Advantage Theory: The (Un)Importance of Tackling

Submission to the Coaching Track

Jeremy Brooks
Carson Carrico
Chris Clement, Concordia Stingers
Tommy Denison, Edmonton Elks

1. Introduction

We begin with a radical thesis: tackling doesn't matter.

All that matters is having enough defenders to fit the run gaps - the rest is commentary. Over the last 10 years Chris has been working on a philosophy of football that he calls Advantage Theory. Tommy Denison worked on his version, Numbers & Leverage, in parallel over the same period until their collaboration began in 2020. They combine for 38 seasons of coaching experience, and have begun a project charting over 10,000 plays over the last 5 seasons and across 4 different leagues. They have come to an overarching and unified theory of the nature of the game. Today we discuss a smaller element of that: when predicting the success of any offensive play the preeminent feature is not tackling. Indeed, it is also not men in the box, formation, scheme, or any of the other areas of previous research. The key element is the advantage state of the offense. Advantage is a principle that applies to all offensive plays, but here we are restricting ourselves to discussion of rushing plays.

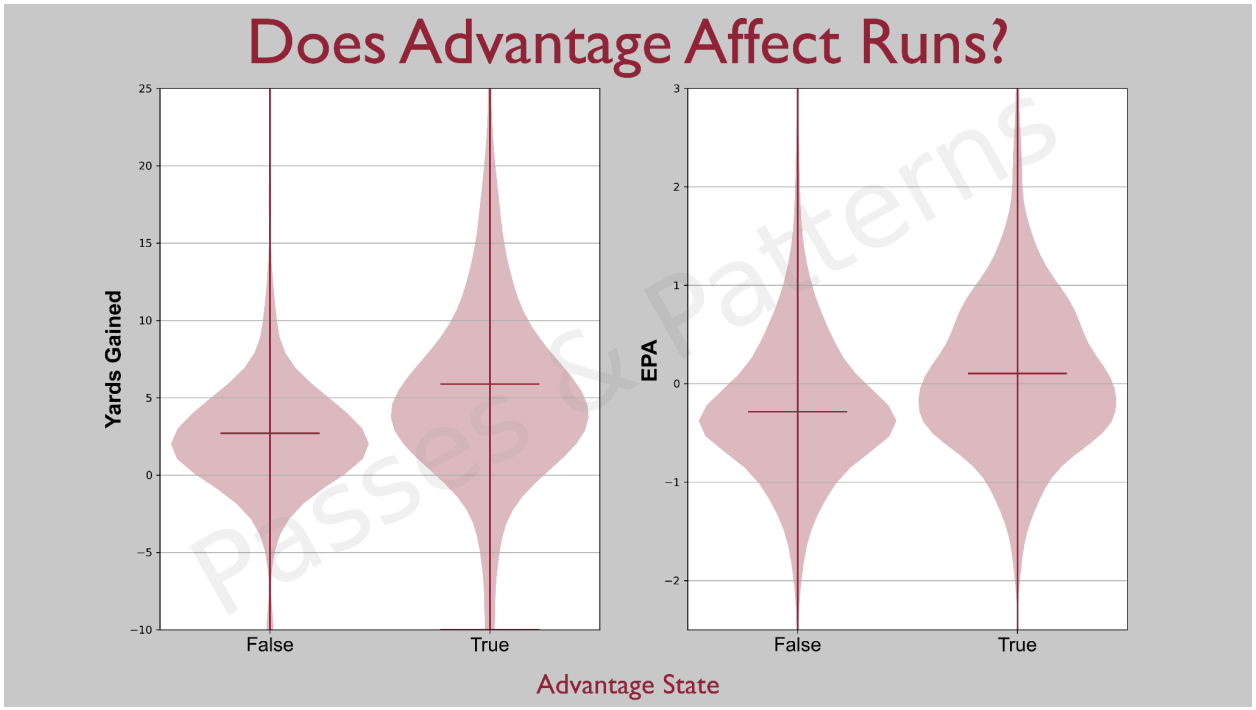
2. Advantage Theory

To understand that tackling doesn't matter requires understanding what does. Filling gaps is the most important thing because runners do not run behind their blockers, they need to find an open gap and go through it. No open gap, no successful run. It isn't necessary to defeat the blocker, penetrate the backfield, and make the tackle. A defense that holds its gaps will hold the runner. For this exercise we restrict ourselves to runs where the quarterback does not create an extra gap, either by running with the ball, blocking, or through certain, though not all, option plays. The fundamentals of advantage theory hold in all those situations, but it complicates the technical elements.

