Professional Dealers versus the GSR: An Empirical Perspective

ARJUN GROVER, UC Berkeley, USA

ACM Reference Format:
Arjun Grover. 2024. Professional Dealers versus the GSR: An Empirical Perspective. 1, 1 (May 2024), 23 pages.
https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 ABSTRACT

We introduce an original dataset of hand-shuffled cards by professional dealers (n=229 riffles, n=21 boxes/strips, n=30 cuts). We then use various statistical techniques to analyze how one theoretical shuffle versus one empirical shuffle randomizes the deck (both relative to uniform binomial randomization). We find that, despite past literature believing the contrary, not every professional dealer is better than the "average person" Gilbert-Shannon-Reeds model, with the model even outperforming some professionals. This leads us to conclude that shuffling randomization is a skill improved by recent, direct practice. Additionally, we show the importance of boxing/stripping the cards, providing randomization equivalent to many riffle shuffles when applied in sequence. Further discussion about the implications of these results concludes the paper.

2 INTRODUCTION

While mechanical card shuffling machines have become much more prevalent in the last few decades (see [8] for details on the randomness of said machines), hand shuffling is still extremely common worldwide, being used in both home games and casinos. In the Los Angeles home game poker scene alone, there are typically many games running every weekend, each ranging from tens to hundreds of thousands of dollars in play. Nearly all of these games will employ a dealer who will hand shuffle cards, like those recorded in this paper, throughout the session. In the casino scene,

Author's address: Arjun Grover, ajgrover26@berkeley.edu, UC Berkeley, Berkeley, California, USA.

Manuscript submitted to ACM

hand shuffling most often comes up during larger tournaments or whenever a table's auto-shuffler needs maintenance. For consistency, casinos will have all tables of a tournament be hand shuffled if they don't have enough tables with mechanical shuffling to be used across every table in play, a very common occurrence at smaller poker rooms or larger tournaments (e.g., the World Series of Poker). The latter maintenance reason comes up much more often than one might expect, as most mechanical shuffling machines have stringent rules on who may repair said machine. This leads to extended periods of time where all hands dealt at that table will be manually shuffled. In these hand-shuffled high-stakes home games and hand-shuffled casino games, a standard shuffle sequence is nearly always applied:

- Riffle: A standard riffle shuffle is conducted by splitting the deck in half and interlacing these two piles together. So-called "clumping" occurs when multiple consecutive cards from one pile are not split up by cards from the other pile, creating a packet from the previous deck order in the new deck order.
- Riffle a second time.
- Box/Strip: A box cut, also known as a strip cut, is conducted by taking the top quarter of the deck and placing it on the table, then taking the next quarter and placing it on top of the first quarter, etc. until the deck is completed. Sometimes fewer or greater than 4 packets are stacked on top of one another (e.g., the top third of the deck is placed on the table, then the next third is placed on that, etc.), with the main idea being to reverse the order of large packets in the deck.
- Riffle again.
- Cut: Take off roughly half of the deck from the top and place it on the table. Then, take the other half and place it on top of the first half.

This sequence of riffle, riffle, box, riffle, cut is empirically studied in the Results.

In contrast, lower stakes home games (typically less than $10,000 in play) almost never complete the full sequence, oftentimes replacing the box with a cut or omitting it altogether. These games will often have players rotate as the dealer or have a staple player who deals every hand. In either scenario, the dealer will typically have a low-medium amount of skill (i.e., at or below the skill level of a professional dealer with six months of experience). These sequences at this skill level are empirically studied in the Results.

Beyond Poker, shuffle randomization has a great deal of importance in many other card games as well. One of the first empirical confirmations of the existence of clumping (Berger 1973) was in the study of Contract Bridge, revealing that the hands with an even suit distribution occurred at a significantly higher frequency than would be expected with a uniformly distributed deck of cards.

Lastly, beyond card games, shuffling has a great deal of mathematical significance. Many of the world's finest mathematicians (e.g., Borel, Shannon, and Poincare) have published research on the exploration of Card Shuffling. Diaconis 2002 [6] and Diaconis, Fulman 2023 [7] explore some of the most significant mathematical and statistical conclusions that have been gained through the study of card shuffling, including developments in Monte Carlo experiments, Zeta values, Descent Theory, Symmetric Function Theory, Hyperplane Walks, Representation Theory, Brown's Semi-Group Walks, and more. While this paper will focus on the empirical side of shuffling rather than the mathematical side, the conclusions reached in this paper, such as the importance of the Box/Cut, have great potential for further mathematical research.

3 PRIOR WORKS

As outlined in [8], randomization after a shuffle and repeated shuffles of a deck of cards is a problem that has challenged probabilists for over a century. Starting with Hadamard 1906 [14] and Poincare 1912 [18], their theoretical analysis showed that repeated riffle shuffles should approach uniform randomization at an exponential rate (randomization as a function of number of shuffles creates a geometric sequence), but weren't able to provide empirical or quantitative conclusions that were applicable to practical problems. Borel and Cheron [4] then studied shuffling in the context of Bridge in 1955, furthering the discourse behind riffle shuffling significantly, including exploring the effects of the packets generated by poor randomization. Williams and Jordan [22] followed this up with a card trick exploiting the existence of packets after seeming "randomization," and Berger 1973 [3] explores how Bridge tournaments moving online fundamentally changed the game (as players were previously exploiting clumping occurring through in-person shuffling which no longer existed in online play). A conclusive mathematical model was then crafted by Gilbert and Shannon in 1955 in Bell Labs [12] and later, independently, by Reeds in 1981 [19] to create the Gilbert-Shannon-Reeds, or GSR, riffle shuffle model. This is the central model upon which we will compare the collected empirical data. The GSR model proceeds as follows:

- Cut the deck into two piles (1 and 2) according to the binomial distribution:
  $\mathbb{P}(\text{Pile 1 has } k \text{ cards}) = \binom{n}{k}/2^n$ for $k \in [0, n]$
  (Note that this is the same mathematical model used for a cut!).
- Sequentially drop a card from each pile with probability equal to the proportion of cards in it:
  $\mathbb{P}(\text{Drop a card from Pile 1}) = \frac{A}{A+B}$
  where A is the number of cards in Pile 1 and B is the number of cards in Pile 2.
- Repeat dropping until both piles are empty, thus recombining the piles into one shuffled deck.

