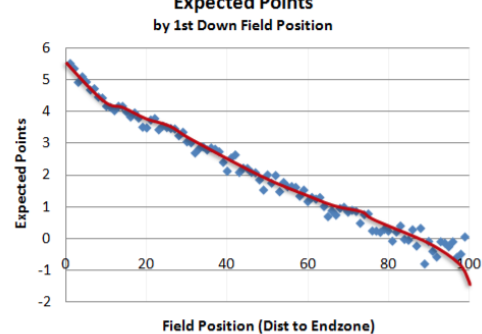
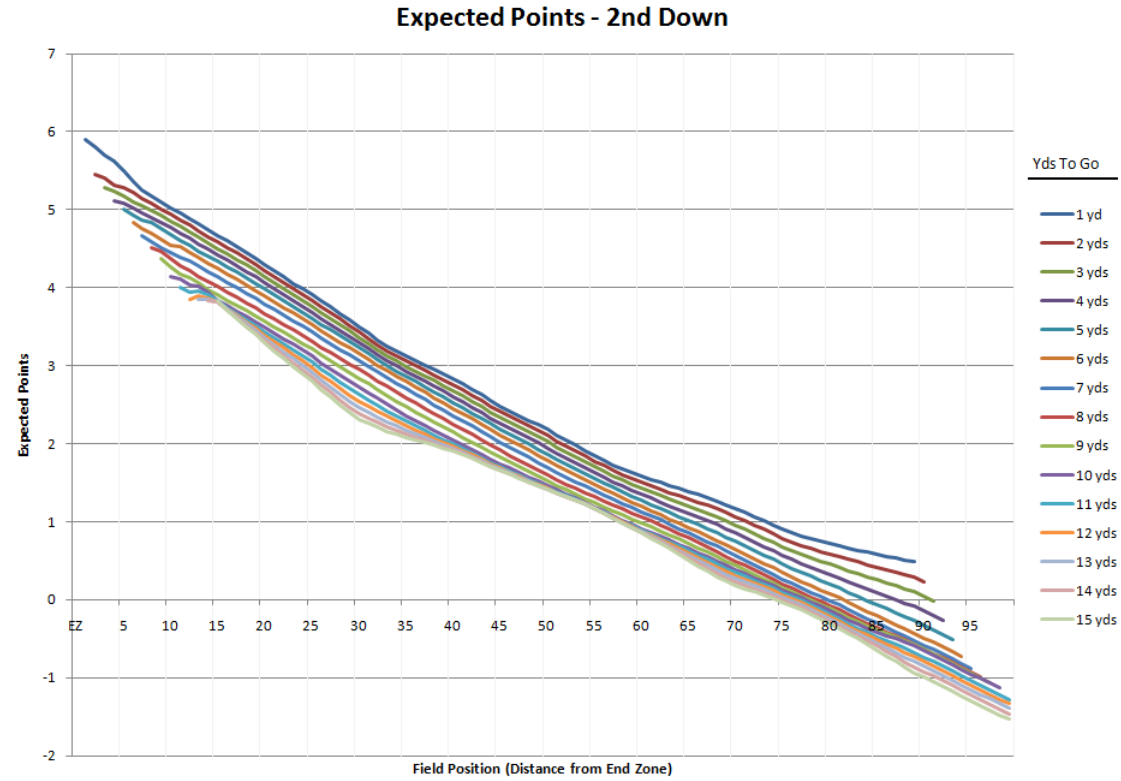
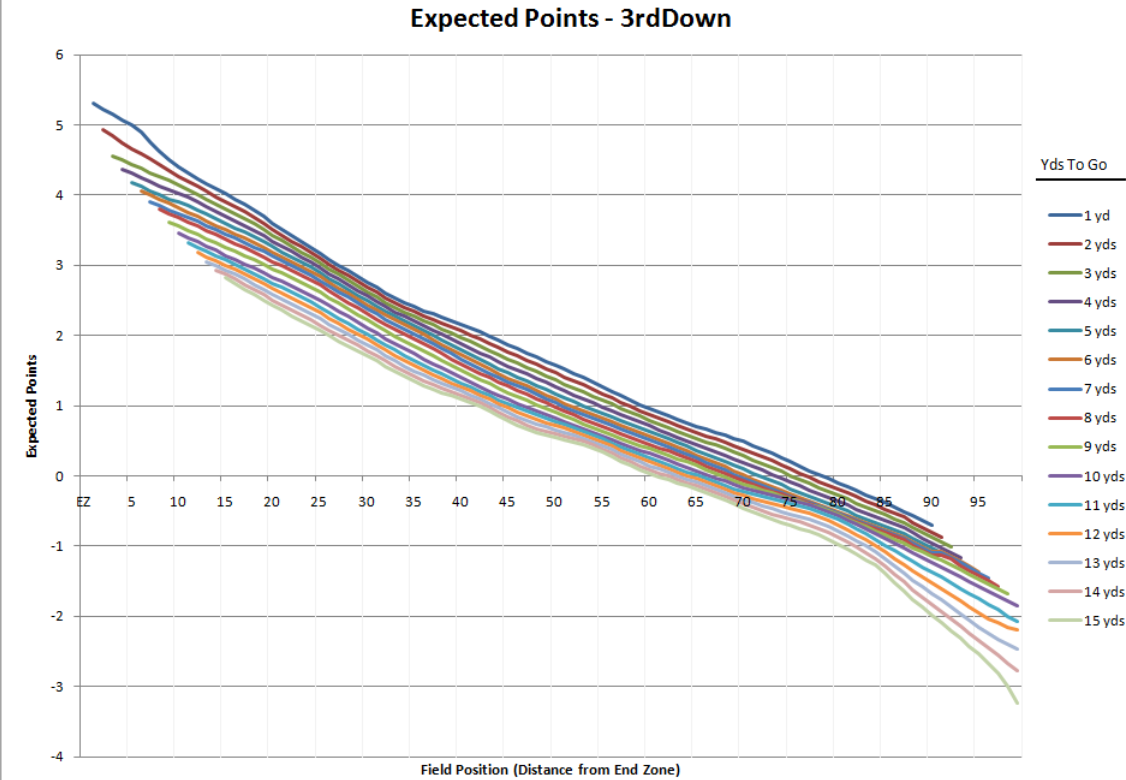
# Expected Points and Expected Points Added Explained: Brian Burke