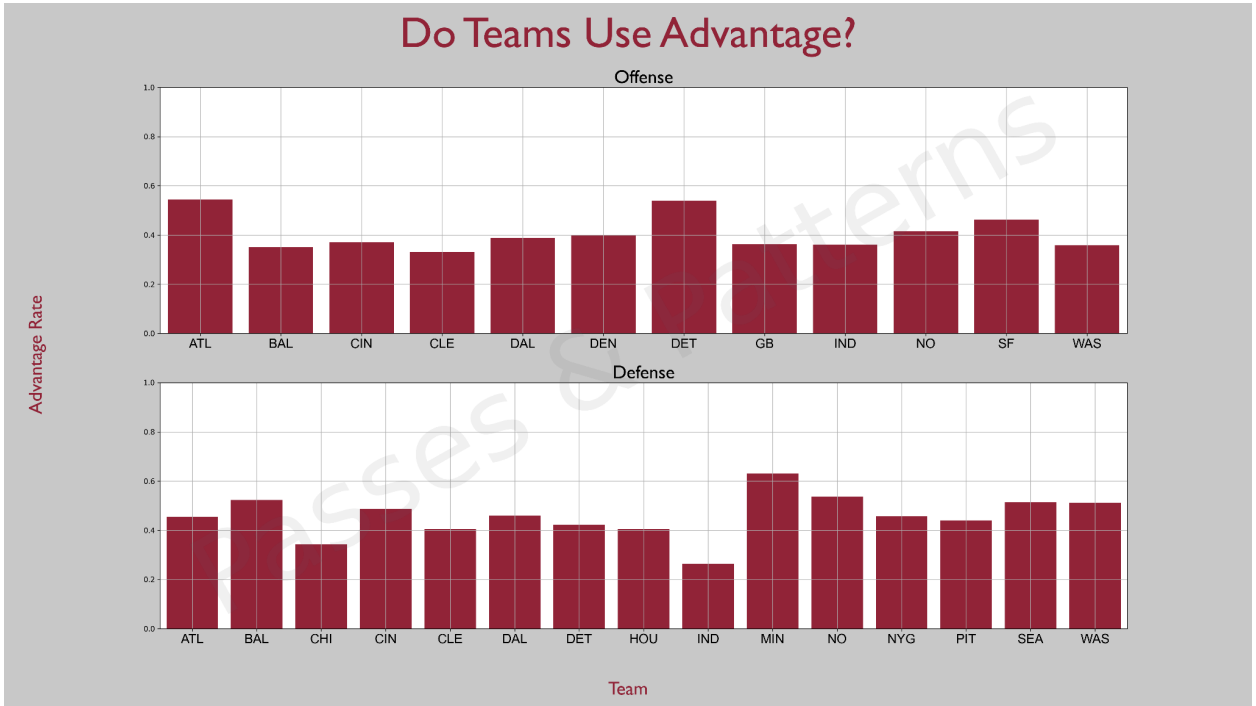
This work uses yards and not EPA as its primary KPI. This is an intentional choice. Plays do not gain EPA, plays gain yards. The EPA value of a given number of yards is context-dependent, and we need to isolate plays from that concept. On any given play the offense is trying to gain as many yards as possible, and the sum of a large number of plays leads to different distributions of outcomes in yards for different concepts, shown in the figure below. It is the responsibility of the playcaller to understand these distributions in terms of yards and thereafter look to maximize EPA. *Ad absurdum*, QB sneaks are a high EPA, low yardage pay. It would be asinine to repeatedly call sneaks because they have high EPA.

In the name of brevity this project focuses only on rushing plays where a handoff is tagged, though the fundamentals of the theory are applicable to all plays. A sample of those plays had their advantage state manually tagged (n=2383), and the results are shown in Figure 1. The difference in the mean run gain is 3.4 yards, with advantageous runs gaining 6.5 yards/play while disadvantageous plays gain 3.0 yards/play. Additionally, the difference in