This model was described and results expanded upon in Diaconis's 1988 book *Group representations in probability and statistics* [5], providing an excellent guide to the algebraic underpinnings of probability at work in shuffling and beyond. Notably in the book was section 4D, an empirical "confirmation" of the GSR model, which has been cited many times since (including by many subsequent famous papers, such as [2]). In this section, Diaconis explores, with Reed, the distribution of a "regular" person and magician's (Reeds and Diaconis, respectively) packet sizes and drops over 100 riffle shuffles. While Diaconis had packet sizes concentrated significantly closer to 1 than Reeds, both had packet sizes generally higher than the $\frac{1}{2^i}$ expected model. Unfortunately, though, no thorough statistical difference testing or more granular data collection was examined (due to computational limitations of the time), nor was the dataset recorded for further analysis. This paper will explore the empirical foundation in much greater depth, providing statistical evidence and clarity to the notion that "[the GSR] is not a good model for the way that Vegas dealers shuffle" [7] by studying actual dealers from a local to Vegas level. The GSR shuffle was created to approximate how regular people riffle shuffle, while Vegas dealers are considerably "better" at shuffling, how much better are they really?

Beyond [5], The only other published empirical card-shuffling work was by Silverman in 2019 [20], where $M = 19$ shuffles were conducted upon $N = 12$ "sets" of "worn-in" cards (no masking as per our methodology). In this paper, Silverman spends a great deal of analysis comparing a rudimentary mechanical shuffling machine to hand-shuffling, with little comparison to the existing theoretical models. It is not noted whether an amateur(s) or professional dealer(s) conducted the shuffles, and the dataset of the digitized shuffles has not been released. Statistical rigor is also not discussed.

4 DATA

Three professional dealers with a wide range of experience were hired. Dealer 1 has dealt professionally for a couple of years across the world, but has since assumed the role of pit boss at a local casino and has not dealt professionally since (for the last couple of years). Dealer 2 has been dealing professionally (at a local casino and freelancing for parties) for about 6 months and actively deals professionally. Dealer 3 has extensive experience, having dealt professionally for many years, including dealing multiple times for the World Series of Poker, and continues to actively deal professionally. Each of these three dealers were hired for one hour and conducted the same experiment.

- The dealer would conduct one riffle shuffle and then fan all of the cards out face up (as depicted in Figure 1). They would repeat this process for 40 minutes.
- The dealer would conduct one "box"/"strip" shuffle and then fan all of the cards out face up. They would repeat this process for 10 minutes.
- The dealer would conduct one cut and then fan all of the cards out face up. They would repeat this process for 10 minutes.

This resulted in over 100 riffle shuffles, 20 strips, and 20 cuts for each of the three dealers.

4.1 Preprocessing and Digitization

After data collection, we were left with over three hours of raw footage. To clean this for digitization, the videos were scrubbed through and a screenshot was taken when the cards were fanned out and all of the cards were visible. While this was relatively easy for Dealers 2 and 3, Dealer 1 oftentimes had multiple cards overlapping, which were only visible by scrubbing between frames (not all visible at once), which created large issues in digitization.

In deciding how to digitize the screenshots from raw images to cards, the first seemingly viable approach was to implement a State-of-the-art YOLOv8 Convolutional Neural Network object detector and classifier (built on TensorFlow) on 52 classes, one for each card. This would enable bounding box recognition and classification of each card, after which a post-processor could output the classifications sorted by x-min (left x-value bounding box position), thus outputting the card order from left to right. Roboflow was used to upload and classify/bounding box training data before exporting as a PASCAL Visual Object Classes (a standardized image data set file type).

Fig. 1. An example screenshot after the dealer completed a shuffle.

Data augmentation in the aspects of Brightness, Exposure, Shear, and Rotation were all applied to expand the dataset by 3x to roughly 100 training points for each class (reducing all images to 640x640 for consistency and to manage compute). Unfortunately, despite training on the largest YOLOv8 model, the classifier really struggled with test data, correctly bounding close to 0% of the data. Believing that this was due to a lack of sample size, further data was added from already existing bounding boxed and classified card models, albeit with different decks. This expanded the dataset to nearly 200 training points per class (roughly 10k images), but yet again, the classifier struggled.

With this failure, we now concluded that the classifier was struggling due to the unique shear exhibited in the collected data which was not modeled enough by our augmentations and low sample-size representations. As a result, we revised our approach to reduce the YOLOv8 model only to classify a single class: any card corner. Then, after the YOLO model successfully bounds each card corner, we would have OCR classify which card each specific corner is in reference to. Combining these OCR classifications with YOLO locations (again sorted by x-min, the left x-value bounding box position), we would be able to output the card order from left to right. After a great deal of augmentations and fine-tuning, we were able to get the YOLO model to bound over 95% of card corners correctly (an example is depicted in Figure 2). Unfortunately, though, OCR really struggled with classification due to the immense shear presented by these boxed corners (despite limiting the dictionary of the OCR model, etc.). Due to time limitations and deadlines approaching, the vast majority of the data was manually classified, leaving automatic digitization as a place for future work. On this note, only Dealer 1's data has been fully digitized. A portion of Dealer 3's riffles and all of Dealer 2 and Dealer 3's boxes and cuts have yet to be digitized, but a large enough sample has been digitized to maintain statistical rigor (further additions will retain consistent results).

4.2 Reproducibility

A public GitHub Repo[13] contains all data and code necessary to reproduce all results found here and expand on this work for any future work. Within the Data folder, one will find a text file called "data_storage.txt" which contains a link to a Google Drive with the videos and screenshots. "Training.ipynb" is what was run in Google Colab to train the YOLO model, with "Digitizer.ipynb," including how these trained models were applied to the collected data. All analysis of the data is included in "Code.ipynb," and the outputted figures are included in the "Figures" folder. Following these resources, one should be able to reproduce all outputted analyses. Please contact the author with any questions about the reproducibility of the data or analysis.

