Understanding Expertise in Elite Competitive eSports: A Comparison of Approaches to Scalable Dimensionality Reduction

Noah Phillips (philln2@rpi.edu)
Department of Cognitive Science, Rensselaer Polytechnic Institute
Troy, NY 12180 USA

Chris R. Sims (simsc3@rpi.edu)
Department of Cognitive Science, Rensselaer Polytechnic Institute
Troy, NY 12180 USA

Abstract

Various methods of dimensionality reduction have been used to apply a quantitative approach to the study of complex skill acquisition. This work builds upon past approaches, offering a comparative analysis of principal component analysis, logistic regression, and linear discriminant analysis to quantify expertise in the domain of competitive video gaming, or "eSports." We present a novel, robust dataset of expert and non-expert gameplay data from professional and amateur players of the *Super Smash Bros. Melee* competitive fighting game. We assess each quantitative model via the metrics of providing accurate expertise classification, predictive utility, and a pragmatic window into the features of complex skill performance that hold the most weight in overall performance outcomes, thereby also providing insights for direction of future training. We posit that linear discriminant analysis provides the best performance for all relevant metrics. The nuances are discussed here, and suggestions for the field are offered for future study of other complex skill domains.

Keywords: dimensionality reduction; principal component analysis; logistic regression; linear discriminant; eSports, complex skill acquisition

Introduction

Understanding the nature of expertise has been a central theme within cognitive science since the founding of the field. In recent years, the increasing availability of large datasets has enabled new approaches for addressing this question (González-Brenes, 2015; Gray, 2017; Gray & Banerjee, 2021; Huang et al., 2017). However, this wealth of data has also brought new challenges. "Big data" in this context often consists of the measurement of hundreds (or thousands) of variables related to human performance, but in practice many measured variables may be irrelevant to the research question. Further, any single variable may only be weakly related to expertise. Consequently, the challenge is not finding a needle in a haystack, but rather making sense of the combined interaction of thousands of needles.

Under these circumstances, a commonly used approach is dimensionality reduction. Here the goal is to discover a small set of latent factors that can help explain expertise in an interpretable fashion, for example, by identifying the dimensions along which experts in a domain differ from non-experts. In the current paper, we explore two different approaches for applying dimensionality reduction within the domain of competitive eSports.

"eSports" refers to the subdomain of video gaming in which players actively seek to maximize their skill level and compete against one another, often in professional tournaments with large sums of money at stake. While video games themselves have been found to causally interact with cognition (Bediou et al., 2023; Green & Bavelier, 2003), it is likely that the realm of eSports further distills these cognitive effects, as its explicit focus on competitive performance and training optimization control for variables such as players' motivation and level of cognitive engagement (Phillips, 2023). In addition, modern eSports automatically collect vast amounts of telemetry data on player inputs and gameplay behavior, offering researchers the full advantages of large, naturalistic datasets collected from individuals at all points along the skill spectrum (Campbell et al., 2018; Pedraza-Ramirez et al., 2020; Reitman et al., 2020). Taken together, these traits make eSports an ideal domain to apply advanced quantitative methods such as dimensionality reduction.

Our research builds upon prior work implementing a similar approach to study another video game: *Tetris*. The work of Lindstedt and Gray (2019) as well as Gray and Banerjee (2021) used *Tetris* data as a medium to implement an exploratory factor analysis on low-level gameplay features, in search of principal components representing underlying skills of expert performance. They found success with this methodology, identifying factors such as "planning efficiency," "pile management," and "zoid control" (Gray & Banerjee, 2021). These factors were then input as predictor variables for a logistic regression model, used to classify players into buckets of expertise level, which in turn highlighted the relative importance of each factor at various points along the skill spectrum. While this approach showed promise for *Tetris*, we point out that, while *Tetris* more than qualifies as a complex task domain with a nearly limitless skill ceiling, *Tetris* lacks a depth of stylistic complexity when compared to other competitive eSports. In other words, games with more room for differing playstyles, alternative strategies, various character types, etc., have more dimensions of variance that do not directly measure performance. We hypothesize that as task complexity increases past the bounds of *Tetris*, techniques such as principal component analysis may lose efficacy as the primary sources of variance are no longer expertise-related. We test this hypothesis by applying the methodology to a

In L. K. Samuelson, S. L. Frank, M. Toneva, A. Mackey, & E. Hazeltine (Eds.), *Proceedings of the 46th Annual Conference of the Cognitive Science Society.*