* Football = sport of strategy + decision making.
* But comparing potential risks + rewards of various options, need to be able to **properly measure value of possible outcomes**
* **Value of a football play** has traditionally been measured in **yards gained** 🡺 FLAWED measure b/c ***not all yards are equal*** (4-yd. gain on 3rd + 3 = much more valuable than 4­ yards on 3rd + 8)
* Any measure of success must consider the “down and distance” situation.
* **Field position** = also an important consideration 🡺 Yards near goal line = tougher + more valuable at midfield + yards lost near one’s own goal line = more costly as well.
* Can measure values of situations +, by extension, the outcomes of plays by establishing an **equivalence** in terms of points.
* Start by looking back through recent NFL history at the ‘**next points scored’** for all plays
* Ex: Look @ all 1st + 10s from an offense’ own 20­-yard line 🡪 team on offense will score next slightly more often than opponent.
* If we add up all ‘next points’ scored for + against offense’s team, whether on current or subsequent drives, can estimate **net point advantage** an offense can expect for any football situation
* For a 1st + 10 @ own 20, it’s **+0.4 net points**, + at opponent’s 20, it’s **+4.0 net points**.
* These net point values = called **Expected Points (EP),** + every down­-distance­-field position situation has a corresponding EP value.
* Suppose offense has a 1st + 10 @ midfield 🡪 worth +2.0 EP.
* 5­-yard gain sets up 2nd + 5 from the 45 🡪 +2.1 EP.
* Therefore, *that* 5-­yard gain *in that particular situation* = a **+0.1 gain in EP**, where gain = called **Expected Points Added (EPA).**
* Likewise, a 5-­yard loss on 1st @ midfield creates 2nd + 15 from own 45 = worth +1.2 EP representing a net difference of -­0.8 EPA.
* Can value turnovers in same way.
* Suppose on 2nd + 5 @ opponent’s 45 🡪 fumble recovered by defense
* 2nd + 5 was worth +2.2 EP, but now opponent has 1st + 10 on own 45, worth +2.1 EP *to them*.
* Result of the play = **­2.1 EP** for original offense for a **net loss of ­4.3 EP**.
* On average, a fumble in that situation means net expected loss of a little more than 4 points.
* **To be of good use for most kinds of analysis, need the measure of success to be linear**.
* In this case, it means +2 EP = exactly twice as good as +1 EP, +4 EP = twice as good as +2 EP, etc.
* **Need linearity when analyzing decisions**.
* What would we rather have: 100% chance of +3 EP, or 60% chance at +6 EP w/ 40% chance of 0 EP?
* To answer this question definitively, each net point of advantage must be equally valuable to a team
* Problem 🡺 We all know being up by 1 @ the end of a game = just as good as being up by 50, so **not all points are equally valuable**.
* Teams well ahead sacrifice point advantages in exchange for running time off the clock, which helps them win.
* **To mitigate that problem, baseline EP values for each down-­distance-­field position situation must be created based on real game situations when points = equally valuable + time is not yet a factor**.
* Baseline EP values = therefore based *only on game situations when score was w/in 10 points in Q1 or Q3.*
* This eliminates situations like ‘trash time,’ + other distortions.
* EP + EPA have a variety of applications
* Can use EP to measure + compare relative value of runs vs. passes in various situations.
* Can tally up EPA for individual players + for teams for a more accurate valuation than from traditional stats
* Perhaps most useful application of EP = analysis of 4th down decisions, which suggests teams should be going for it far more often.
* Example of what EP values look like on 1st down:



* Expected Points on 1st downs = easy to compute b/c there’re so many 1st + 10s compared to any other down-distance combo
* For 2nd + 3rd downs, not nearly as simple as averaging next scores for each field position
* There’s considerable noise (or sample error) b/c there’re relatively few cases of each down-distance combo.
* To get reasonable estimates for later down situations, use a smoothing technique [**LOESS**](http://en.wikipedia.org/wiki/Local_regression) = fancy way of drawing a crayon through a collection of noisy data points
* Challenge = to get estimates consistent across 3 dimensions: field position, down, distance to goal



