IDENTIFICATION AND MEASUREMENT OF LUCK IN SPORT

Darren O'Shaughnessy [a,b,c]

[a] *Ranking Software, Melbourne*
[b] *Hawthorn Football Club, Melbourne*
[c] *Corresponding author: darren@rankingsoftware.com*

Abstract

In Australian Rules Football, play is started each quarter, and after each goal, by the umpire bouncing the ball in the centre of the ground to be contested by two ruckmen while several shorter players rove the vicinity. There are about 28 of these events each match, and it is common wisdom that winning clear possession here is a Key Performance Indicator. When a team wins the ball from this restart, it immediately enters its attacking zone 70% of the time, and scores directly from 21% of its clearances without an opponent touching the ball.
Whilst evaluating Hawthorn Football Club's successful 2014 season, I discovered three remarkable disruptions to this common sense interpretation:

1. Despite full-time stoppage coaches and massive investment in strategies, the win-loss ratio of clubs in an entire season's data was essentially a coin flip. Only one club won centre clearances at a rate outside ±1.5 standard errors in the mean compared with a Bernoulli trial model, and the standard deviation of the 18 clubs' SEMs was 1.06.
2. The centre clearance statistic had zero reliability; in fact teams were slightly more likely to get a negative differential in the match following a positive one.
3. Over the season, there was a strong negative correlation (r = –0.68) between clubs' clearance win rate and the conversion of those opportunities into scoreboard equity (O'Shaughnessy, 2006). This suggests teams that invest extra resources in winning the "coin-flip" have depleted resources at more vital locations.

Hawthorn FC embarked on a *gedankenexperiment* where the coaches were asked to consider how they would structure their resources if, instead of competing to win the clearance, the umpire simply flipped a coin and gave the ball to the winner.
This paper explores the implications of classifying some events as mostly luck – a spontaneous breaking of symmetry – and how a sports team or player might measure performance accounting for known sources of variation.

Keywords: Luck, performance analysis, Australian Rules Football, AFL

1. INTRODUCTION

As the field of performance analysis matures, identification of key performance indicators in sporting contests has expanded. While some new metrics – such as Expected Goals in soccer – are explicitly attempting to reduce random variation, coaches and analysts have often been reluctant to re-evaluate the meaning of traditional KPIs that have always been part of their reporting and coaching methods.

In reality, all actions that are counted on the field of play are affected to some extent by stochastic processes, from the bounce of the ball to the workings of the neuromuscular system (Mauboussin, 2012) to the intervention of officials. The state of play prior to each action provides significant context to the action's outcome, but is itself affected by both systematic and random factors.

Additionally, statistical models that attempt to predict team success based on combinations of these KPIs are affected by several confounding factors including

- Small sample sizes of perhaps a couple of dozen games per year
- Opposition effects: each contest is a dynamic exchange (e.g., Gréhaigne & Godbout, 2014) where the opponent has as much effect on the collection of outcomes as the team. This is unlike most other fields of statistical analysis
- Multi-collinearity and interconnectedness between indicators

We would to identify those actions which are indicative of skill, and those we can classify as luck. Naturally this is a continuum and the proportion of luck in a mixed indicator tends to decrease with the square-root of the number of data points. We would also like to measure the effect of these on the scoreboard. This paper offers some insights into the process in one sport.

2. METHODS

THE CENTRE CLEARANCE

One distinctive aspect of Australian Rules Football (colloquially called AFL) is the amount of umpire intervention. There are three field umpires, two boundary umpires, two goal umpires, video reviewers, supervisors and assistants. This befits the somewhat anarchic play with 36 players on the field, no offside rule, full contact, and dozens of player substitutions each quarter. At the beginning of each quarter, and after each goal, an umpire takes possession of the ball in the centre of the oval and either bounces it or throws it into the air[1]. This is called the Centre Bounce, one of three types of *stoppage* where the umpire impels the ball back into play.