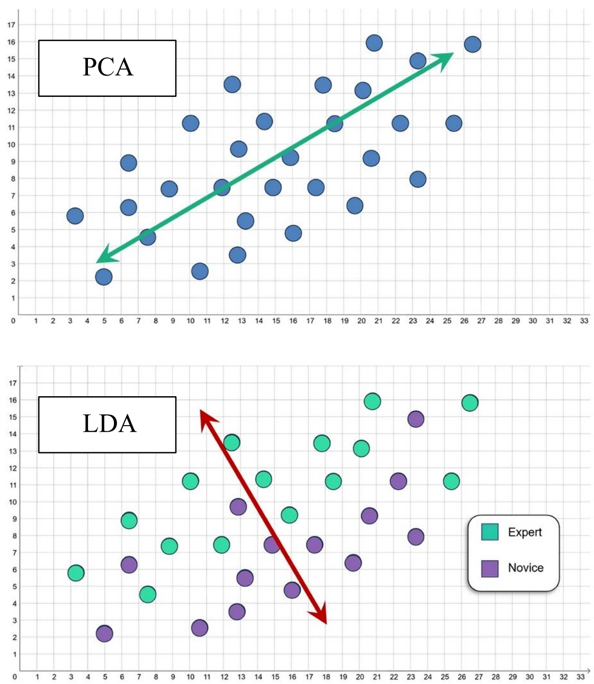

Figure 1: A visual representation of the key difference between principal component analysis (PCA) versus linear discriminant analysis (LDA). In PCA, the first dimension represents the highest vector of *overall* variance in the data, which may or may not be related between groups differences, such as level of expertise. LDA, however, identifies the dimension of variance in the data that maximizes discrimination between multiple known groups.

novel dataset from another competitive eSport, *Super Smash Bros. Melee.* We then present an alternative methodology utilizing linear discriminant analysis to improve upon prior approaches, creating a scalable framework for expertise-based big data analysis as well as providing practical insights for the pursuit of real-world skill acquisition.

Methods

The dataset, R code, and all other supplemental materials used in this work can be accessed via our Open Science Framework repository, linked here: https://osf.io/qvfhx.

Our Novel Dataset

Super Smash Bros. Melee is a popular competitive eSport from the fighting game genre of video games. Players face off against each other, and the goal of the game is to deal enough damage to one's opponent to send them flying off the stage and into the abyss below. Players utilize a series of different attacks, intricate movement options, and a wide range of tactics and strategies to outplay their opponent.

Although the game was originally released in 2001 for the Nintendo GameCube home video game console, the grassroots competitive community has engineered a robust system by which the game can be played online via personal computers, facilitating competitive play between continents (*Slippi* website: slippi.gg). As an auxiliary result, these matches automatically record a wealth of data of each players' inputs during the games they play. Replay files for all matches played online are automatically saved, and public repositories exist where players of any skill level can upload their replays to be harvested by data scientists. In addition, the same input tracking software has been utilized at many of the largest in-person tournaments and world championships, for which public data repositories exist as well. By sampling data from only the final matches of any given tournament, we can effectively isolate gameplay data specifically from the best, highest-level experts in the world.

We have curated data from these two repositories to create a combined dataset consisting of two distinct groups, one of amateur, non-expert players, and the other of top-level professional, expert players. Additional confounding variables, such as which in-game character our players are using, have been standardized to uniformity to ensure that between-groups comparisons are as ecologically valid as possible. Data have been parsed into our dataset such that each observation in our dataset represents data from one player in one particular game. A particular player may show up more than once within the dataset, if more than one of their games existed in the repositories. In total, our parsed dataset contains a matched sample of 364 amateur-level non-experts and 364 top-level professional experts.

The data consists of 69 different quantitative measures of gameplay performance, such the number and types of inputs each player is making, as well as summary statistics measures such as how often the player is "getting the first hit," thereby winning a neutral exchange and creating an opening to capitalize on, how much damage the player is dealing to their opponent per opening they create, and how many of such openings they require to defeat their opponent. A sample of features are described in Table 1.

Principal Component Analysis

The methodology for our principal component analysis (PCA) was based largely on the work of Gray and Banerjee (2021), which itself expanded and improved upon previous work by Lindstedt and Gray (2019). The overall logic is as follows: use gameplay data to acquire a list of low-level features which provide numerical representations of various metrics of skilled performance. Next, apply exploratory factor analysis to these features in an attempt to identify latent factors (Costello & Osborne, 2005) consisting of multiple corelated features within the data. This is accomplished using PCA as a method of dimensionality reduction. Once these factors have been identified, their loading vectors can be examined to see which features weigh most strongly upon each factor, and we may assess whether these factors point to intelligible, underlying skills of gameplay.