5 METHODOLOGY

Since the actual card values are irrelevant and each shuffle is independent, it is possible to "mask" the card ordering by resetting the values to 0-51 at any given shuffle stage and then applying the masking function on the shuffled deck (or after any future number of shuffles). This enables $\frac{n(n+1)}{2}$ shuffle sets, each with $n - i + 1$ shuffles respectively, where i denotes the number of shuffles you would like in each set ($i \in [1, n]$). Therefore, whenever we analyze how the deck changed after k riffle shuffles, we take advantage of the masking function to calculate the loss between a naturally ordered deck and a masking function from the starting index deck to the natural numbers applied to the deck k indices after the starting index deck (thus acting as if the deck is originally ordered at the start of each shuffle). When analyzing the standard shuffle sequence, the first two shuffles are represented by a mask, the box/strip is applied by randomly selecting a masking function, the third riffle is represented by the riffle succeeding the original two, and the cut is applied as a

Fig. 2. An example of a test-set image's YOLO classifications.

randomly selected masking function from the conducted cuts (all from the same dealer). Since each shuffle is independent, selecting sequential riffles is identical in result to randomly selecting riffle masks (as done for the box/strip and cut), the former is just employed for ease of implementation.

To compare the empirical loss (represented in green) to the theoretical model (represented in blue),

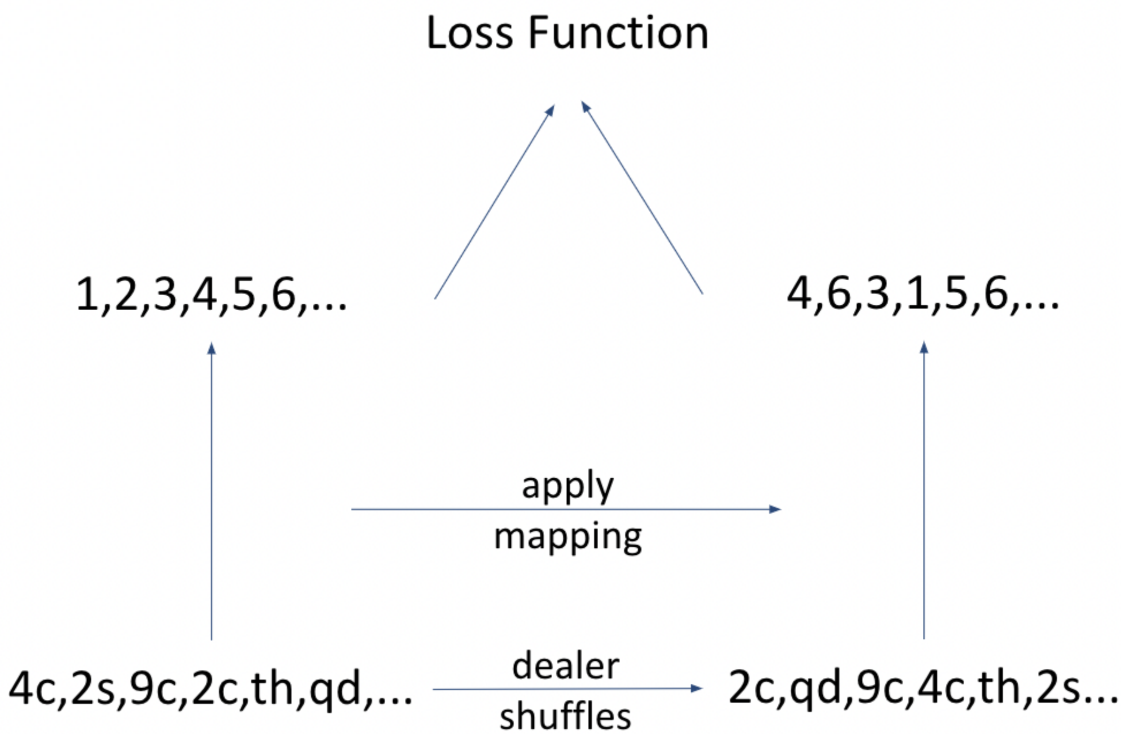


Fig. 3. A depiction of the novel masking technique.

we have implemented the GSR shuffle into Python to represent riffle shuffles. Boxes/Strips are represented by the following model:

- Select a random number k from 1 to 8 by flipping a coin 8 times, i.e., $k = max(1, Binom(8, 0.5))$. These parameters were selected since 4 piles are the most common from real boxes/strips. Each pile will be of size $\lfloor \frac{n}{k} \rfloor$.
- Pull off a pile from the top of the deck of the aforementioned size and put it into a new pile.
- Keep pulling off piles from the top of the old deck and place them on top of the new deck until there are no more cards from the old deck.

This effectively acts in the same manner that a box is conducted. Lastly, the cut is modeled by binomial card selection as in the first step of the GSR. In addition to these models, a uniform random distribution (as in Aldous Diaconis, 1986[1]) is also depicted as a reference point (represented in orange).

First, the most common loss functions in past card randomization literature are computed upon one shuffle for each dealer. Let x be the deck order before the shuffle and y be the deck order after. Then, each respective loss function is calculated as:

- L2 Loss: $\sum_{i=1}^{n}(x_i - y_i)^2$
- Spearmann Rank Correlation: $1 - \frac{6\sum_{i=1}^{n}(R(x_i)-R(y_i))^2}{n(n^2-1)}$
- KL-Div: $\sum_{i=1}^{n} x_i \log(\frac{x_i}{y_i}) - x_i + y_i$
- Kendall's Tau: $\frac{2}{n(n-1)} \sum_{i<j} \operatorname{sgn}(x_i - x_j) \operatorname{sgn}(y_i - y_j)$

After calculating these four respective loss functions on a single shuffle, they are then applied upon sequences of shuffles and a "low-stakes home game" variation. Lastly, an analysis of convergence from only riffling is conducted. To maintain thorough statistical conclusions, each of these loss diagrams has had a respective ANOVA test computed upon it. Specification of the model is as follows:
Begin with an OLS model:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij}$$

where Y_{ij} is the loss function for the j-th observation in the i-th group, μ is the overall mean, τ_i is the effect of the i-th group, and ϵ_{ij} are i.i.d normal errors s.t. $\epsilon_{ij} \sim \mathcal{N}(0, \sigma^2)$. Then, let the ANOVA be defined by the F-Statistic:

$$F = \frac{MS_{\text{Between}}}{MS_{\text{Within}}} = \frac{\frac{SS_{\text{Between}}}{df_{\text{Between}}}}{\frac{SS_{\text{Within}}}{df_{\text{Within}}}}$$

where

$$\text{SS}_{\text{Between}} = \sum_{i=1}^{k} n_i(\bar{Y}_i - \bar{Y})^2, \qquad \text{SS}_{\text{Within}} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$$

$$df_{\text{Between}} = k - 1, \qquad df_{\text{Within}} = N - k$$

For each respective ANOVA test, the title specifies which loss function and which variables the test was conducted on. The ANOVA was selected as an appropriate measure for statistical significance since it allows multiple groups to be compared, its assumptions are satisfied (normality and homogeneity of variances due to the large sample size and independence of observations from the nature of the data), and it provides clear quantitative measures of how strong the evidence against the null hypothesis is. To clarify what randomization entails more visually, heat maps were also created (via seaborn) to represent what movement looks like and how clumping may occur.