At the Centre Bounce, only one designated *ruckman* from each team is able to contest the ball in the air. The ruckman who hits the ball (gaining a *hitout* statistic) attempts to direct it to one of his *rovers*. Three rovers from each team are allowed within the area, and other players arrive from 25m away within five seconds of the bounce. The rover with *first possession* either attempts to clear the ball from the congested area himself with a handball or kick, or passes to a teammate who seeks to achieve this. The first player who effectively passes the ball to a teammate in sufficient space, or successfully clears the centre area without disadvantaging his team, is awarded a *centre clearance* or *centre break*. If neither team can release the ball due to the carrier being tackled, the umpire directs a *secondary stoppage* or *ball-up*.

Figure 1 is a Sankey diagram showing the average flow of ball possession from the Centre Bounce. Where numbers do not sum to 100%, there is leakage due to the play ending without clearance. There are numerous opportunities for individual skill and team strategy to influence the play, from the jump and tap of the ruckman to subtly blocking opponents, commanding space, sharking the ball from the opposition ruck, stealing it at ground level or gaining a free kick. While the best ruckmen can achieve hitout rates of over 65% long-term, *hitouts to advantage* where a teammate has immediate space to make a decision are rarer – about 28% of all CB hitouts are classified this way (not shown in Figure 1, but comprising 56% of the 50% Gathered).

Teams invest considerable resources into stoppages, with professional teams often employing a full-time stoppage coach, engaging ruck consultants and developing playbooks of formations and strategies intended to give the team a significant advantage from these neutral restarts.

For the Centre Clearance analysis in the next section, this paper will consider only those 86% of Centre Bounces that assigned a clearance to a team. Notations from every 2014 match were used (n = 4915 from 207 matches). 70% of *possession chains* from clearances are taken into the attacking zone within 50m of goal without an opponent touching the ball, and of those 30% lead directly to a score.

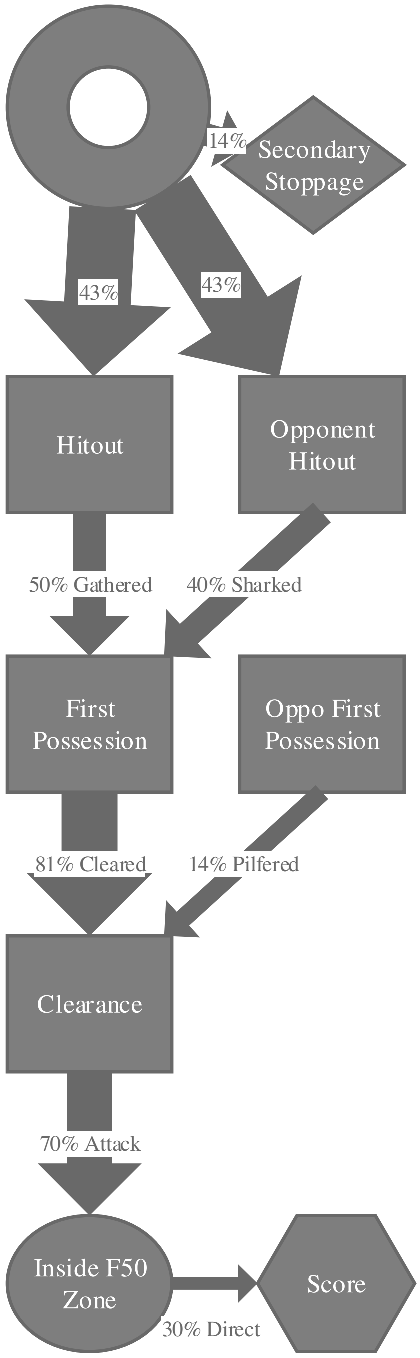

Figure 1: Anatomy of a Centre Clearance

---

[1] The ball is thrown up when conditions are unfavourable for bouncing, or if the umpire has failed to bounce the ball satisfactorily; these differences are not discussed in this paper