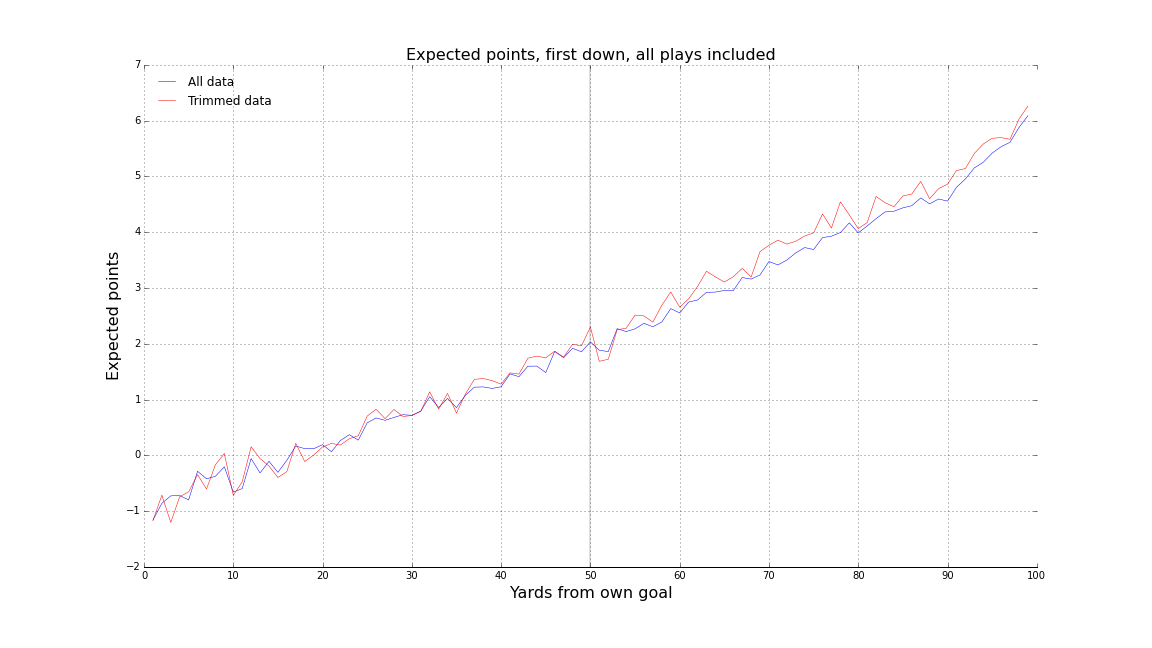


# Expected Points Part 1: Building a Model and Estimating Uncertainty

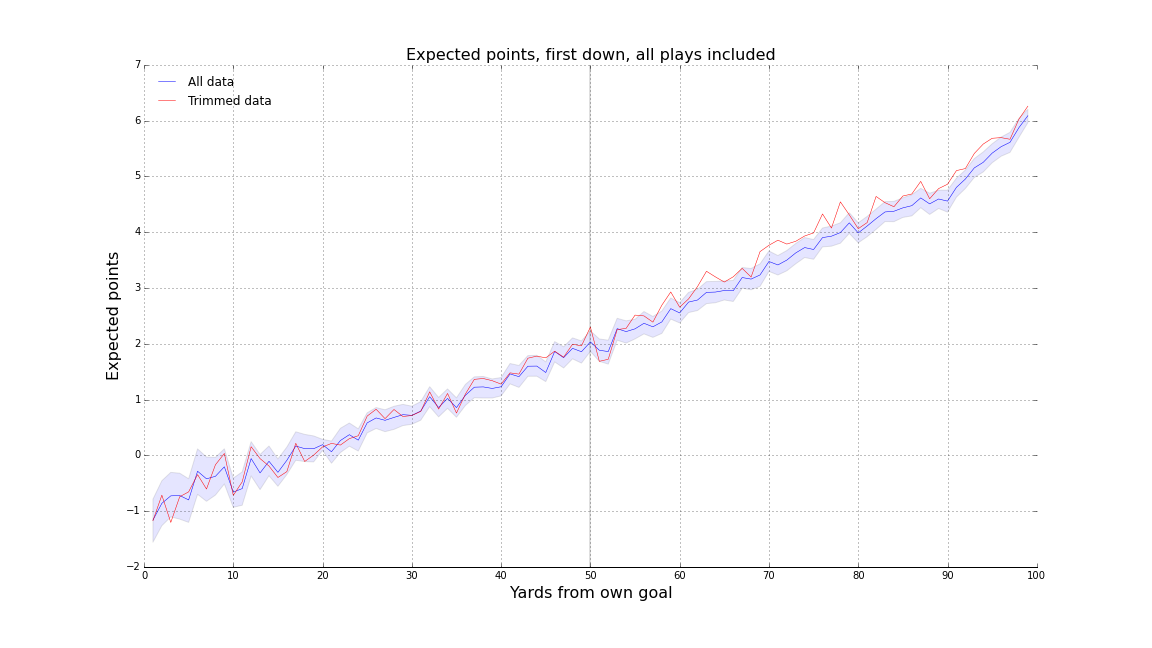
* **Expected points** got renewed attention from [FiveThirtyEight article](http://fivethirtyeight.com/features/kickers-are-forever/) on rise of kicking accuracy over time + how 4th-down decision-making could be affected.
* However, lots of EP models already exist
* 2 YouTube tutorials ([1](https://www.youtube.com/watch?v=JclgcQgPOcE), [2](https://www.youtube.com/watch?v=IDLCulWNGyk)) on building an expected points model
* Basic idea behind EP:
* **Given any combo of down, yards to go, + distance from end zone, the expected value of the points from that position = the average of every *next score* from that position.**
* Next score could be on that play via a FG or TD, could be several-to-many plays later through a successful drive.
* Could also be negative 🡪 next points scored by other team.
* Can imagine EP from one's own 1-yard line = probably negative, b/c even if you punt the ball away, opponent will probably have very good field position to start next drive + will likely get at least a FG
* Similarly, can imagine expected points on 1st + goal from opponent's 1 = somewhere between 3 + 7 b/c you'll have nearly 4 tries (barring turnovers) to score a TD or FG
* Reason to build these kinds of models = **to place a value on every position on the field to allow for in-game decision-making.**
* By being able to compare EP from a variety of possible outcomes, can choose play call that allows for maximizing # of EP
* May be game scenarios when more interested in maximizing EP (early, when an individual play may not have much impact on overall win probability)
* Building the model itself w/ Python = easier w/ indexing + grouping capabilities of [pandas](http://pandas.pydata.org/) 🡪 data manipulation + taking the mean.
* all of the code in an IPython notebook on [Github](http://github.com/treycausey/thespread/tree/master/notebooks/expected_points.ipynb) ([NBViewer](http://nbviewer.ipython.org/github/treycausey/thespread/blob/master/notebooks/expected_points.ipynb))