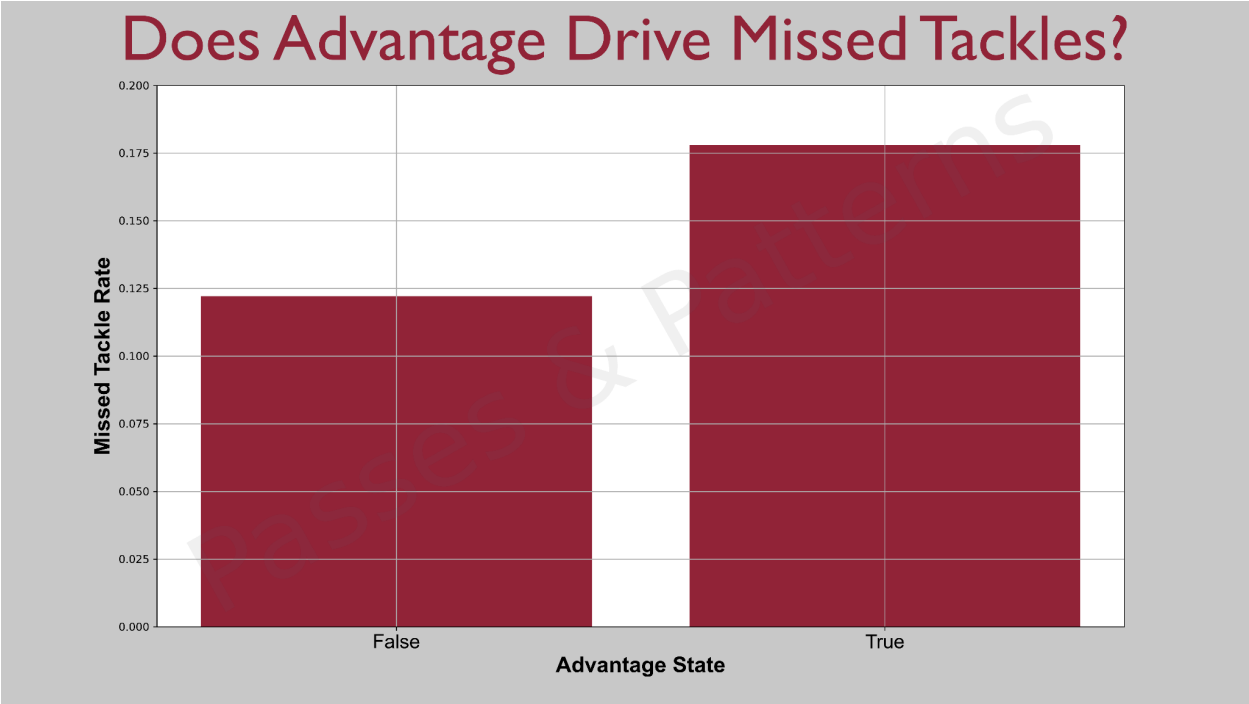
mean EPA is 0.39 EPA/play, a result consistent with Chris & Tommy's other work in other leagues.



While advantageous runs are much more effective than disadvantageous runs, and are, to some extent, within the offense's control, teams do not structure their offenses to consistently pursue advantageous runs. The variation in advantage rate between offences is small because teams are neglecting its importance.

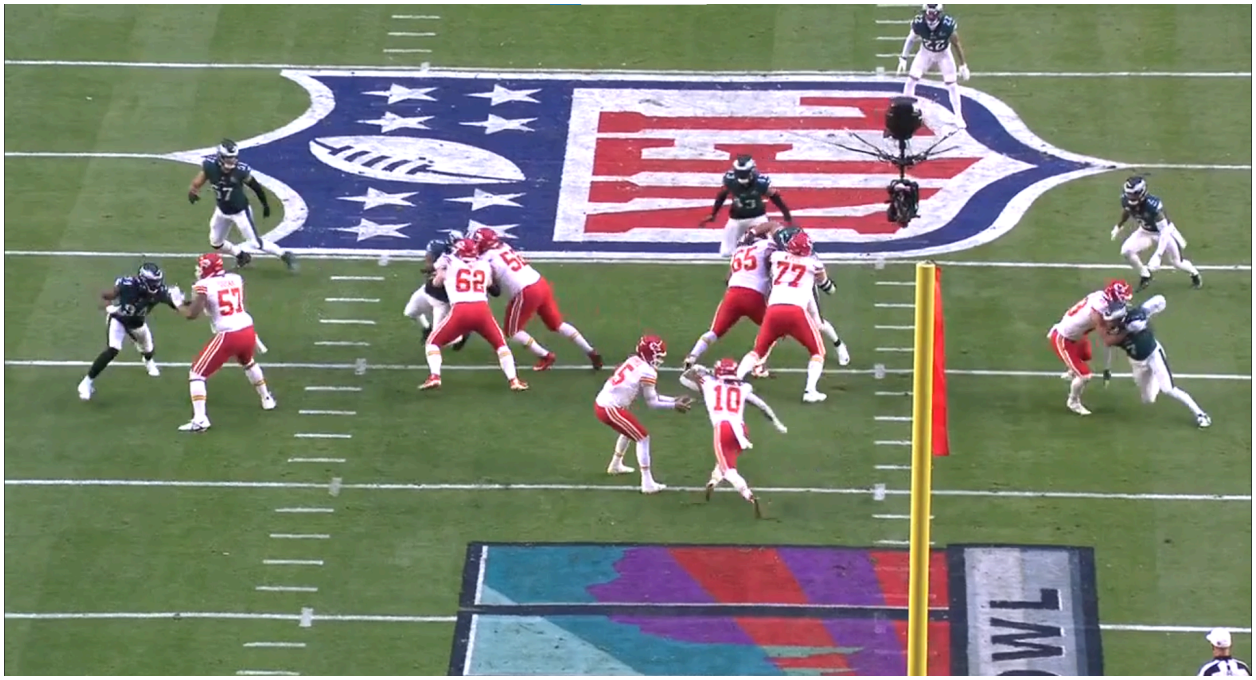


Additionally, advantage is a strong predictor of missed tackles, as when an offence runs with advantage missed tackle rate increases from 12% to 18%. Missed tackles, along with yards gained, are results of running plays with advantage, *ergo* it is missing the point to focus on missed tackles, the real problem is upstream. Teams, both offensive and defensive, need to focus on controlling the advantage state play to play.