SCOREBOARD EQUITY

O'Shaughnessy (2006) uses the term scoreboard equity to describe the net value of the current state of the ball, with respect to a team's chances of scoring next. Given sufficient time, the Markov process that describes ball movement will be absorbed as a score, worth either +6, +1, -1, or -6 to the team. Scoreboard equity is the expected mean value of the ensemble of absorbing states, weighted by the probability of that score occurring next. Each AFL play can be regarded as a mini-game where teams attempt to optimise equity, in order to maximise their probability of winning the match through realised scores.

Simple data analysis shows that the average equity of a centre clearance in recent years has been +1.06; if the opponent clears the ball the team has equity -1.06. This implies a swing of 2.12 scoreboard points depending on which team wins use of the ball from the centre bounce.

For this paper, the scoreboard equity has been evaluated at known standard states, calculated via parametrisation of empirical data over the seasons 2012-2014. For the purposes of the model, we can regard these states as known intermediate equity values at the location $\boldsymbol{x}$: ball-up by umpire, throw-in by boundary umpire, set shot by either team, running shot by either team. Unlike the 2006 paper where play was tracked until the next score, the model used here truncates the equity calculation to the next stoppage or shot at goal. This should reduce the noise and team-specific effects due to tracking further play after the standard state.

BACKGAMMON ANALYTICS

The game of backgammon, from which the equity theory was adapted, offers reliable ways of analysing the mix of skill and luck employed by both players. The first and most direct method is to compare the player's choice of move with a perfect evaluation, and report the error rate. For instance, if a player has to make 250 decisions during a match, and makes 22 errors with a total loss of equity of 1.4 points versus perfect play, his error rate is reported as -5.6 points per 1000 moves. If his opponent's error rate is -7.3, he can be regarded as playing better whether he won or lost. This measure is not perfect, as it does not take into account the relative difficulty of the decisions, only the extent to which match-winning probability was diminished by each choice.

The second method employed by backgammon analysts is to measure the luck dealt by the dice, then subtract it from the result of the match to leave a residual skill effect. The algorithm evaluates the equity of each of the 21 possible dice rolls, and collects the difference between the mean of those 21 possibilities and the equity of the actual dice roll. This luck contribution can also be scaled using the effect on match-winning probability. In simplified terms, if a player won 11-8 with total luck of +2.6 points, while his opponent's dice produced total luck of -1.4, then the residual displayed skill suggests that he should have lost by one point instead of winning by three.

With a perfect evaluator, these two methods of skill measurement should report the same conclusion. Backgammon software is now superior to human experts, but is imperfect given the size of the decision tree. Nonetheless the two measures are very similar in most real-world situations, validating the use of luck measurement to expose differences in skill.

It is well known among backgammon experts that in a typical match lasting an hour, luck dominates skill. Between opponents of similar expertise, the match is far more likely to be decided in favour of the luckier player than the more skilful one. Since the introduction of sophisticated backgammon software in the past two decades, this fact has become part of the common knowledge and the culture. In contrast in professional sport, there is very little evidence that coaches or commentators have understood this basic law of statistics: if the competitors are close in skill, luck will usually determine the winner of the contest. This applies at the level of player versus player competing for a loose ball, and at the level of team versus team over a match.

In sport there is no such thing as a perfect evaluator, even in principle as players do not take turns using their own random activity generators. However, in the following sections some obvious sources of on-field luck will be partitioned and measured on an equity scale, reducing the error in the measurement of skill difference, compared to the scoreboard.

3. RESULTS

CENTRE CLEARANCES

Winning the majority of centre clearances is clearly a factor in winning a match of AFL. As mentioned in the previous section, the effect of the average centre break is an immediate boost of +1.06 points of equity, with the winner scoring next 65% of the time.