6 RESULTS

To first represent the data, Figure 4 is used as a visual representation of how often (the proportion of time) a card originally in position k falls into any given position after Dealer 1 conducts one riffle shuffle. Brighter-colored rows indicate that a card originally in position k falls into that row's position after a shuffle more frequently than it would fall into positions with darker-colored rows. An evenly colored graph would represent that a card initially in position k is equally likely to be in any position after a shuffle. In this figure, $k \in [2, 22, 28, 49]$ are represented to depict how a card initially in the top, middle, and bottom of the deck moves about after a shuffle.

Recall that Dealer 1 was out of practice, having not professionally dealt in a few years. In this figure, it is immediately obvious that Card 2 (a proxy for all cards roughly contained in the top third of the deck), the third card from the top, stays in the exact same position a large majority of the time (emphasizing how poor a single shuffle is). Similarly, card 49, the third from the bottom, stays in the exact same position a majority of the time as well. Although this representation only shows a single card, it represents clumping well as the graph is created across many shuffles aggregated, the independence of which shows how card movement is minimal. Additionally, the probability is very smooth between positions, meaning that multiple cards are often dropped together (if it were evenly split between the two piles, we would see position parity being significantly associated with probability, as seen in later dealers who are better at randomization). It is interesting to note that the bottom card has a more spread-out distribution than the top card, most likely because the dealer runs out of cards early when riffling, dropping a large top packet nearly every time (versus the bottom, where each packet still exists). In addition to these two cards, a heatmap of the two middle cards is also shown, representing how a card on either side of the initial cut ends up after the riffle. Card 22 typically ends up on the bottom of the top cut pile, while Card 28 typically ends up on the top of the bottom cut pile (i.e., the dealer is typically cutting between 22-28, which lines up with the GSR). From these heatmap representations, one would easily conclude that the deck is not well randomized at all, which is statistically analyzed in the following figure.

Looking into Figure 5, each of the loss functions previously mentioned in the Methodology section are applied to all of Dealer 1's single riffle shuffles (with masking). As seen by the ANOVA tests next to each graph, every loss function's empirical distribution and model distribution are

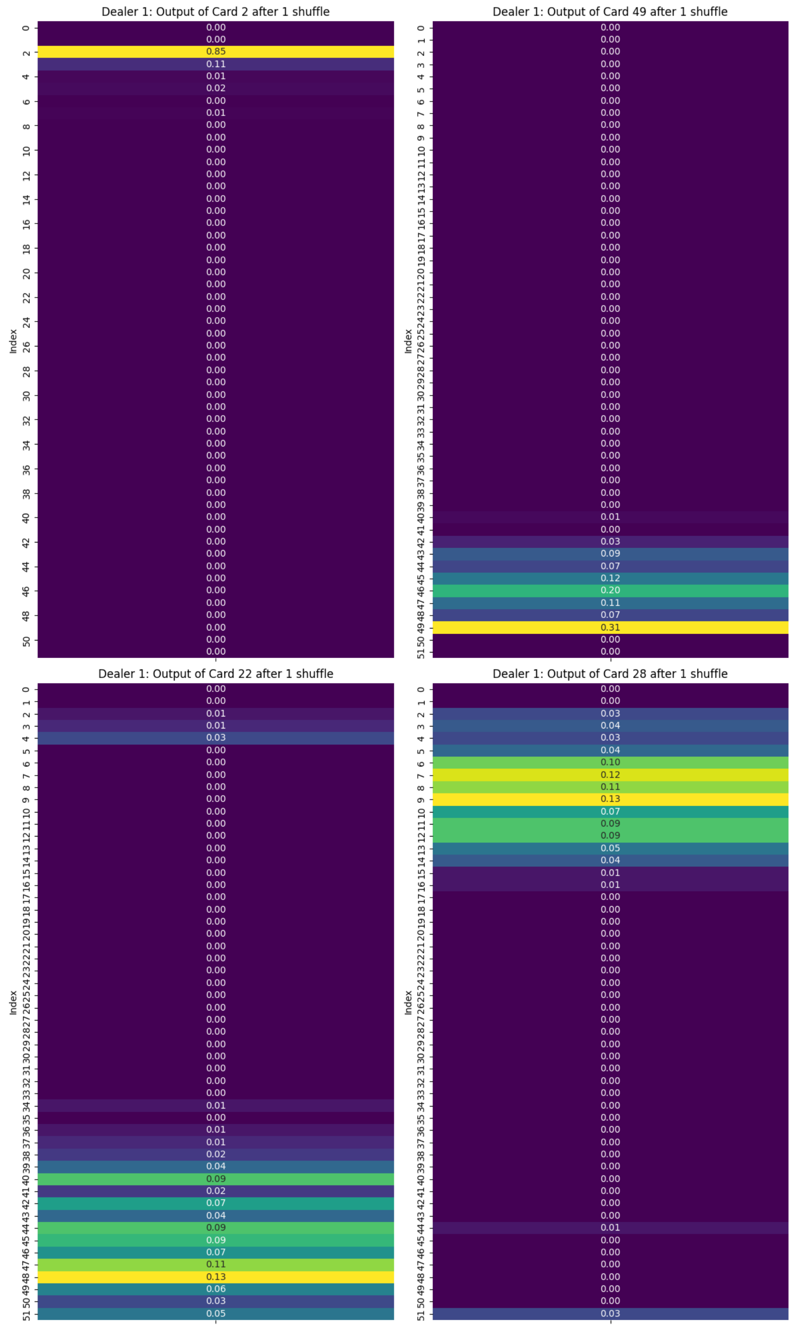
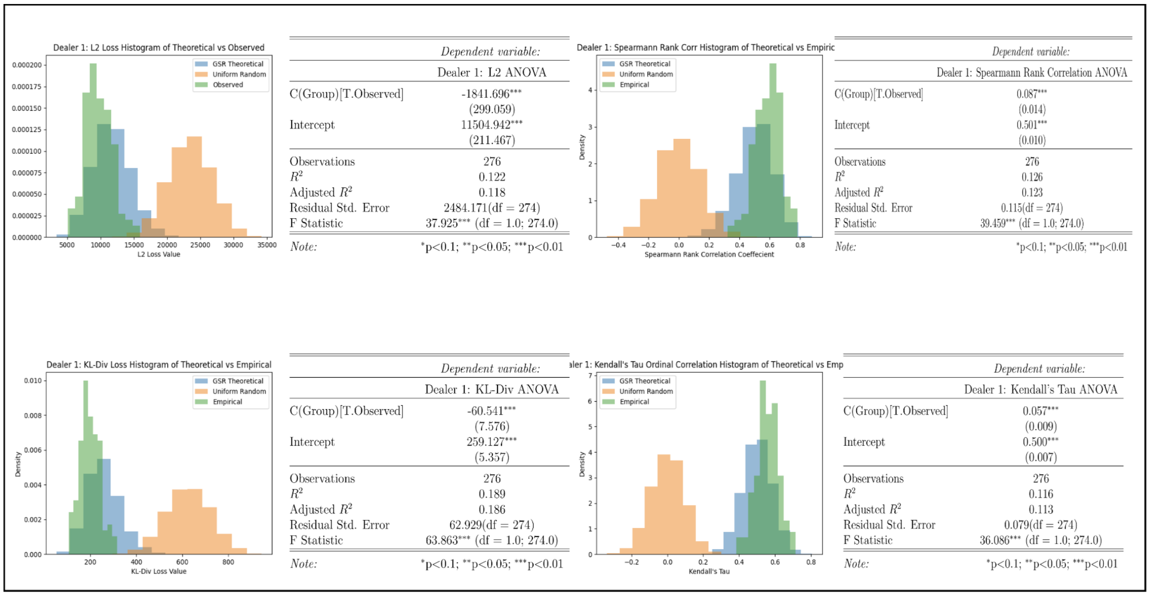