**Exploring the assumptions**

* # of assumptions that go into building this kind of model.
* 1) Throw out plays where score difference > 10 + those from 2nd + 4th quarters b/c teams operate differently when facing/delivering a blowout or when half is about to end.
* winning team may just run RB repeatedly towards end of game, not really trying to gain yards or score more, which could distort effects of these plays on points scored.
* Can presenting how assumptions change analyses = 1 way of measuring effects of assumptions, but also a good way to see how **robust** your conclusions are to changes in the data.
* EP as a function of field position on 1st down w/ + w/out these plays removed.



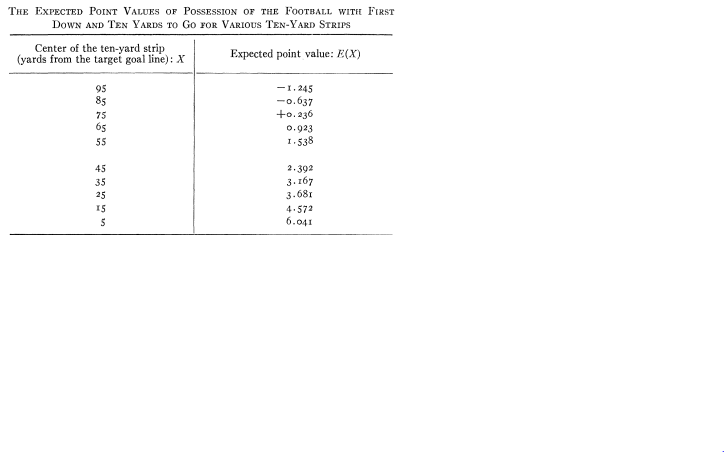
* Surprisingly, not much of a difference 🡺 trimmed data produces slightly higher estimates of EP than complete data in the opponent's half of the field.
* But *how much* of a difference is "not much"?
* Can use a local regression **smoother** **LOESS** to remove bumps + get better sense of 'true' EP contained in those noise lines.
* Can also use [**bootstrap**](http://stats.stackexchange.com/questions/26088/explaining-to-laypeople-why-bootstrapping-works) to build CI’s around those EP values
* EP above only represent plays we've actually seen happen, but **they are just an estimate**.
* Want to make some **inferences** about the **range** of possible outcomes we DIDN'T see.
* Assume the plays we saw = drawn from some distribution of outcomes from alternate universes
* Can simulate this distribution by taking repeated samples w/ replacement from plays we DID see
* **This procedure doesn't assume anything about the distribution of the statistic we're interested in**
* CI built up gives some idea of how much variation we might expect in our estimator (EP) if we were to keep sampling from the distribution that generated the observations we already have.
* Look @ the 95% CI for the original EP:



* EP estimated using only 1st & 3rd quarters + close games falls outside of our CI quite often in the opponent's half of the field
* Note also uncertainty around EP is at its greatest the closer you get to your own end zone, + the least the closer you get to your opponent's endzone
* This = intuitive, but always good to know if an estimator has constant variance or not.