For our purposes, the quantitative measures present within our *Smash Bros. Melee* dataset serve the role of a feature list. One potential problem is that some level of redundancy exists within our dataset, as some constructs are effectively measured multiple times, such as players' inputs being

Table 1. Sample gameplay features and definitions for *Super Smash Bros. Melee* replay data (Slippi)

Gameplay Feature Name	Description	Expected Indication for Expertise Level
digitalIPM_ratio	The number of button presses the player performed per minute. Doesn't include analog stick inputs.	Expert players should tend to exhibit higher inputs per minute scores.
neutralWinRatio_ratio	The number of openings to inflict damage the player created, proportional to their opponent.	This depends primarily on how closely matched in skill the two players are.
damagePerOpening_ratio	The amount of damage, on average, the player was able to inflict on their opponent per opening.	Experts should score higher, hitting harder and exhibiting greater efficiency.
Lcancel_fail	The number of times the player failed to execute an "L cancel" – an advanced technique which speeds up the recovery time of certain attacks.	Experts should score lower, as this feature represents an objective, unforced error.
usmash	The number of times the player performed an "Up smash" – a high-risk, high-reward attack option.	Experts should score lower, as they exhibit a less risky playstyle.

measured both as a sum total per match as well as a ratio of inputs per minute. As high collinearity of dependent variables has been shown to affect our planned analyses (Henson, 2002), the 69 variables were assessed for redundancy, and variables with a collinearity threshold above 50% were removed, resulting in a final list of 40 features of gameplay.

These 40 features served as the input for our principal component analysis, and all future analyses. In following with Gray and Banerjee (2021), we applied a varimax rotation to our PCA to aid in factor interpretability. We next created a scree plot to assess how many of our rotated principal components should be retained for the next steps of our analysis. Considering recommendations regarding eigen values of features greater than 1, explaining a sufficient level of overall variance in the data, and trying to visually identify "the elbow" in the scree plot, we chose to retain the first 12 factors from our PCA. These 12 factors account for 53.8% of the overall variance in the dataset, with the first five factors themselves accounting for 6.4%, 6.0%, 5.7%, 4.9%, and 4.8% of overall variance, respectively.

Finally, the feature loadings for each factor were analyzed to look for patterns pointing towards underlying skills being captured by the PCA. Unfortunately, unlike in Gray and Banerjee's deep dive into *Tetris* gameplay data, we were not able to interpret cohesive meanings of each factor based on the features which primarily comprised them. It is critically important to note that this failing is certainly not due to a lack of knowledge of the inner workings of our data, or of the task domain our data come from. The first author of this paper is a former professional player of the competitive *Super Smash Bros. Melee* eSport, and therefore possesses a deep and nuanced understanding of this domain. Additionally, the authors have consulted with international top-level players and game analysts who belong to this competitive community, who have all similarly been unable to interpret the meaning of the factors resulting from principal component analysis. We hypothesize that – compared to *Tetri*s – a relatively open-ended, free-flowing fighting game such as *Super Smash Bros.* simply achieves a level of complexity for which a large amount of variance in the data cannot be attributed directly to skill level. We will elaborate upon this hypothesis later, but for now we continue with the next step of the planned analysis: logistic regression.

Logistic Regression

Regression on Component Factors The next section of our analysis involves implementing logistic regression as a means to identify gameplay factors that may be predictive of gameplay skill level. Our first step is to feed the principal component scores acquired from our exploratory factor analysis in as input variables to a logistic regression model in order to perform binary classification among two distinct groups: experts and non-experts. After running the initial regression model, we refined the model using a stepwise, bidirectional model selection which optimizes for the lowest possible Akaike information criterion (AIC). The resulting model retained 8 principal component predictors, and achieved an AIC of 713.45, with a McFadden's pseudo R^2 = .31. The next step was to examine which factors within the regression model did the best job of predicting expertise status for any given individual. Comparing the estimate values and z-values for each retained component, we found that the third principal component was the best single predictor of skill level. In line with our proposed analysis, we then reexamined the features that comprise component three to see if there is an intuitive rationale as to why those features would be particularly effective in predicting skill.