Fig. 4. Heatmap of the distribution of various cards after a single shuffle by Dealer 1



	Dependent variable:
	Dealer 1: L2 ANOVA
C(Group)[T.Observed]	-1841.696***
	(299.059)
Intercept	11504.942***
	(211.467)
Observations	276
R^2	0.122
Adjusted R^2	0.118
Residual Std. Error	2484.171(df = 274)
F Statistic	37.925*** (df = 1.0; 274.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	Dealer 1: Spearmann Rank Correlation ANOVA
C(Group)[T.Observed]	0.087***
	(0.014)
Intercept	0.501***
	(0.010)
Observations	276
R^2	0.126
Adjusted R^2	0.123
Residual Std. Error	0.115(df = 274)
F Statistic	39.459*** (df = 1.0; 274.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	Dealer 1: KL-Div ANOVA
C(Group)[T.Observed]	-60.541***
	(7.576)
Intercept	259.127***
	(5.357)
Observations	276
R^2	0.189
Adjusted R^2	0.186
Residual Std. Error	62.929(df = 274)
F Statistic	63.863*** (df = 1.0; 274.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	Dealer 1: Kendall's Tau ANOVA
C(Group)[T.Observed]	0.057***
	(0.009)
Intercept	0.500***
	(0.007)
Observations	276
R^2	0.116
Adjusted R^2	0.113
Residual Std. Error	0.079(df = 274)
F Statistic	36.086*** (df = 1.0; 274.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

Fig. 5. Distributions and statistical analyses of one shuffle by Dealer 1

significantly different from one another. Surprisingly, this dealer is significantly worse at randomization than the GSR would assume (loss in the opposite direction of the uniform)!

Looking into Dealer 2's heat map, Figure 6 shows a significantly better result than that seen by Dealer 1. Recall that Dealer 2 has been professionally dealing for a relatively short amount of time, but is actively dealing. Cards at the top of the deck are now more evenly distributed, but the pattern of cards on the bottom being more randomly distributed than cards at the top is maintained. Likewise, cards in the middle (this time Cards 22 and 30, respectively) again are polarized by the cut, this dealer having a larger range with splits typically occurring in positions 22-30. Notably, a parity pattern of packet size begins to appear in the middle position shuffles, revealing that Card 22 is typically dropped on the bottom, separated singly by a card from the opposite packet up through position 39, where two card-sized packets begin to occur. This significantly better appearance of randomization is then quantified by Figure 7, where every loss function's empirical distribution and model distribution are not significantly different from one another. This shows that the GSR model accurately represents this dealer's shuffles!