The result of a quarter-by-quarter linear regression of score margin vs centre clearance differential yields a slope of 1.00±0.17, while a match-by-match regression of the same data shows a slope of 1.28±0.48. This indicates that the advantage from each clearance persists through the match, but does not show any extra effect that might be due to collinearity between overall team skill and an ability to win centre clearances.

The most startling result came in the analysis of variation in team clearance percentages through the entire 2014 season, as shown in Table 1 and Figure 2. Comparing with a Bernoulli trial model where each team has 50% probability of winning the clearance, 17 of 18 clubs were within ±1.5 standard errors in the mean. The standard deviation of that collection of SEMs is just 1.06, suggesting that the entire season's variation from 50/50 very closely resembles noise.

To test the reliability of the Centre Clearance indicator week to week, each team's centre clearance differential was regressed against its differential in its next match. The correlation coefficient was approximately 0.015, even after adjusting for any home ground advantage effect (home teams averaged 12.4 clearances to 12.0 for the away team). In other words, teams did not seem to be able to maintain any advantage over time, nor did they correct their deficits any more than regression to the mean would suggest.

Club	Centre Clearances				Equity from Centre Clr			
	Won	Lost	%	SE in Mean	Won	Lost	Diff	Final Position
Essendon	292	228	56.2%	+2.81	1.05	-0.71	1.77	7
Adelaide	311	280	52.6%	+1.28	1.17	-0.76	1.94	10
Sydney	274	252	52.1%	+0.96	1.23	-0.81	2.04	2
Fremantle	274	253	52.0%	+0.91	1.02	-1.25	2.28	5
WC Eagles	278	260	51.7%	+0.78	1.02	-0.96	1.98	9
W Bulldogs	280	263	51.6%	+0.73	0.93	-1.09	2.02	14
Carlton	279	273	50.5%	+0.26	0.93	-1.06	1.99	13
Port Adel	288	286	50.2%	+0.08	1.42	-0.84	2.26	3
North Melb	278	280	49.8%	-0.08	1.12	-1.04	2.16	4
St Kilda	269	277	49.3%	-0.34	0.73	-1.49	2.22	18
Melbourne	219	226	49.2%	-0.33	0.81	-1.31	2.12	17
Collingwood	247	257	49.0%	-0.45	0.99	-1.01	2.00	11
Richmond	264	279	48.6%	-0.64	1.02	-1.04	2.05	8
Gold Coast	271	291	48.2%	-0.84	1.19	-1.04	2.23	12
GWS	282	305	48.0%	-0.95	0.82	-1.22	2.04	16
Geelong	259	288	47.3%	-1.24	1.37	-0.97	2.34	6
Hawthorn	300	334	47.3%	-1.35	1.39	-1.11	2.51	1
Brisbane	250	283	46.9%	-1.43	0.81	-1.39	2.20	15