If such an intuitive explanation did exist, then perhaps we could conclude that, although the first two principal components are measuring dimensions of variance in the data other than expertise level, this third component is measuring the underlying skills which do directly relate to performance outcomes. Unfortunately, the features comprising this factor do not align in any apparently meaningful way, and thus we again cannot in good faith present an objective narrative of what skills this or any other factor are measuring.

Regression on Raw Data Variables As an alternative approach we also fit a separate regression model on our feature list of raw gameplay data. In this case, we fed our 40 gameplay features into the model, after which we once again applied a bidirectional stepwise model selection. The resulting model retained 25 gameplay feature predictors, and achieved an AIC of 616.57, with a McFadden's pseudo R^2 = .44. In examining this raw data model, we finally see that the gameplay features which are the strongest predictors of expert group membership align very well with our domain knowledge of this competitive eSport. This regression model shows that players who exhibit features of greater technical skill, such producing higher inputs per minute (see Figure 2 and Table 1), and exhibiting better metrics of efficiency, such as dealing more damage to their opponent per opening chance, are more likely to be experts. Conversely, higher metrics reflecting unforced errors, such as failing to execute moves correctly or failing to choose the correct moves in a given situation, are strong predictors of being a non-expert.

It seems that, generally, applying the combination of factor analysis and logistic regression approach which bore fruit for the simpler task paradigm of *Tetris* did not excel at either of its goals when applied to the competitive eSport of *Super Smash Bros. Melee*. It failed to give us meaningful insights regarding the underlying skills relevant to gameplay performance measured by each PCA factor, and even for the factor measured most predictive of expertise group, the features comprising this factor could not in good conscience be interpreted with any objectively meaningful patterns. What we offer next is an alternative to the above methodology, which we believe not only effectively accomplishes the prior stated goals, but also, critically, can scale with increasing complexity of a given task domain.

Linear Discriminant Analysis

Descriptive Discriminant Analysis Our first goal of linear discriminant analysis is to serve a descriptive role (Fisher, 1936) in identifying the gameplay features that maximize the discrimination between our expert groups. We start by centering and scaling our data, as the natural scale of our gameplay metrics can range from less than one to sometimes over a thousand. We then passed our modified dataset to the lda() function in R's MASS package, with the prior probabilities of membership to each group being equal (50/50), due to our known expert groups being of equal sample size. This operation assigns a numeric weight, or coefficient, to each gameplay feature, with a greater coefficient magnitude indicating that said feature is particularly useful in discrimination between expert groups, and therefore that this feature may be of particular importance to competitive performance more broadly. In examining the list of coefficients, we are relieved to see that the weights align nearly perfectly with our prior domain expertise of competitive *Super Smash Bros. Melee*. Precisely the features we expected to be correlated with gameplay skill level seem to be the exact variables that are most effective in discriminating between our expert groups. This serves as a promising step in validating LDA as a useful tool to accomplish our first analytical goal.

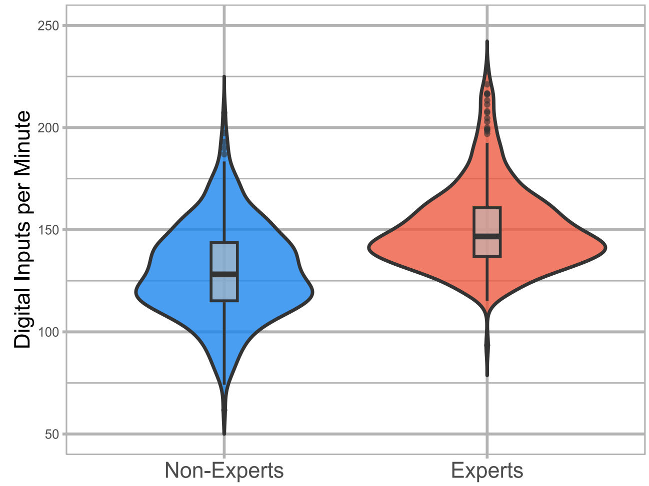

Figure 2: Violin plots for digital inputs per minute, one of the gameplay features which most strongly predicts overall skill level. The plots include scattered data points as well as a nested box and whisker plot for each expertise group.