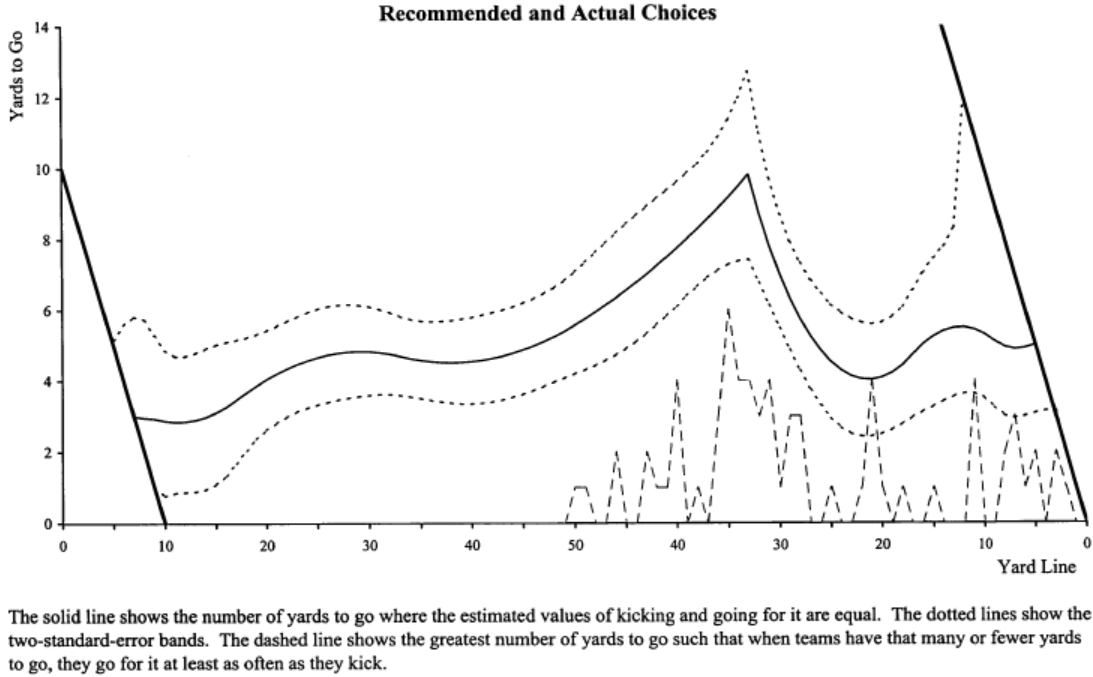
# OPERATIONS RESEARCH ON FOOTBALL

* Based on a census of 8,373 plays from 1st 56 games of 1969 season, this note calculates, for various field positions, EP values of possession of the football on 1st + 10 + discusses strategic implications of these values
* Several operations-research studies have been performed[1] on strategy + tactics for baseball.
* No comparable coherent effort has been published concerning professional football (1971), although a brief note appeared in 1954 [2]
* Study = data on the 56 games played in 1st half of 1969 NFL, wherein each 8,373 individual plays in was coded, punched, + entered into a CPU + all analyses were made on this data base.
* Comparative analyses of results from the total sample w/ results from subsamples, as well as routine computations of standard errors, indicated that, for most questions, this sample = sufficiently large.
* # of statistical analyses were performed, most in results, or to indicate that optimum strategies appeared to have been evolved, apparently by intuitive methods.
* However, for 1 item , interesting numbers + an indication of a change that should be made in present football strategies were derived
* Analysis concerns the **expected value of possession of the football** on 1st + 10, @ any particular point on the playing field.
* Basic formula for expected value is, of course,  w/ # of possible outcomes = 103, w/ 1st 4 being touchdown (X = +7), FG (X +3), safety (X = -2), + opponent's TD (due to fumble or interception) (X -7).
* Remaining possible outcomes consist of eventually turning over the ball to the opponents at 1 of the 99 possible points on the field.
* It is then assumed that the appropriate value of X here is **-E(X) 🡺** leads to a system of 99 equations in 99 unknowns.
* Since not enough data were available to determine these 99 probabilities w/ adequate accuracy, field was divided into 10 strips, namely, 99-91 yards to go, 90-81 yards to go, etc.
* These data sets are ID’ed by their mid-points in Table I.