Table 1: Centre Clearance Results by Team

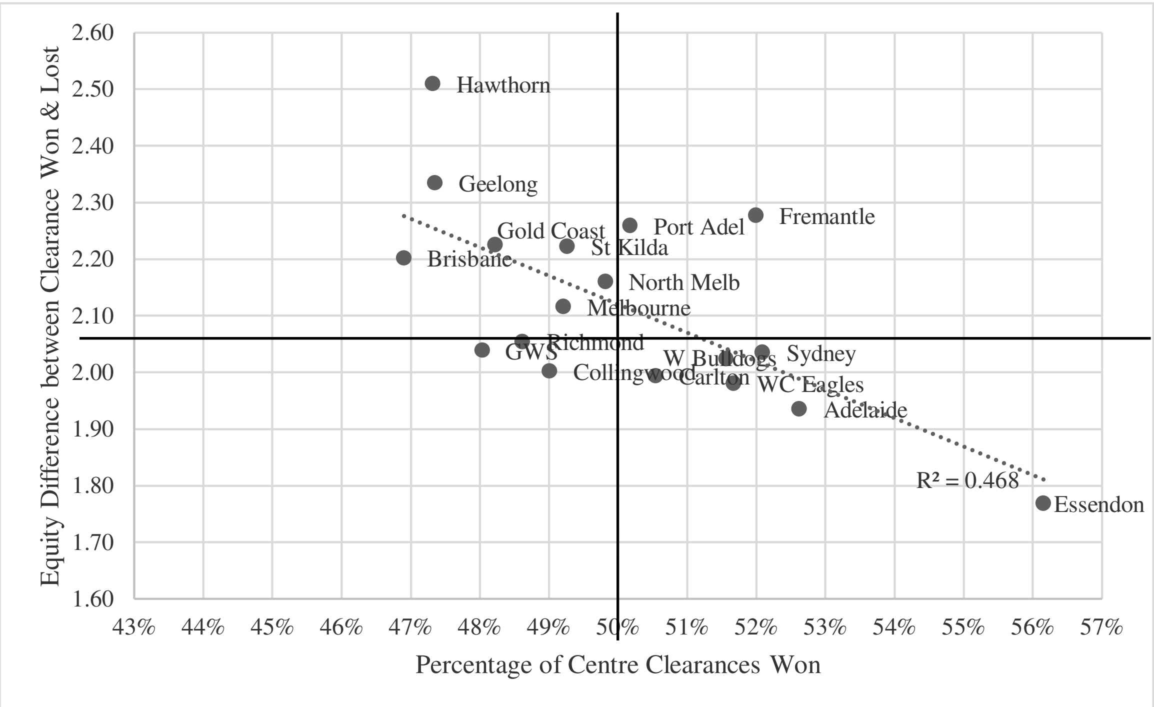

Figure 2: Equity Difference through Centre Clearance versus Centre Clearance Win%

There appears to be a strong negative correlation between winning a high percentage of Centre Clearances (Essendon being the outlier) and a reduced effectiveness of those opportunities as measured by equity. Additionally, there was no relationship between club quality as measured by their final position and any of the Centre Clearance indicators, apart from the obvious that equity metrics are higher for better teams.

MEASURING LUCK

A visualisation of the scoreboard margin as a sum of "mostly luck" and "mostly skill" effects has been developed. As an example, Table 2 shows the 2014 Preliminary Final which was won narrowly by Hawthorn, 15.7 (97) to 13.16 (94).

In this match, virtually all of the measurable luck-heavy indicators were in Hawthorn's favour, including conversion of scoring opportunities. The results of shots at goal have a heavy impact on the ultimate result, but are strongly dependent on random effects of the ball being dropped onto the swinging leg.

A traditional media analysis of this match would have emphasised the dominance at clearances and "clutch" shooting at goal as strengths of a team that went on to win the premiership the next week. Instead if we choose to monitor more stable indicators of on-field performance, our primary conclusion would be that Port Adelaide was very unlucky to lose.

Mostly Luck	Raw	Equity
Centre Clearances	16-12	+4
Ball-up Clearances	11-8	+3
Throw-in Clearances	21-14	+7
Shots at Goal	15 / 23	+17
Oppo Shots at Goal	13 / 31	+5
Unrealised Equity		+2
Mostly Skill		
After Won Clearance		-32
After Lost Clearance		-8
After Kick-In	7 / 15	+1
After Oppo Kick-In	6 / 7	+4

Table 2: Partition of Play by Phase

For a typical AFL match, using a continuous Markov process or approximating its absorbing states by a Poisson distribution, we can estimate the total variance of the final margin as approximately 1060 ($\sigma \cong 32.5$), varying slightly with the pace of the game. Approximately 34% of this variance is explained by the success or failure of shots on goal, and another 9% is explained by the clearance count. Considering the thousands of micro contests that happen around the ground and affect who emerges with the ball, this is a sizeable first step towards reducing the noise in the analysis of core performance.