Predictive Discriminant Analysis Our second goal is slightly more ambitious, with possibly farther-reaching implications. LDA also has application as a predictive tool for future group-based predictions. The aim here is to use our model to create classification rules which can predict group membership based on novel gameplay of unknown skill level. By first training the model using data from known experts and non-experts, we can produce *linear classification functions* to make these future predictions. With these functions, we can then calculate probability values for each group based on new data.

Our methodology for predictive discriminant analysis is based largely on the procedure laid out by Boedeker and Kearns (2019). Running our data through their syntax produces a new dataset complete with class predictions, posterior probabilities, and typicality probabilities for each observation. These new statistics give us several insights, such as highlighting possible interactions between typicality and expertise. Based on the suggestion of Huberty and Wisenbaker (1992), we considered any observation with a typicality score less than .1 to be an outlier, meaning that this player was not the "typical" representation of its predicted class. Across our 728 observations, we found 41 outliers among the experts and 89 outliers among the non-experts. That there were more than double the number of atypical cases in our non-expert group than in our expert group may suggest that there is a more stable archetype of highly skilled gameplay, compared to a more chaotic and variable swath of non-experts.

Another key benefit of this methodological approach is allowing for the identification of so-called *fence riders* (Huberty & Olejnik, 2006). Fence riders are cases where the probabilities of being classified as either group are very similar, such as being classified as an expert with 51% certainty. Fence riders provide LDA with yet another unique advantage: to be able to identify new groups within the data.

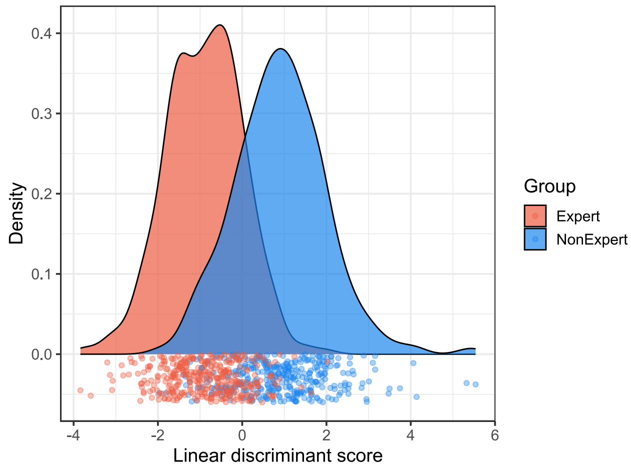

Figure 3: Distribution of linear discriminant scores for the expert and non-expert groups. The plot markers show individual players (their position jittered to increase visibility). Overall, LDA achieves 80% classification accuracy (repeated *k*-fold validation).

Especially when dealing with continuous spectra such as skill level, a large number of fence rider cases may indicate that an additional intermediate skill level group should be added to the analysis. In our case, the number of fence riders was low (4 non-experts, 6 experts), suggesting that our data collection was successful in sampling from two distinct skill groups, and further verifying the existence of fundamental differences between experts and non-experts of this domain.

Classification Accuracy Perhaps the most important output from our model is the previously mentioned linear classification functions which we can use to assess the classification accuracy of our model, as well as to predict expertise level for novel data of unknown skill level. In following the recommendations of the field, we evaluate the accuracy of our model using *k*-fold cross validation, with k = 10 and repetitions = 20 (Boedeker & Kearns, 2019; Rodríguez et al., 2010). Our model achieved a hit rate of .796, and with consideration of our prior probabilities, achieved a Huberty's *I* index of .593, indicating a large effect size when compared to Huberty's conservative threshold of .35 (Huberty & Lowman, 2000) (see Figure 3 for a visual depiction of linear discriminant scores). This combination of 79.6% classification accuracy and a notably large effect size – in addition to LDA's strongest predictive features aligning nearly perfectly with our expectations stemming from robust domain knowledge – shows great promise in this approach being the correct tool for analyzing complex skill data.

Prediction for New Cases The next capability of our linear classification functions pertains to classification of novel data not present within the original dataset. This manifests in two distinct and valuable use cases. The first of which is to simply collect new data and run them through our preexisting functions, to predict their skill level category. Our model easily allows for this, and generates posterior and typicality probabilities for each new data point. We can opt to use these new classifications to update our prior probabilities as well, supporting a complete Bayesian framework.