3. Advantage method

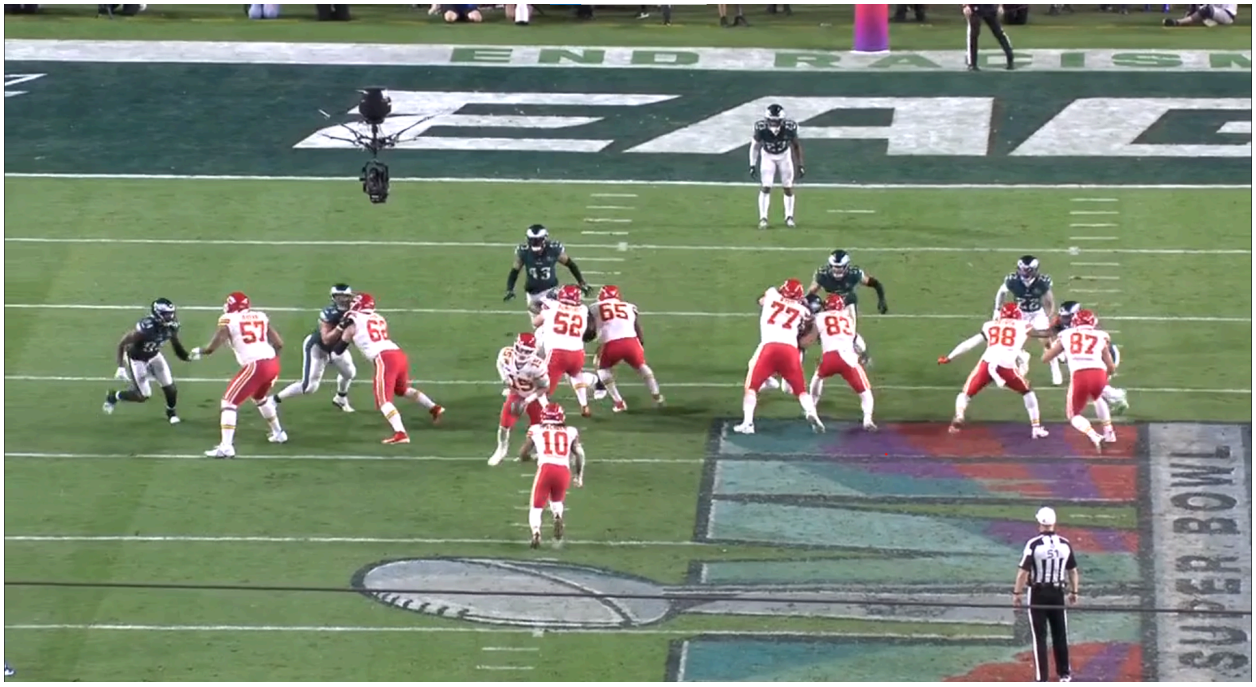
When manually tagging advantage to plays, the tagger aims to find if the offence has more gaps created than defenders reasonably positioned to account for them. This is work heavily guided by experience about the ground that defenders can cover, and as such is slow to generate. Figure 3 gives an example of a run without advantage.



There are 6 defenders in a zone scheme[1], creating 7 gaps. There are also 7 defenders positioned to defend these gaps. To those schooled in football the run fit is plain to see and all the gaps will be filled posthaste. Although this example is straightforward to see the gap fits, any run where there are enough defenders to fit the gaps is considered to not have advantage.

While determining advantage is a question of how many players are participating in the run on either team, typical "box count" metrics do a poor job of understanding whether or not to count players on the periphery - safeties and apex players, and prior works have found that existing metrics are not good predictors of rushing success. There is a nuance that so far can only be appreciated with the benefit of experience and practice.

[1] This paper does not interest itself with nonsense discussions of zone, duo, split zone, search, stretch, zone slice, inside zone, mid zone, wide zone, outside zone, zone lock, or any distinctions without differences..



