the effect of a penalty, if the player who receives them has a cumulative average of more than about 2.2. It has the effect of a bonus, if the player's cumulative average is less. Since it is an average, it need hardly be pointed out that half of the players lie above this point, and the other half below it.

In lockstep scoring, the, devalued dealer votes penalize, not just the player whose defintion was lost, but everybody else as well. The player is whose definition was lost is deprived of the expected score from other players' votes v, and receives d in compensation. But $d - v$ is always negative: it ranges between –0.133 and –0.087*. And the other players receive votes that are devalued by $\frac{d}{n}$.

One could preserve the status quo and keep the compensation neutral with respect to the number of players by awarding, not two, but $\frac{2(n-1)}{n(1-\frac{2}{n})}$ devalued dealer votes. This is a number between about 2.10 and 2.15*, and will come out at 2 points irrespective of the number of players.

But this is messy, and in fact the status quo is not really worth preserving; because there is no getting away from the fact that *compensatory dealer votes are a bad idea*.**

We keep compensatory dealer votes in the scoring system because they are entrenched in the game, but player whose definition goes missing would be advised to opt instead for being treated as if he or she had simply sat out the round.

Summary

All of the point-value adjustments for special cases are cumulative, because it is always possible that several will arise at once. In summary, the special cases are taken care of as follows. Say that

a players submit a definition but fail to vote, or DQ
b players vote for their own definition,

* For realistic numbers of players, that is, between 15 and 24.
** The paper *Dealer Points Considered Harmful* treats this issue at length.

the dealer awards a total of c compensatory dealer votes,
d players submit a definition but cast only a single vote,
e players do not submit a definition and cast only a single vote,
g players guess correctly and also submit definitions that are combined with the real one.
In the round there are m fake definitions D_1, D_2, ... D_m, and these resulted from original submissions numbering s_1, s_2, ... s_m; and these definitions receive votes totalling υ_1, υ_2, ... υ_m.

Then the value of a vote is
$1+\frac{1}{n}\left(2a+b-c+d-e-\sum_{i=1}^{m}\upsilon_i(s_i-1)+g\right)$ points.

Computational Shortcut

The formula is pretty daunting and it only covers the special cases that it knows about. What we need is a computational shortcut that is easy to do and is robust enough to handle exceptional situations or combinations of circumstances that I may have neglected.

Fortunately, this is not too hard. The fundamental goal is to keep the total score for a round at $2n$ where n is the number of players. The shortcut is as follows:

1 Count votes as 1 each, irrespective of whether they are double-counted, etc., precisely as is done in the standard system.

2 Treat players who vote for their own definition, DQ, or fail to vote, precisely as is done in the standard system.

3 *New.* Do not award 2 points to each player who guessed correctly. The dealer has only 2 dealer votes to award for correct guesses. Divide these 2 votes equally among the players who guessed correctly*.

4 *New.* If nobody guessed correctly then award the votes to the dealer.

5 Compute the sum of all votes awarded and call it T. Now count the number of players, including the dealer, and call it n. The value of a vote is $\upsilon = \frac{2n}{T}$.

* But a player whose definition was combined with the dictionary definition gets no share of the dealer's votes for guessing correctly; the other players don't get that share either.

6. To compute the lockstep score of a player, take that player's tally of votes and multiply by υ.

It is easy to see that υ will scale the value of votes so that the round total is always $2n$. In the absence of special cases, $T = 2n$, because there are n players, each of whom has 2 votes to cast (including the dealer), and so $\upsilon = 1$.

Should players fail to vote for whatever reason, then T will drop below the expected value of $2n$ and υ will be greater than 1. This will raise the value of the remaining votes to compensate for the fact that some have "gone missing".

If extra votes are introduced into the round, either through double-counting of votes for combined definitions, or because of compensatory dealer votes for lost definitions, then T will be greater than the expected value of $2n$ and υ will be less than 1, thus devaluing all of the votes cast to compensate for the fact that there are too many of them.