The second use case has even greater potential for practical application in the real-world. Any competitive player could first input their own data into the model to see which expertise category they are placed into. Next, using the most predictive features identified by the descriptive portion of the LDA, the player can analyze their own personal shortcomings when compared to others in their skill categories. This provides critical, evidence-based insight into which areas of their performance they ought to devote more time to improving. Then, the player can run what-if simulations of their data to see how much improvement would be necessary to move up to the next skill level. It could be that a player labelled as a non-expert is actually quite close to reaching expert territory, and that if only the player could improve a particularly lagging metric by a few points, they could experience a marked leap in performance. This application could prove to be a vital tool for skilled competitors in domains beyond competitive eSports, as well as for the coaches who train them, both as a means of identifying lagging aspects of performance and as a motivational driver to help the individual break through learning plateaus.

General Discussion

This paper presents several different approaches to implementing dimensionality reduction to study the underlying skills present within the complex task domain of competitive eSports. Building on previous methodology used to study players of the video game *Tetris* (Gray & Banerjee, 2021; Lindstedt & Gray, 2019), we first applied principal component analysis to a novel dataset of players of various skill levels who compete in the *Super Smash Bros. Melee* eSport. Contrary to findings in *Tetris*, the clusters of correlated features picked up by our largest principal component factors did not seem to represent specific underlying skill constructs that matched up in any meaningful way with deep, organic domain knowledge. These principal factors were then used as variable inputs into a logistic regression model, which performed binary classification on gameplay data into either expert or non-expert groups. This model achieved mediocre performance (pseudo R^2 = .31), and when examining the strongest single predictor (principal component three), we once again find no explanation for which latent skills could be represented by the correlated features comprising this factor. This most predictive factor only accounted for 5.7% of the variance in the dataset, further supporting the notion that the dimensions of variance picked up on by PCA are not likely to relate primarily to expertise, particularly when studying a complex and open-ended task domain such as a competitive eSport.

The second approach – the one we favor – makes use of a much simpler, more robust, and user-friendly methodology in that of linear discriminant analysis. This approach addresses both of our goals, namely to be able to describe key differences between experts and non-experts, and to create a model that allows us to predict group classifications either for brand new cases, or for preexisting data hypothetically modified to see how specific feature alterations would affect expertise classifications. Our predictive discriminant model achieved a high level of prediction accuracy (80% hit rate), with a large effect size (Huberty's $I = .593$). Additionally, this approach supports a complete Bayesian framework, with the ability to set prior probabilities as an input and to output posterior probabilities and typicality observations for each case. These measures allow researchers to easily identify and examine typicality outliers within each group, as well as scan for fence riders that may indicate the existence of distinct intermediate expertise groups which should be considered. Finally, and perhaps most critically, the variable coefficients produced by our LDA paint a picture of expertise that aligns very closely with our own knowledge of the task domain. The gameplay features identified by LDA to be the best predictors of expert group discrimination were the precise features that we would expect to correlate strongly with overall game skill. We believe that one of the most vital steps in applying these quantitative methods to studying skill acquisition is to ensure that one's numerical methods can, to a reasonable degree, resemble insights derived from true experts. That our LDA was uniquely able to align with the knowledge of real experts shows great promise in its utility.

Why did the PCA approach that seemed appropriate for the game of *Tetris* go so wrong for us? We believe the answer lies predominantly with task complexity. *Tetris* is a game that requires a great deal of skill by all means, but its structure is markedly different than a one-on-one competitive fighting game such as *Super Smash Bros. Melee*. We posit that, unlike *Tetris*, other eSports simply contain a greater degree of possible variance which is not intrinsically tied to expertise level, or necessarily even to performance outcomes at all. We draw an analogy to the broad domain of *music*. What are the primary sources of variance in such a complex and varied category? One may first think of different genres, or styles, or instruments. One could build a lengthy list before even thinking to consider "the skill of the performer" as a key factor in the variance of music. In this sense, we believe that the more complex a domain is, the more dimensions comprise its final form or output beyond merely expertise. Following this line of reasoning, the employment of principal component analysis is not likely able to provide valuable insights for the specific study of expertise in task domains of sufficient complexity, as the primary dimensions of variance picked up by the PCA won't be connected to expertise itself (see Figure 1 for a visual representation of this point).

We surmise that the game of *Tetris* may be near the limit of task complexity for which the previous methodology may bear fruit. This, again, is not to say that *Tetris* is a simple, uncomplicated domain. We merely argue that a much greater degree of the overall variance in Tetris performance is directly related to skill level and performance outcomes. When compared to other eSports, or professional musicians, or traditional sports, there is much less room in *Tetris* play to allow for additional markers of complexity such as *style*. It speaks less so to the complexity of *Tetris* itself, and much more so to the nearly unimaginable limits of depth that exist in other tasks. If we are to develop a robust and generalizable framework for understanding skill acquisition, it must be capable of scaling with task domain of arbitrarily increasing complexity, which we argue can be accomplished with LDA.