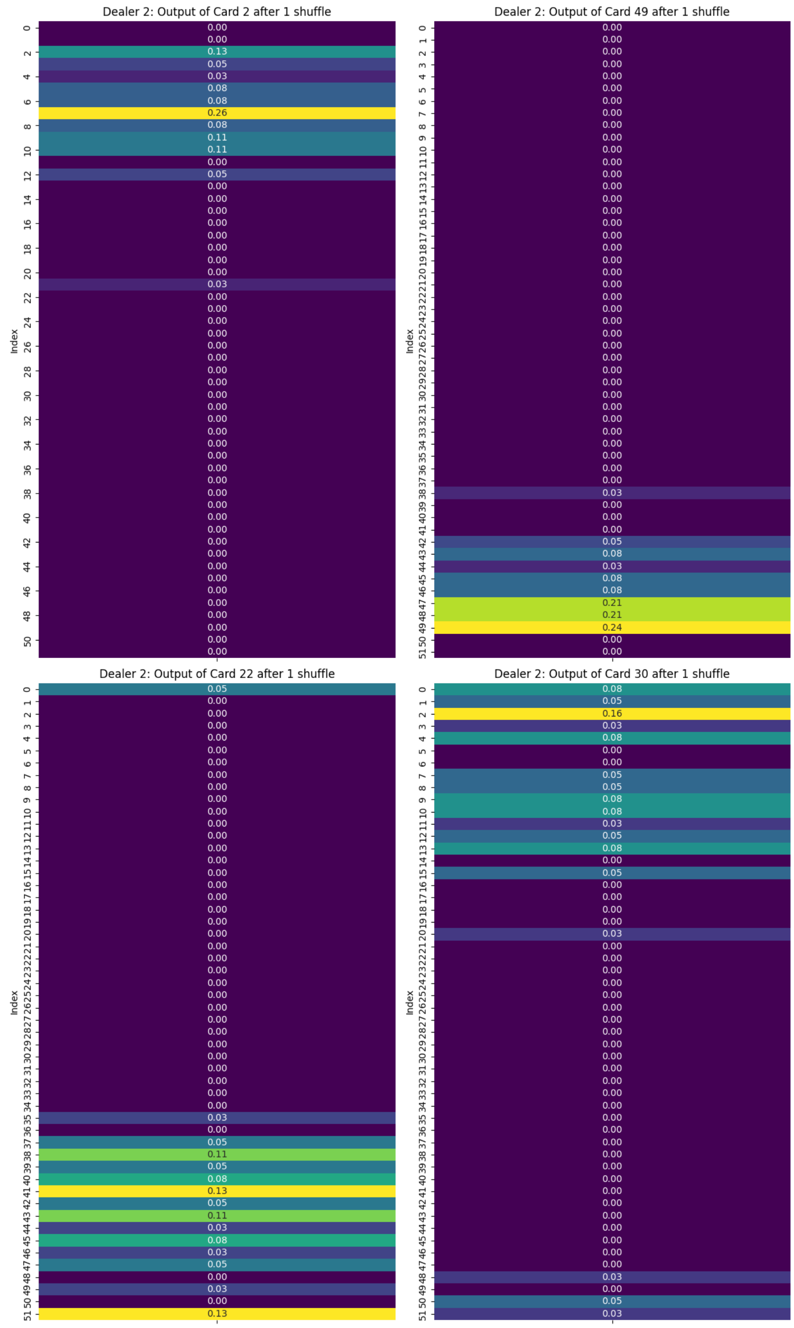


Fig. 6. Heatmap of the distribution of various cards after a single shuffle by Dealer 2

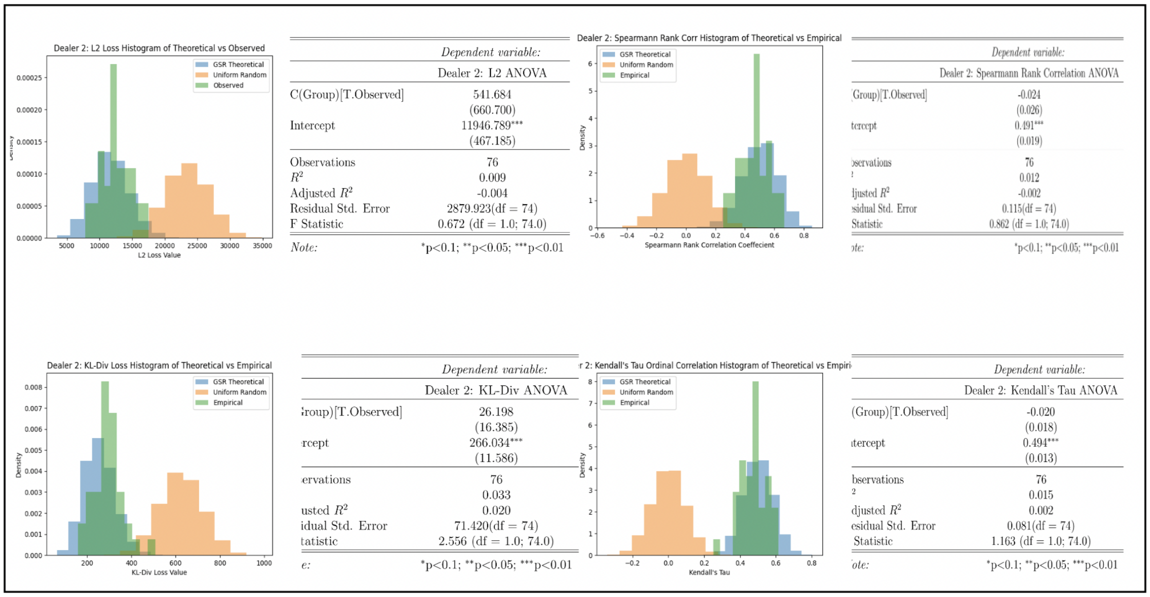


Fig. 7. Distributions and statistical analyses of one shuffle by Dealer 2

Lastly, looking into Dealer 3's heat map, Figure 8 shows an even better result than that seen by Dealer 2, with a greater and more evenly spread range of positions for cards on the top and bottom and cards in the middle more evenly distributed between positions (aka varying packet sizes). Recall that Dealer 3 is an expert, having dealt professionally for many years and actively dealing. This leads to excellent randomization as depicted by Figure 9, where every loss function's empirical distribution and model distribution are significantly different from one another in the positive direction (closer to uniform). This dealer is significantly better than the model would assume (as is consistent with Diaconis's hypothesis[7])!