Practical implementation

Here are a few sample rounds to show how lockstep scoring would work in practice.

Round 1732. One missing vote: the remainder get scaled up.

Player	Def	Votes	DPs	Standard Score	Lockstep Votes	Lockstep Points
Barrs			2	2	2	2.051
Bourne	12	17 & 18		1+0=1	1+0=1	1.026
Carson	9	*13* & 15		0+2=2	0+1.5=1.5	1.538
Crom	16	4 & 14		1+0=1	1+0=1	1.026
Cunningham	3	*13* & 14		2+2=4	2+1.5=3.5	3.590
Emery	6	1 & 5				
Heimerson	2	*13* & 18		2+2=4	2+1.5=3.5	3.590
Hirst	7	10 & 15		1+0=1	1+0=1	1.026
Keating	4	4 & 18		5+0=5	5+0=5	5.128
Lodge	18	2 & *13*		6+2=8	6+1.5=7.5	7.692
Madnick	10	14 & 16		1+0=1	1+0=1	1.026
Savage	17	2 & 7		2+0=2	2+0=2	2.051
Schultz	11	4 & 18				
Scott		1 & 3				
Shefler	5	17 & 18		2+0=2	2+0=2	2.051
Shepherdson	15	1 & 4		2+0=2	2+0=2	2.051
Stevens	8	4 & 18				

Player	Def	Votes	DPs	Standard Score	Lockstep Votes	Lockstep Points
Stone	1	3 & 4		3+0=3	3+0=3	3.077
Widdis	14	5 & 12		3+0=3	3+0=3	3.077
Dealer						
Total				31+8+2=41		40

In this round, there were 20 players including the dealer. But one player's emails went missing entirely (no definition and no votes) and he was awarded 2 dealer votes. This introduced 2 extra votes into the round, but that only served to replace the votes he did not cast (so no adjustment required). Another player voted without submitting a definition (no adjustment required), and one player voted for his own definition (so one vote went missing: to account for this, the value of a vote needs to be scaled up slightly, to 1.026).

The word was not too difficult to guess: four players voted for it. So in lockstep scoring those players get $1\frac{1}{2}$ dealer votes for a correct guess instead of 2.

With this round scored in lockstep, players did less well with correct guesses (because the word was easy to guess), but better if they snagged votes (because one vote "went missing", making votes slightly harder to get, and so more valuable). But in total, lockstep scoring awarded 40 points, one less than the 41 of standard scoring.

Round 1650. D0: dealer gets 2 points, everything else as standard

Player	Def	Votes	Standard Score	Lockstep Votes
Abell	12	11 & 15	1	1
Bourne	5	12 & 15		
Carson	1	2 & 19		
Crom	18	8 & 10	1	1
Cunningham	2	4 & 15	2	2
Emery	13	8 & 16		
Heimerson	8	15 & 17	6	6
Hirst	9	8 & 19	2	2
Kiwiro		17 & 18		
Kornelis	17	2 & 9	6	6
Lodge	19	3 & 11	3	3
Madnick	16	3 & 17	4	4

Player	Def	Votes	Standard Score	Lockstep Votes
Mayer		7 & 16		
Savage	15	7 & 8	7	7
Schultz	4	15 & 17	1	1
Scott	11	15 & 17	2	2
Sheffler	6	8 & 19		
Sheperdson	10	8 & 16	1	1
Wetzstein	3	16 & 17	2	2
Widdis	7	9 & 15	2	2
Dealer				2
Total			40	42

This was a very straightforward and uncomplicated round. It's included only to demonstrate what happens when the dealer fools everyone. The only difference is that the dealer scores 2 for the D0, which keeps the total score for the round at $2n = 42$. There is no need to scale the point value because everybody who submitted a definition voted, and nobody voted for their own def, so no votes "went missing". Lockstep scoring awards two points more than standard scoring, to the dealer.

This page intentionally left blank.