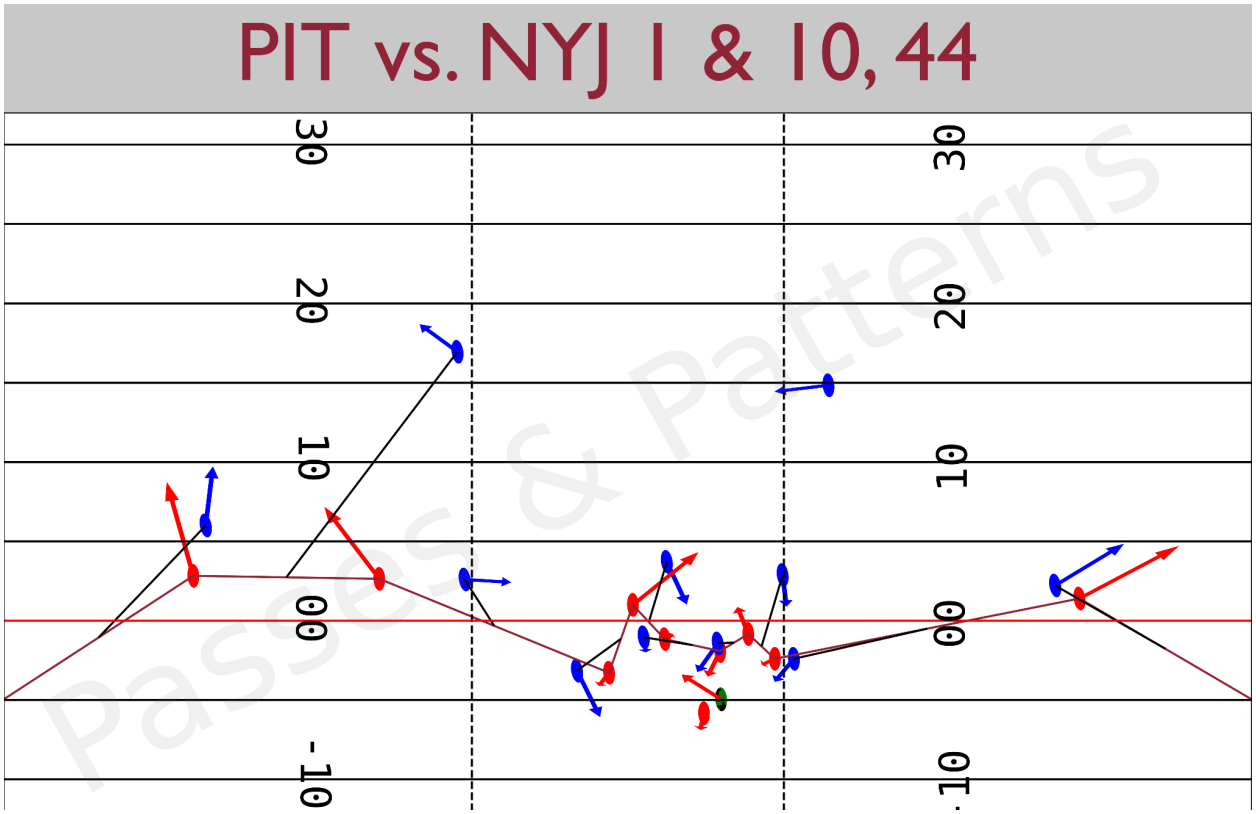
Conversely, Figure 4 shows a run *with* advantage. 8 blockers create 9 gaps, with only 8 defenders to fill them. The 3 linebackers have four gaps to fill - the boundary A, and the field B, D, and F. The defence simply cannot fill all those gaps. Without a significant failure from a blocker or a superlative defensive play this will be a positive play.

Advantage theory is highly resistant to factors commonly considered important to rushing success - scheme, front, personnel, most down & distance situations. Chris & Tommy have employed their methods to great effect with different teams, with demonstrable results.