* The smallest # of DP’s in any set = 57 (95 yards to go) + largest = 601 (75 yards to go).
* This condensation led to a system of 10 equations in 10 unknowns w/ results presented in Table I.
* Analysis was based on a study of 2,852 1st + 10plays.
* Independent calculation was performed for the subset of 1,258 1st downs immediately following turnover + average absolute difference was < 1/4 of a point.
* **Psychological differences as well as statistical fluctuations may have affected this difference.**
* There are some obvious inadequacies in the analysis.
* The several hundred situations starting on the 20 probably should’ve been separated into an 11th category.
* Apparently, 1st + 10 on own 20 has an expected value very close to 0, which indicates wisdom on the part of the rules-makers.
* It would have been nice to check to see how exactly this comes out.
* More significantly, the value of a kick following a score was ignored, although this is presumably negative.
* Free kick from the 20 following a safety obviously has negative expected value, but this occurs w/ sufficiently low frequency as to have little effect on the analysis.
* More important, kickoff from the 40 following a score is a very frequent occurrence + appears to have a negative expected value.
* The #’s in Table I are obviously in error by perhaps several tenths of a point but nonetheless have some obvious qualitative value.
* Basketball: good strategy is built about accepted fact that a rebound = worth about 1 point.
* There’s a similar qualitative value in having an approximate knowledge of the value of having the ball at a particular point on the field.
* More significantly, it appears that 1 technical decision is regularly being made incorrectly in
* Specifically, the **negative** value of having the ball with 1st + 10 very close to goal line has been ignored (i.e. situation where a team has ball, 4th + goal, in which case a FG is routinely attempted)
* Even in desperate situations (say 6 points behind in the 4th, when a FG is rather worthless), the usual tactic under a 4th + goal situation would be an attempt to pass
* Rules say “if a pass is incomplete in the end zone, the ball is brought out to the 20”
* We recommend the ball be rushed under these circumstances
* If the decision is to kick, there can be 2 possible outcomes.
* 3 points made w/ probability varying from about 0.98 if ball is well centered + team has a good kicker to perhaps 0.6 if angle is very bad + kicker is poor.
* From this, expected value must be subtracted the negative expected value of the ensuing kickoff
* If the kick is missed, the ball is brought out to the 20 where expected value is ~0
* If ball is run, there is a probability of perhaps 0.2 or 0.3 of making virtually 7 points + if run fails to make a TD, the ball is turned over so close to the enemy’s own goal line that there is something between 1-2 points of **negative** value for the opposition.
* Unfortunately, there is insufficient data to pin down exact variation in this negative value w/in the last few yards of the field.
* But even if the negative value is only 1.25 points, which appears to be the average for the 1-9 yard lines, it seems clear the analysis on a 4th + goal situation = very different from 4th + 4 on, say, the 20.
* In the latter case, the kick seems well justified.
* A similar analysis is pertinent in evaluating the 'coffin-corner' punt
* Another other area in which most teams seem to be making an incorrect choice of strategy**: strategy of calling TO’s during last 2 minutes of either half**
* No statistics have been gathered to support this, but the error (as we assume it to be) derives from a naive misunderstanding of **decision theory**.
* If one is considering calling a TO, the **Type I error (FP)** is in calling it when one shouldn’t, + the penalty arises when the ball is turned over to the opponent before time expires.
* This may happen through a fumble or interception, running out of downs followed by a punt, a missed FG, or an actual score after which the opponents again get the ball.
* **This Type I error appears to be a very frequent occurrence.**
* Ex: Cleveland-Oakland on November 8, 1970, in which Cleveland, attempting to break a tie, called a TO + subsequently lost the ball by interception
* BLANDA had enough time to win the game by kicking a FG
* The **Type II error (FN)** consists of not calling a time out when one should w/ the naive assumption that appears to have been that the penalty for this is running out of time 🡺 **This is wrong**.
* Penalty for a Type II error = running out of time **when one still has time outs left to call**, believed to be an extremely rare occurrence, if indeed it has ever occurred.
* We assert the Type I error (calling TO when one should not + being penalized for it) has been made many times; whereas, the Type II error (failing to call a TO when one should) = extremely rare.
* Specifically, therefore, we recommend a team behind 7 or less should NEVER call TO where > 30 seconds left (if it has the ball) or where > 1 minute left if opponents have the ball.
* We are referring here ONLY to TOs called for the exclusive purpose of stopping the clock.
* 1. GEORGE LINDSEY, review of EARNSHAW COOK, Percentage Baseball, Second Edition, The MIT Press, Cambridge, Mass., 1966, in Opns. Res. 16, 1088-1089 (1968).
* 2. CHARLES M. MOTTLEY, "The Application of Operations-Research Methods to Athletic Games," Opns. Res. 2, 335-338 (1954).

# Going for It on Fourth Down (Burke)

* 4th + goal from the 2 in Q1 🡪 Most coaches kick FG, a virtually certain 3 points.
* 4th + goal from the 2 is successful = 3/7 times, assuring the same # of EP, ***on average***, as the FG.
* Plus, if attempt at a TD is unsuccessful, opponent is left w/ the ball on the 2 or even 1 yard line.
* If FG is successful, opponent returns a kickoff which leaves them usually around the 28
* Should be obvious that **on balance**, going for TD = better decision 🡺 case made by economist David Romer, author of a 2005 paper called "Do Firms Maximize, Evidence from Professional Football."
* Paper = an analysis of 4th down situations in the NFL + is quite possibly the most definitive proof coaches = too timid on 4th down.
* Romer's theory = coaches don't try to maximize team's chances of winning games as much as maximize job security.
* Coaches know if they follow conventional wisdom + kick but miss, though is = the players just didn't make it happen.
* But if they take a risk + lose, *even if it is on balance the better decision* = criticized, or job security will be put in question.
* Coaches = concerned about Monday morning criticism
* Down by 3 very late in the 4th against winless + fatigued Dolphin defense, former Ravens coach Brian Billick chose to kick a FG on 4th + goal 1 ft. from the end zone.
* Dolphins went on to score a TD in OT + Billick's explanation: "Had we done that [gone for it] after what we had done to get down there + [not scored a TD], I can imagine what the critique would have been today about the play call."
* Billick, a 9-year veteran head coach + Super Bowl winner, was more concerned about criticism from Baltimore Sun columnists than the actual outcome of the game.
* Would rather escape criticism than give team best chance to win.
* Romer's paper considers data from 3 years of games.
* To avoid the complications of particular "end-game" scenarios w/ time expiring in 2nd or 4th quarters, he considers only plays from 1st quarter of games.
* So recommendations **should be considered a general baseline for the typical drive**, + NOT a prescription for every situation.
* Romer's bottom line:
* 
* Solid line = when it’s advisable for a team to attempt the 1st down rather than kick.
* According to the analysis, it's almost ALWAYS worth it to go for it w/ < 4 yards to go + recommendation peaks at 4th + 10 from an opponent's 33
* Romer basically measures the EV of the NEXT score.
* Say it's 4th + 2 from the 35 🡪 He compares the value of attempting a FG from the 35 w/ the point value of a 1st + 10 from the 33 (multiplied by probability of actually making the 1st down.)
* He also recognizes a FG isn't always worth 3 points, + a TD isn't always worth at least 6.
* **The ensuing kickoff gives an EPV to the opponent 🡺** There is a point value to having a 1st + 10 from own 25
* **1 weakness of the paper = it dismisses the concept of risk as unimportant**.
* Romer says long-term point optimization should be the ONLY goal, so coaches should *always* be risk neutral
* But if the level of risk aversion were *actually* considered, we might find coaches = more rational than he concludes.
* But the paper makes a very strong case coaches should go for it on 4th down far more often than they currently do, + job security for coaches seems to be primary reason why they don't.
* At a meeting w/ some researchers making the case for more aggressive 4th down decision making, Bengals coach Marvin Lewis responded, "You guys might very well be right that we're calling something too conservative in that situation. But what you don’t understand is that if I make a call that's viewed to be controversial by the fans + by the owner, + I fail, I lose my job."
* It would be great if a coach came along + rarely kicked + it would be gamble, but if Romer + others are right, chances are the coach would be successful + the rest of the NFL would have to adapt.