With this understanding of single shuffles, we may now begin to explore how shuffle sequences work. Looking into the dealer with the worst single shuffle randomization, Dealer 1, we see that after a full shuffle sequence, Figure 10 depicts each card seemingly falling into any given position almost identically, with some hotspots here and there (can be thought of as noise). This heatmap looks significantly better than anything found in the previous single shuffle examples, nearing

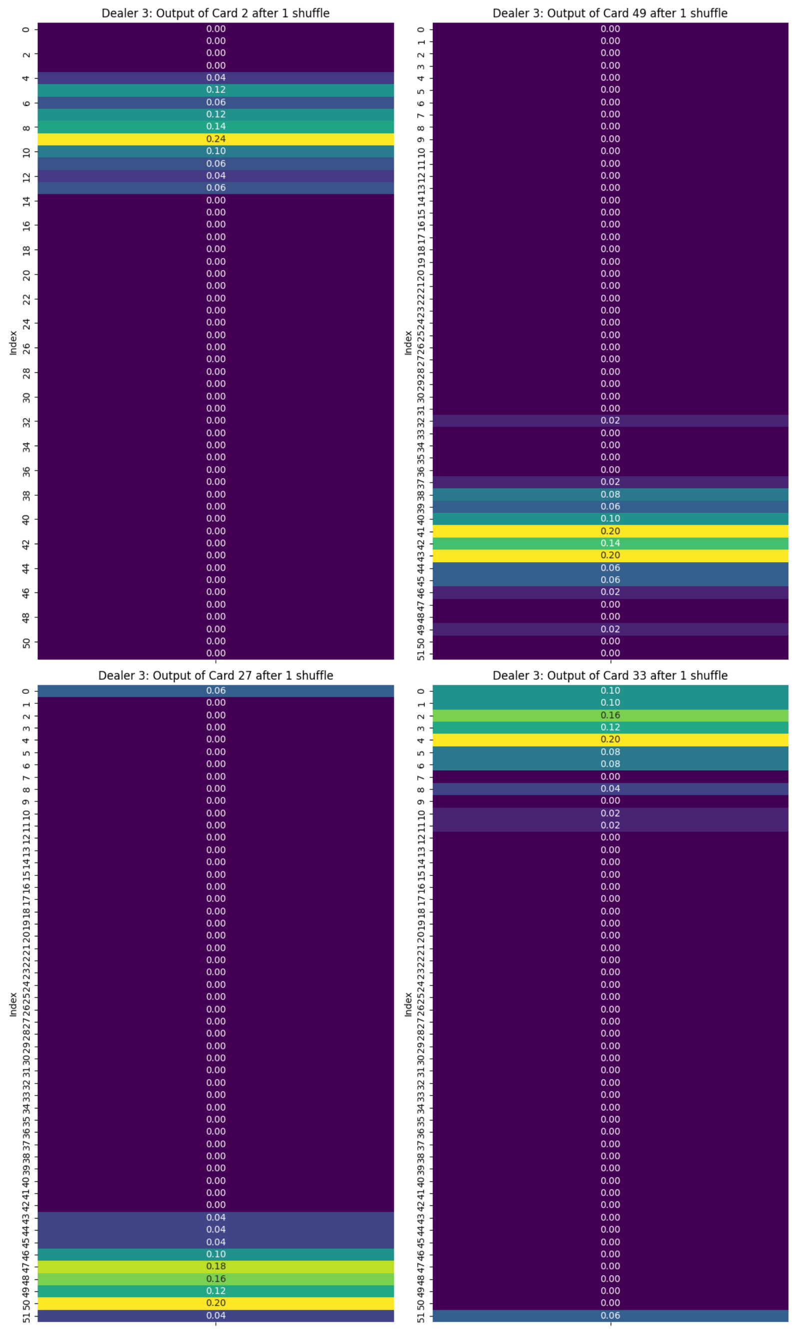


Fig. 8. Heatmap of the distribution of various cards after a single shuffle by Dealer 3

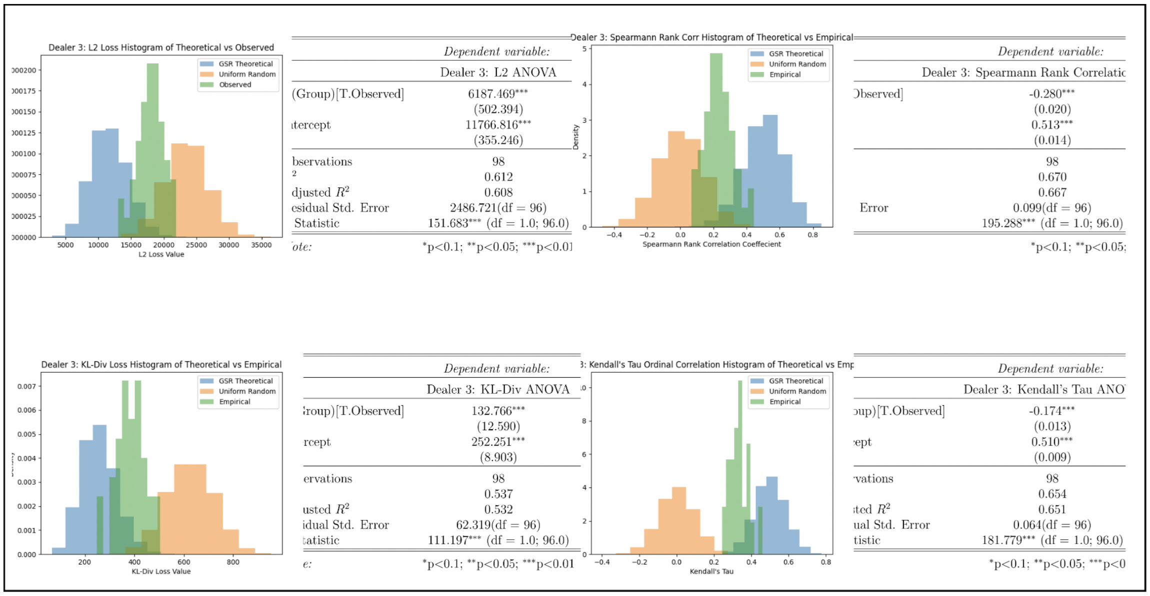


Fig. 9. Distributions and statistical analyses of one shuffle by Dealer 3

something that one might think of as uniformly random. This hypothesis would be proven correct by looking into Figure 11, which shows how the distributions vary across the sequence. L2 loss is used as a proxy for the other losses since they all nearly mirror each other in results. As seen in the diagram, the extremely poor initial shuffle heads quickly toward the uniform distribution (exponentially approaching it as mathematically confirmed by [14] and [18]), which when combined with the box and final cut lends us to almost identically the uniform distribution. Likewise, the GSR mirrors the empirical distribution at this stage, with all 3 achieving no statistical significance under the ANOVA, as we would expect. This diagram lends us to conclude that the standard shuffle sequence successfully randomizes the cards, even if the dealer is seemingly "bad" at randomization (Dealer 2 and 3 are naturally also random by the same metrics). Looking at the convergence across the sequence, though, it is immediately apparent how drastically the box/strip made a difference to the loss, leading to a follow up question: would the distribution still be random if the box/strip was removed?