4. DISCUSSION

The graph in Figure 2 caused surprise at Hawthorn FC, and requires substantial thought by experts in the sport as well as statisticians. The coaching group discussed what it meant if the centre clearance was effectively a coin flip, and how they might structure their defence, midfield and attack to respond to an event they have very little control over. The strong hint in the data that teams might over-invest in winning the clearance at the expense of resources in more impactful locations and roles is also a lesson to consider. Additionally, if the coaches cannot effectively intervene during the match to bring about a dominance of centre clearances – despite their obvious importance to the result – then they are free to stop worrying about that KPI during the game and focus their skills on pattern recognition and problem solving that computer algorithms cannot tackle.

Of course there is luck in every action on the field, and plenty of it in those phases labelled as "mostly skill". For any given contest, it is likely that luck dominates the difference in skill between two elite players. In any five-minute period, there is probably not enough data recorded to reliably inform decisions in the coaching box, yet almost every club speaks in terms of periods of dominance similar to this time-scale. Our obligation as statisticians is to patiently seek the signal in the data – quite the opposite of players who are having to continually react and physically respond in a complex environment. Coaches must straddle both camps, recognising their history as reactive players but forced to "think slow" and use their substantial knowledge of the game they are scrutinising to guide decisions. Classifying KPIs by importance to the result, and by the ability to reliably influence them, is a critical phase in sports analysis.

The sports analytics community has accepted measures like Expected Goals in recent years, recognising that the difference between a shot at goal succeeding or failing can be inches, and is well-described by stochastic models with sufficiently accurate data. These events happen at the end of a possession chain, and many people can now accept the epistemological argument that the difference between what *did* happen, and what *might have* happened (an array of counterfactuals) is a realisable metric.

It is harder to comprehend classifying events at the start of a possession chain in the same way, because we see a cascade of counterfactuals leading to an alternative reality. That way lies madness, we think, or at least chaos. Taking lessons from chaos theory, histories are expected to diverge at an exponential rate based on minuscule differences in initial conditions, and there are several decisions each second whence an alternative history could start its evolution.

The theory of backgammon offers a solution, as long as we are willing to briefly shut our eyes to the continuing play and just perform the mathematics, much like quantum physicists in the Feynman mould. The difference between the average path from what happened and the average path from what might have happened is expressible in scoreboard terms.

5. CONCLUSIONS

Lefgren, Platt & Price (2012) cleverly analysed American football data to discover that coaches would change their gameplans in reaction to a loss, even if it was uninformative because they were expected to lose to that opponent. On the other hand a lucky or substandard win over an inferior team would not prompt as many changes. Sports analytics must become better at putting information in front of coaches after it has been normalised, had its variance reduced if possible, and is ordered by importance to success. Approaches such as that outlined in this paper could be applied to AFL and other sports, modifying the way data is collected, counterfactuals are considered, and reporting supports decisions.

Acknowledgements

I wish to thank Hawthorn Football Club coaches and particularly David Rath, Head of Coaching Services, for their valuable discussions around this work.

References

Gréhaigne, J.-F., & Godbout, P. (2011). Dynamic Systems Theory and Team Sport Coaching. Quest, 66 (1), 96-116.

Lefgren L, Platt B, Price J. (2012). Sticking with what (barely) worked. *NBER Working Papers № 17477*, National Bureau of Economic Research, Cambridge, MA.

Mauboussin, M.J. (2012). The Success Equation: Untangling Skill and Luck in Business, Sports, and Investing. Published by Harvard Business School Publishing.

O'Shaughnessy, D. (2012). The Luck of the Score: Partition of Variance for Performance Analysis of Team Sports. *Proceedings of the World Congress on Performance Analysis of Sport IX,* 37.

O'Shaughnessy, D. (2006). Possession versus Position: Strategic Evaluation in AFL. *Proceedings of the 8th Conference on Mathematics and Computers in Sport*. (Published in a condensed form in *Journal of Sports Science and Medicine*, 5, 533-540)