We find that the LDA approach is not only an improvement upon the previous methodology used to study *Tetris*, but also the best tool when compared against other known options. One such alternative to LDA would be to apply logistic regression directly to the raw feature values. The two approaches have similarities, but also important differences (Efron, 1975). Compared to LDA, logistic regression places fewer assumptions on the data (such as normality or equal covariance). The most significant difference is that logistic regression estimates the probability that a case is a member of a binary class; in contrast, LDA seeks to identify a continuous latent dimension along which cases differ. The latter is more appropriate to the current setting, where we seek to model expertise as falling along some continuum rather than as a binary construct. LDA allows us not only to predict expertise, but critically, to also understand the factors that contribute to *gradations* in expertise along this dimension.

Similarly, one might consider the use of a neural network such as a perceptron for expertise classification. Although some work has found multilayer perceptrons to slightly outperform LDA under specific circumstances (Pardo et al., 2006), results have often been comparable between the two approaches (Altman et al., 1994; Bertels et al., 1999). More critical than even accuracy, however, is the issue of interpretability. As opposed to a black-box neural network, the feature scalings, linear discriminant scores, and various probability metrics given by LDA provide a much clearer picture of how the analysis is effectively quantifying expertise, giving useful insight for practical intervention. We believe this to be a key advantage of LDA: the ability to strike an optimal balance between accuracy *and* interpretability.

Conclusion

Compared against alternatives, linear discriminant analysis provides an effective tool for dimensionality reduction to be used for the study of skill acquisition and expertise in complex task domains. LDA achieved superior results in identifying underlying features of skilled performance congruent with the intuitions of genuine experts with organic domain knowledge. LDA offers additional unique insights such as the ability to easily identify typicality outliers and emergent intermediate groups while operating within a robust Bayesian framework. While this case study utilized competitive eSports as the task domain of choice, the tools offered by LDA can be applied to data of any domain, unbounded by task complexity. High-level performers, as well as those who train and coach them, can harness these quantitative insights to enhance training paradigms, leading to real-world improvements in performance, plateau breakthroughs, and achievements in new heights of expertise.

Acknowledgments

This work was supported in part by grant N00014-22-1-2109 from the Office of Naval Research. Special thanks to Jas "Fizzi" Laferriere for pioneering the *Slippi* replay system, which made this large-scale data collection possible, the Slippi development and Melee Stats discord servers for analysis and interpretation advice, and to Dr. Emily Klein for lending her invaluable brilliance in writing, and all else.

References

Altman, E. I., Marco, G., & Varetto, F. (1994). Corporate distress diagnosis: Comparisons using linear discriminant analysis and neural networks (the Italian experience). *Journal of Banking and Finance*, *18*(3), 505–529. https://doi.org/10.1016/0378-4266(94)90007-8

Bediou, B., Rodgers, M. A., Tipton, E., Mayer, R. E., Green, C. S., & Bavelier, D. (2023). Effects of action video game play on cognitive skills: A meta-analysis. *Technology, Mind, and Behavior*, *4*(1). https://doi.org/10.1037/tmb0000102.supp

Bertels, K., Jacques, J. M., Neuberg, L., & Gatot, L. (1999). Qualitative company performance evaluation: linear discriminant analysis and neural network models. *European Journal of Operational Research*, *115*(3), 608–615. https://doi.org/10.1016/S0377-2217(98)00161-1

Boedeker, P., & Kearns, N. T. (2019). Linear Discriminant Analysis for Prediction of Group Membership: A User-Friendly Primer. *Advances in Methods and Practices in Psychological Science*, *2*(3), 250–263. https://doi.org/10.1177/2515245919849378

Campbell, M. J., Toth, A. J., Moran, A. P., Kowal, M., & Exton, C. (2018). eSports: A new window on neurocognitive expertise? In *Progress in Brain Research* (1st ed., Vol. 240). Elsevier B.V. https://doi.org/10.1016/bs.pbr.2018.09.006

Costello, A. B., & Osborne, J. W. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research and Evaluation*, *10*(7).