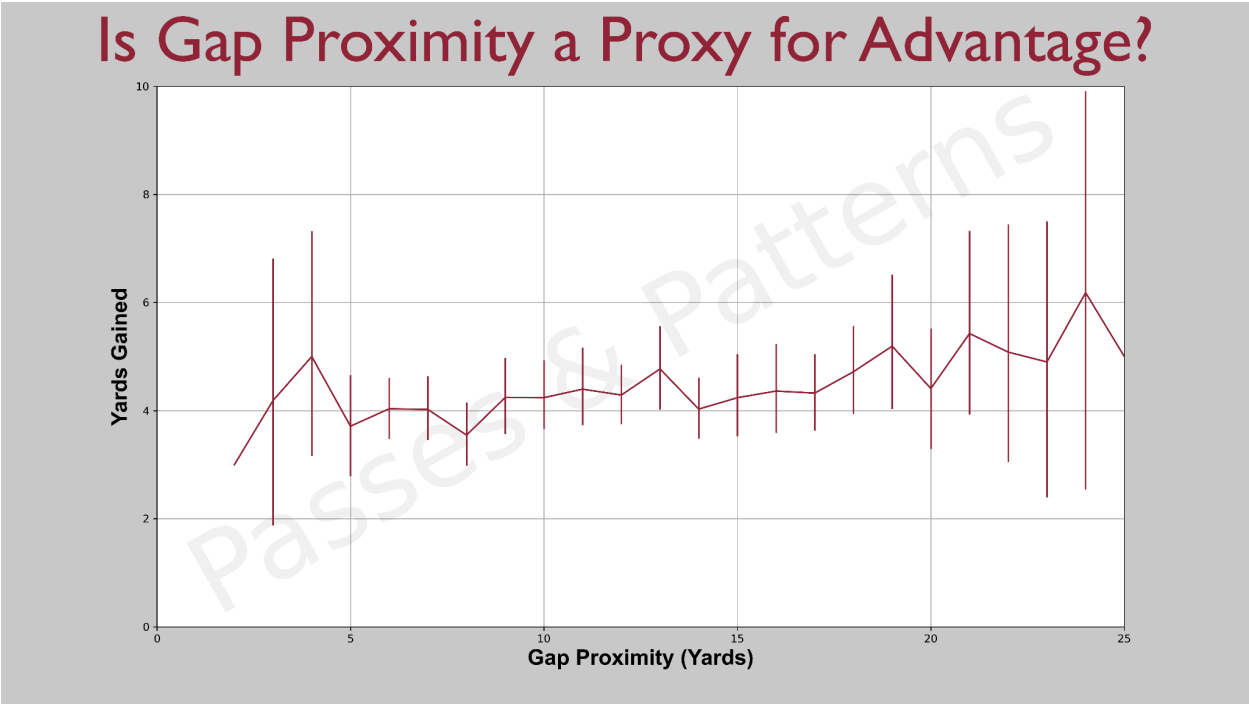
4. BDB method

What follows next is the economist's dilemma: how to turn something that works in practice and make it work in theory. Our first approach focused on identifying the specific frames tagged as the moment of handoff, that being both a clear pont of reference, and the decision point on option reads. Thereafter we know that there are 9 potential blockers, and thus 10 gaps. We add two dummy points on the sidelines at the depth of the ball carrier and use a travelling salesman algorithm to find the most efficient path to define the 10 gaps.

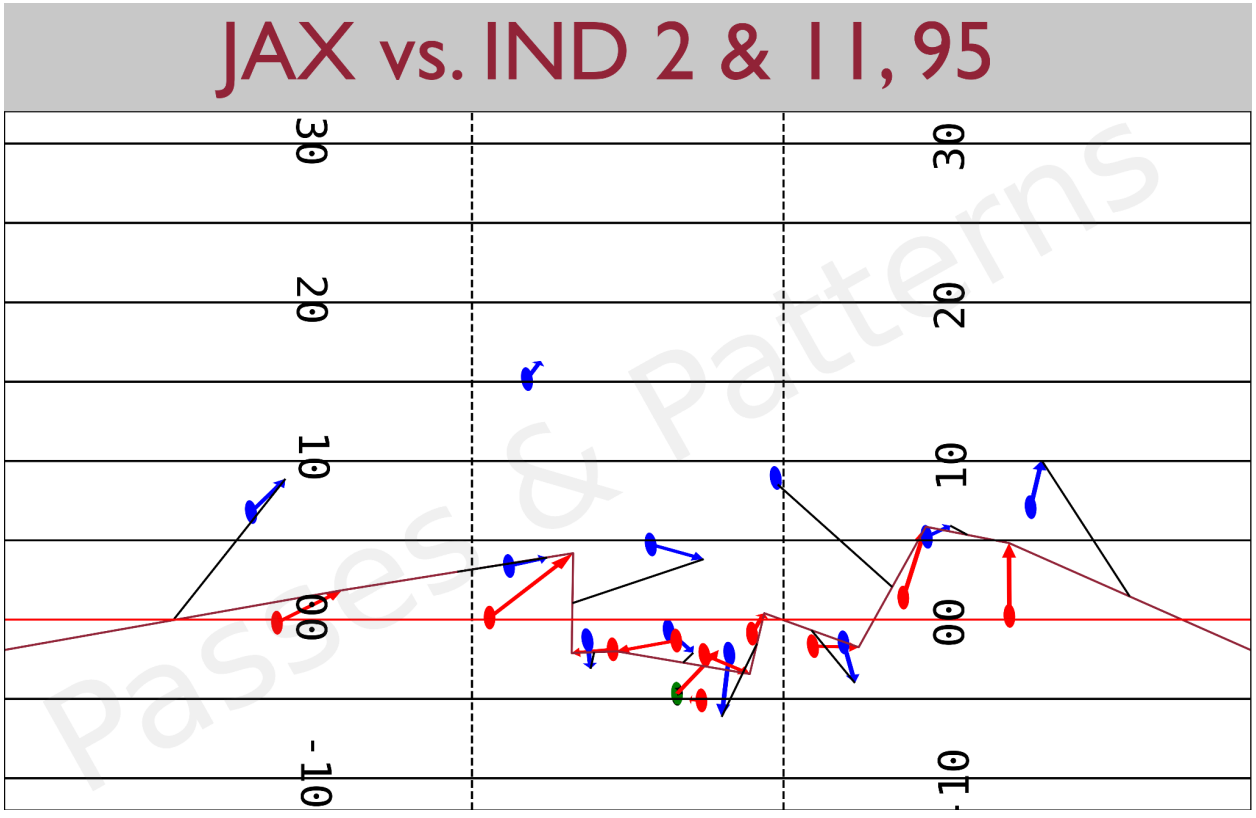
Having defined the gaps we need to find the defenders to fit them. Using the defender's positions we treat the assignment of defenders to gaps as an Integer Linear Programming (ILP) problem, seeking to minimize the sum of the distances between defenders and their gaps. To see what this looks like in practice, take the below visualisation, taken from a randomly-selected play from the dataset. A series of line segments from player to player demarcate the gaps in the blockers, and lines connect each defender to a gap.