# Decis

# ion Theory in Football By Brian Burke

* In **Decision Theory** 🡺 generally 2 kinds of analysis + rarely are the 2 things the same
* **Descriptive analysis** = what people ACTUALLY do
* **Prescriptive analysis** = what people SHOULD do
* When I use a win probability model to evaluate 4th down decisions = prescriptive analysis.
* Trying to explain what coaches are actually doing = descriptive analysis.
* Coaches != CPUs + are subject to all imperfections of human decision making.

**NFL Orthodoxy**

* NFL football = evolved as extremely conservative game 🡪 i.e. coaches adhere to wisdom passed down from previous generations + are reluctant to deviate from the established orthodoxy.
* In the real world, away from sports, this approach usually makes sense, as it’s not bounded by sidelines, end zones, + 15-minute quarters + is highly uncertain + far less predictable than we'd like to think.
* Makes sense to adhere to what is known to work rather than try to engineer an optimized outcome in a highly uncertain environment.
* **But in football, we have the stats, know the probabilities**, + **know the possible consequences**
* 'Conservative’ = therefore often NOT the best approach
* Possible that coaches adhere to same orthodoxies is b/c they aren't conscious of the level of certainty available to them

**Minimax**

* 1 of the more conservative approaches is the [**minimax**](http://en.wikipedia.org/wiki/Minimax)**criterion** 🡪 says *pick the option that assures you the highest* ***minimum utility****.*
* Let's say you have the choice between going on a picnic + going bowling + you’d really rather go on the picnic, but it might rain, so the payoff matrix would look like this:

|  |  |  |
| --- | --- | --- |
|  | No Rain | Rain |
| Picnic | 4 | 0 |
| Bowling | 1 | 1 |

* If it doesn't rain, picnic pays off, but if it rains you've lost the afternoon.
* Bowling = not as much fun, but it wouldn't matter if it rains.
* Minimax says go bowling b/c its minimum payoff = 1 while picnic’s minimum payoff = 0

**Minimax-Regret**

* **Minimax-regret criterion** seeks to minimize potential regrets.
* Imagine coming out of bowling alley greeted by a sunny blue sky.
* In this case, if you go bowling + it doesn't rain, **you've gained 1 unit of utility but lost out on 4 units**, for **a net regret of 3**.
* If you go on a picnic + it does rain, you've **gained 0 utility but lost out on 1 unit**, for a **net regret of 1**
* **T**o minimize regret, choose the picnic.
* We haven't mentioned the weather forecast yet 🡪 **These methods = best relied upon when there is a very high level of uncertainty in the "states of nature" that will determine the payoffs**.
* Consider football 🡪 coach has 3 plays that make sense for a given situation + opposing defense can call 1 of 3 kinds of defenses for a Hypothetical Football Payoff Matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Def X | Def Y | Def Z |
| Play A | -4 | 4 | 12 |
| Play B | -2 | 3 | 8 |
| Play C | 3 | 2 | 1 |

* *This is not game theory. We're not looking for a Nash equilibrium*.
* OC is thinking of the defense as a "state of nature" 🡪 something he has no control over + is difficult to predict.
* In this case, both Plays A + B have the possibility of negative payoffs.
* **Play C guarantees at least a payoff of 1, + therefore would be the MINIMAX decision.**
* **Regret method says something different.**
* Assume defense had called Def X 🡺 best payoff possible given Def X = 3 w/ Play C, so had we called Play C, there’d be no regret.
* But had we called Play B, we would’ve earned a -2 payoff, which = a regret of -5.
* In other words, we *could* have had 3, but we got -2
* Had we called Play A, we would’ve earned a -4, which = a regret of -7.
* If we repeat the regret calculation for each possible defense, we get a whole new regret matrix:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Def X | Def Y | Def Z |
| Play A | -7 | 0 | 0 |
| Play B | -5 | -2 | -4 |
| Play C | 0 | -2 | -11 |

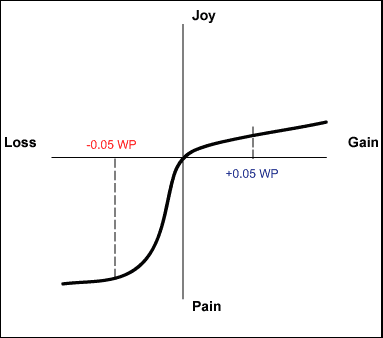
* Given this regret matrix, the **minimax-regret criterion** looks for the choice that **assures** us of the **best worst-case scenario**.
* For Play A, worst regret is -7, for Play B = -5, + for Play C = -11.
* Therefore, we'd pick Play B b/c = least costly in terms of maximum possible regret.
* Of course, coaches/anyone else would never actually draw up a matrix + do math to make a decision
* But just like in the picnic-bowling example, our brains are attempting poor analog versions of these kinds of decision criteria, + emotions play a large role.