Efron, B. (1975). The efficiency of logistic regression compared to normal discriminant analysis. *Journal of the American Statistical Association*, *70*(352), 892–898. https://doi.org/10.1080/01621459.1975.10480319

Fisher, R. A. (1936). The Use of Multiple Measurements in Taxonomic Problems. *Annals of Eugenics*, *7*(2), 179–188. https://doi.org/10.1111/j.1469-1809.1936.tb02137.x

González-Brenes, J. P. (2015). Modeling skill acquisition over time with sequence and topic modeling. *Journal of Machine Learning Research*, *38*, 296–305.

Gray, W. D. (2017). Game-XP: Action Games as Experimental Paradigms for Cognitive Science. *Topics in Cognitive Science*, *9*(2), 289–307. https://doi.org/10.1111/tops.12260

Gray, W. D., & Banerjee, S. (2021). Constructing Expertise: Surmounting Performance Plateaus by Tasks, by Tools, and by Techniques. *Topics in Cognitive Science*, *13*(4), 610–665. https://doi.org/10.1111/tops.12575

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*(6939), 534–537. https://doi.org/10.1038/nature01647

Henson, R. K. (2002). The logic and interpretation of structure coefficients in multivariate general linear model analyses. *Annual Meeting of the American Educational Research Association*, *2*, 1–96.

Huang, J., Yan, E., Cheung, G., Nagappan, N., & Zimmermann, T. (2017). Master Maker: Understanding Gaming Skill Through Practice and Habit From Gameplay Behavior. *Topics in Cognitive Science*, *9*(2), 437–466. https://doi.org/10.1111/tops.12251

Huberty, C. J., & Lowman, L. L. (2000). Group overlap as a basis for effect size. *Educational and Psychological Measurement*, *60*(4), 543–563. https://doi.org/10.1177/0013164400604004

Huberty, C. J., & Olejnik, S. (2006). *Applied MANOVA and discriminant analysis*. John Wiley & Sons.

Huberty, C. J., & Wisenbaker, J. M. (1992). Variable Importance in Multivariate Group Comparisons. *Journal of Educational Statistics*, *17*(1), 75–91. https://doi.org/10.3102/10769986017001075

Khondoker, M., Dobson, R., Skirrow, C., Simmons, A., & Stahl, D. (2016). A comparison of machine learning methods for classification using simulation with multiple real data examples from mental health studies. *Statistical Methods in Medical Research*, *25*(5), 1804–1823. https://doi.org/10.1177/0962280213502437

Lachenbruch, P. A. (1975). Zero-mean difference discrimination and the absolute linear discriminant function. *Biometrika*, *62*(2), 397–401. https://doi.org/10.1093/biomet/62.2.397

Lindstedt, J. K., & Gray, W. D. (2019). Distinguishing experts from novices by the Mind's Hand and Mind's Eye. *Cognitive Psychology*, *109*, 1–25. https://doi.org/10.1016/J.COGPSYCH.2018.11.003

Pardo, M., Sisk, B. C., Sberveglieri, G., & Lewis, N. S. (2006). Comparison of Fisher's linear discriminant to multilayer perceptron networks in the classification of vapors using sensor array data. *Sensors and Actuators, B: Chemical*, *115*(2), 647–655. https://doi.org/10.1016/j.snb.2005.10.033

Pedraza-Ramirez, I., Musculus, L., Raab, M., & Laborde, S. (2020). Setting the scientific stage for esports psychology: a systematic review. *International Review of Sport and Exercise Psychology*, *13*(1), 319–352. https://doi.org/10.1080/1750984X.2020.1723122

Phillips, N. (2023). Competitive eSports as a new paradigm for cognitive science: current state and future directions. *Proceedings of the Annual Meeting of the Cognitive Science Society, 45*, *45*, 3537–3544. https://escholarship.org/uc/item/9nt8j13s

Reitman, J. G., Anderson-Coto, M. J., Wu, M., Lee, J. S., & Steinkuehler, C. (2020). Esports Research: A Literature Review. *Games and Culture*, *15*(1), 32–50. https://doi.org/10.1177/1555412019840892

Rodríguez, J. D., Pérez, A., & Lozano, J. A. (2010). Sensitivity Analysis of k-Fold Cross Validation in Prediction Error Estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *32*(3), 569–575. https://doi.org/10.1109/TPAMI.2009.187

Sherry, A. (2006). Discriminant Analysis in Counseling Psychology Research. *The Counseling Psychologist*, *34*(5), 661–683. https://doi.org/10.1177/0011000006287103