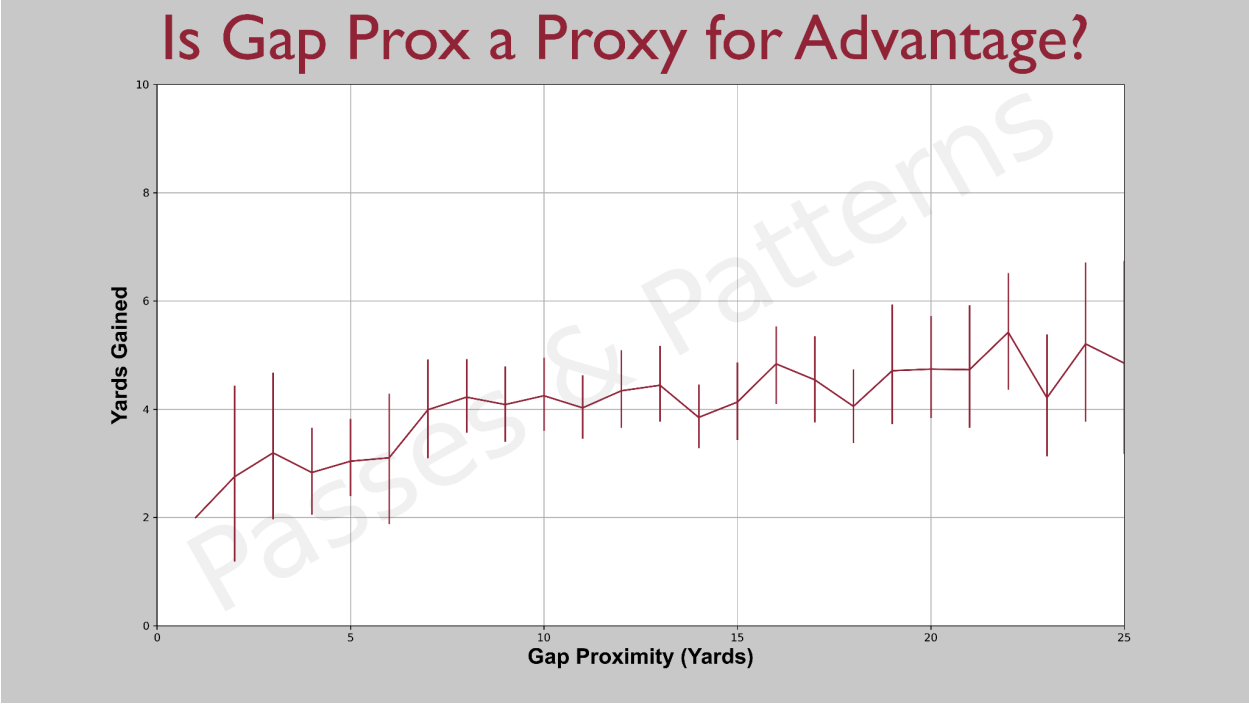
With gaps defined and defenders assigned to them we need to assess the defense's ability to effectively fill those gaps as compared to the ballcarrier's ability to attack those gaps. Our initial method was to identify the gap with the greatest difference between the defender's distance to the gap and the rusher's distance to the gap, as a proxy for the ballcarrier's ability to get through the gap before the defender can fit into it. We hope to see that plays where the rusher has a headstart on the defender to a gap will have an increase in yards per play. Ideally we can find a proxy that serves a continuous measure, an improvement on the current binary evaluation of advantage, which does run into some ambiguous situations sometimes. This would also allow us to automate the process of identifying advantageous plays, which is presently an onerous manual task.