**Expected Utility**

* What if we reduce the uncertainty in the defense?
* Can't predict exactly which one we'll see, but can estimate probabilities we can expect each defense
* **Expected utility** of a choice = the **weighted average of the possible payoffs**.
* For simplicity, say each defense is equally likely w/ 1/3 chance.
* Can estimate EU for each play choice
* Above, **EU for Play A = (1/3)(-4) + (1/3)(4) + (1/3)(12) = 4,** EU for Play B = 3, + EU for Play C = 2.
* **EU method says Play A is the best choice.**
* The 3 methods above each call for a different decision + each is logical + consistent *in its own way,* but there is only 1 truly correct method in football, only 1 prescriptive analysis.
* Remember, in football we can KNOW the probabilities + the payoffs, or at least have a solid league-wide baseline for them.
* \*\*\***The EU method is the only correct method\*\*\***
* The math behind EU analysis couldn't be any easier = 5th grade arithmetic
* Challenge = **Knowing the utility function**.
* Yards, + even points, don't equate to utility.
* 7-yard gain = usually good, but relatively useless on 3rd + 8
* FG doesn't help late in the 4th when down by 7.
* Fortunately, there’s **win probability (WP) =** only correct utility function for ANY game, including football
* Winning = all that matters, whether by 1 or 100 points.
* **WP = also perfectly linear** 🡪 essential to valid EU analysis.
* A 0.40 WP = exactly 2X as good as a 0.20 WP, 0.80 WP is 2X as good as 0.40 WP.

**Prospect Theory**

* Even if coaches were to somehow use expected WP analysis when making decisions (say by using 'quick reference' cards like they sometimes do for 2-point conversion decisions), it's likely they still wouldn't be very rational.
* [**Prospect theory**](http://en.wikipedia.org/wiki/Prospect_theory) = **people fear losses more than they value equivalent gains**.
* Humans evolved w/ a tendency to try to avoid loss + are usually more upset w/ ourselves when we misplace a $20 bill than happy when 1 falls out of the laundry.
* This tendency has been borne out time and time again in clinical experiments and other studies.
* In football, this means decisions = warped b/c coaches would fear a loss in WP more than an equivalent gain in WP.
* According to prospect theory, the "joy" from a 0.05 gain in WP < the "pain" from a 0.05 loss in WP.

[](http://1.bp.blogspot.com/_ksxjg7CFQxA/SoNufIMregI/AAAAAAAAJgc/OA_d7U3BTrU/s1600-h/prospect+theory.png)

* This asymmetry would affect tactical decisions in many ways, but most obvious may be 4th down doctrine
* Say a team finds itself in a situation where punting = result in a 0.50 WP, but EU analysis says going for the conversion = result in a net 0.55 WP.
* If goal = win the game, **correct decision = to go for it. Period.**
* The analysis isn't so straightforward for the coach (even if he could do all the math on the spot).
* Say failed conversion results in a 0.45 WP + successful conversion results in a 0.65 WP.
* A 50% chance at successful 4th down conversion therefore results in a net 0.55 WP.
* But a coach sees 0.45 WP as a possible loss of 0.05 WP + sees the 0.65 as a gain of 0.15 WP.
* B/c he fears the loss far more than he values the potential gain, even one 3X as large, he'll prefer the sure-thing option and punt.
* Further, **it's possible to actually measure risk aversion of coaches by comparing WP advantages in situations where they went for a conversion to WP advantages in situations where they forego the conversion attempt.**

**An Advantage**

* The coach who can resist this human tendency + make decisions based purely on EU will have an advantage
* How big? 🡪 Just by following a pure EU analysis on 4th down, a coach would win **an average** of an extra 1.4 games per year (calculated this based on a play-by-play database from the past 9 seasons)
* For each 4th down in which a team kicked either a FG attempt or punt, calculate the difference between going for it + kicking.
* Wherever the difference was positive, sum the increase in WP for going for it
* Grand total for nearly 2400 games was +203.1 WP, which equates to an increase of 0.17 WP for every game.
* But since there’s 2 teams in every game, this means we need to halve that = 0.086
* Bottom line: A pure EU approach to 4th down decisions would increase a team's chances of winning a game from 0.50 WP to about 0.59 WP 🡺 equivalent to an extra 1.4 wins per season (0.086\*16).
* That's a bold claim, but if we trust the WP model (nothing more than a smoothed empirical observation of how often teams actually won in given game situations in real NFL games), claim is not so bold.
* It's not a *perfect* model, but **the**[**errors are unbiased**](http://www.advancednflstats.com/2009/07/win-probability-model-accuracy.html) (**overestimates as much as it underestimates**)
* Still, if a coach only followed the EU recommendations when WP for going for it was > 0.05 more than WP for kicking, his team would still benefit by an extra 0.8 wins per season.
* That's nothing to sneeze at in a 16-game season.