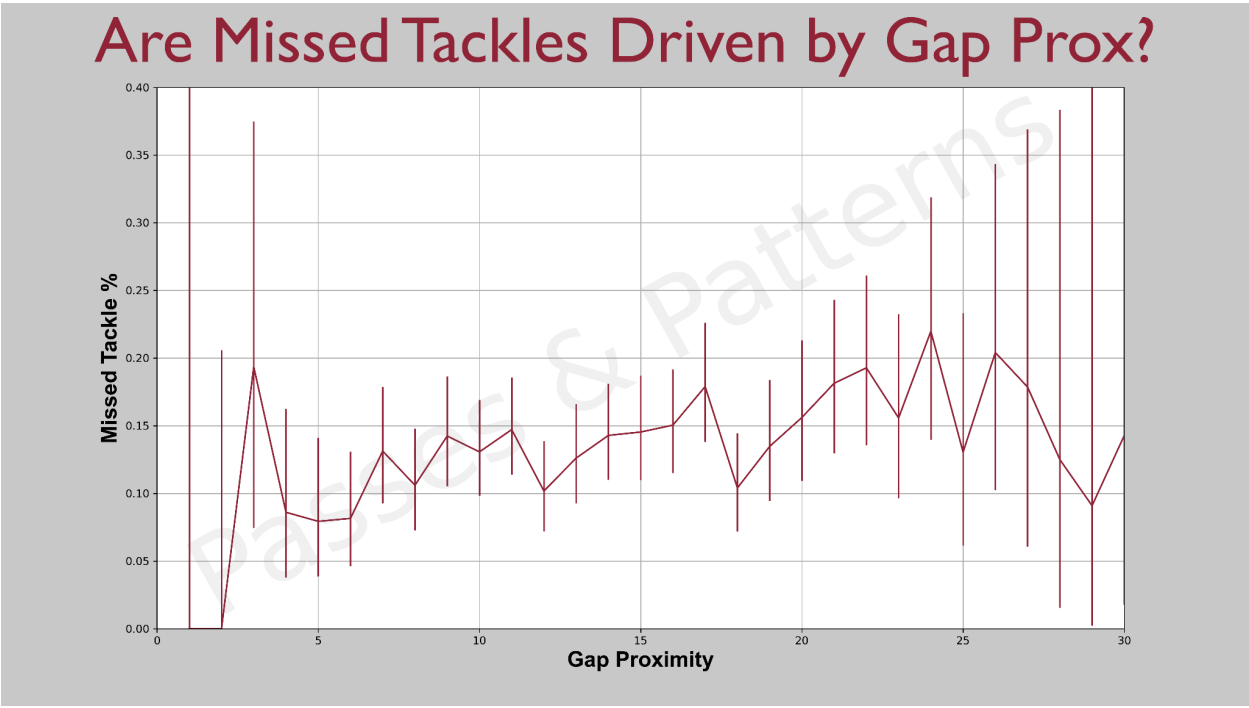
Unfortunately, this method doesn't work. The average yards per play increases negligibly as relative gap proximity increases. We move on to a second method to take the motion of players into account. If, instead of using their actual position, we include their velocity vector we could have a better sense of where gaps are going to be, rather than where they are at a given instant.



Once more, we look at the same proxy, rusher distance from each gap (as measured from the end of the rusher's velocity vector) compared to each gap defender's distance, as illustrated in the figure below.



A partial success, including the velocity vector allowed us to differentiate between plays. We can see consistent improvement as the gap proximity delta increases, from 2.5 yards to 5 yards. Moreover, as pertains to the original thesis of this work, missed tackles are much more prevalent as gap proximity delta increases.



5. Conclusion

Our second method of approximating advantage showed progress and was a weak proxy for advantage state. The most valuable next step is to find better ways of measuring advantage state, which requires some technical developments. The first is to find a better proxy for advantage than the current simple approach of comparing the distance from the gap of the rusher or the defender and taking the simple difference.

The current approach of determining the most efficient gaps is a Travelling Salesman problem, and thus NP-hard. This makes it difficult to scale to frame-by-frame data. The gap assignment problem is an ILP problem, and is also NP-hard. The current approach also relies on creating gaps for each blocker, and treats those gaps as equally accessible when some may not be. A next step would be to automate a method for cancelling out players. Wide receivers and defensive backs who are clearly separate from the run scheme can be cancelled out 1:1 without affecting the advantage state, and this would greatly simplify the two NP-hard problems.

Another problem lies in the gap assignments. The current ILP approach simply optimizes for total distance to gaps, but on the individual player level this can sometimes result in implausible gap assignments, such as where the assigned gap is occluded. A restriction preventing a defender from being assigned a gap such that the path to that gap crosses another gap.

Additional work needs to be done for plays where the QB creates an additional gap, which is a non-obvious set of plays. It is insufficient to merely assume that QB runs and option plays create an additional gap, as many do not do so, and there is a degree of domain expertise required to parse this distinction.

Finally, a maximalist approach would be to have a qualified individual tag a large number of plays as being advantageous or not, and then to develop a supervised learning model to attempt to automate this task. The extent to which this would generate useful results is uncertain, the authors have their doubts about efforts to predict rushing gains in prior Big Data Bowl competitions. However, seeing as the number of coaches with experience in Advantage Theory numbers in the single digits, this would present its own challenges. Advantage Theory has been shown to hold for both runs and passes.

As to the utility of this work, it is simple. There are 3 yards lying on the ground every play, and nobody is picking them up. The value of understanding advantage and employing it regularly is greater than any number of missed tackles.

6. Appendix

We appreciate the assistance provided by Charlie Taggart of the Hamilton Tiger-Cats and Keagan Hall of the McMaster Marauders

Code and additional data at:
https://github.com/christophermclement/NFL-BDB-VI