This question is explored by Figure 12, which examines the same standard shuffle sequence but omits the Box/Strip. It is now clearly apparent how big of a difference the Box/Strip made, with both the model and empirical distributions remaining significantly different from the uniform distribution. Likewise, this omission reveals how little the final cut changes the randomization of the deck (the F-Statistic is barely altered and very significant in both cases), as clumping is

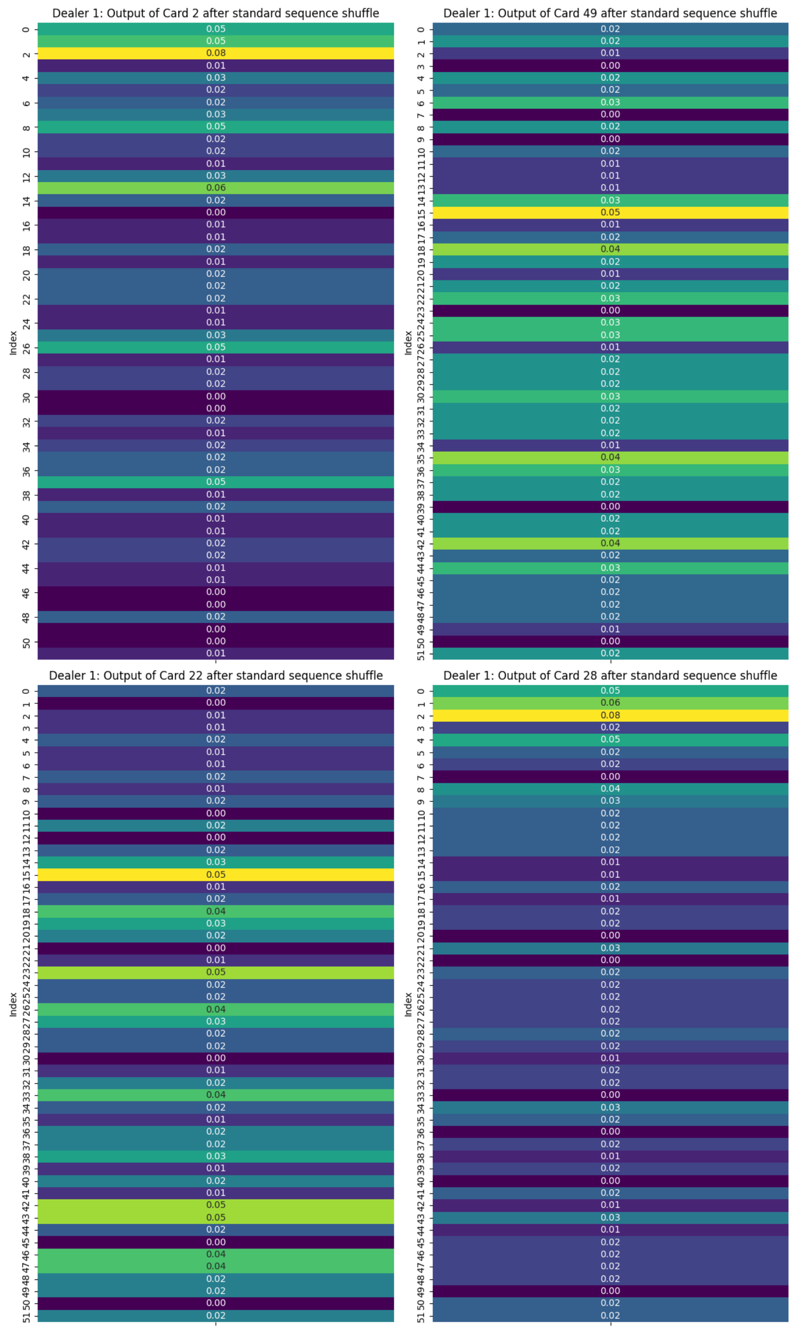
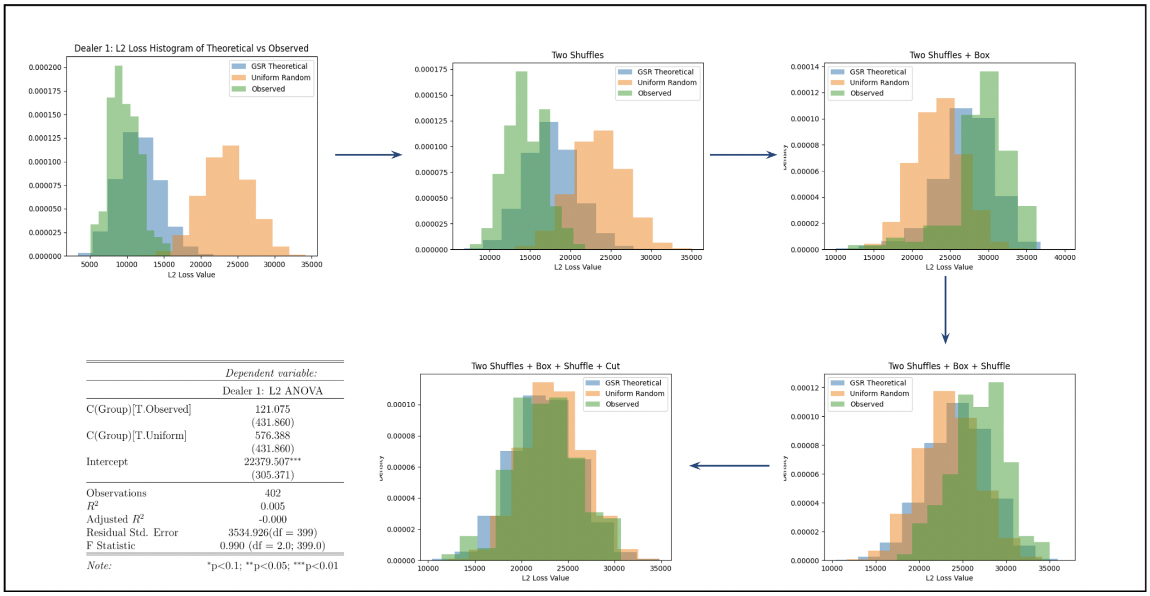
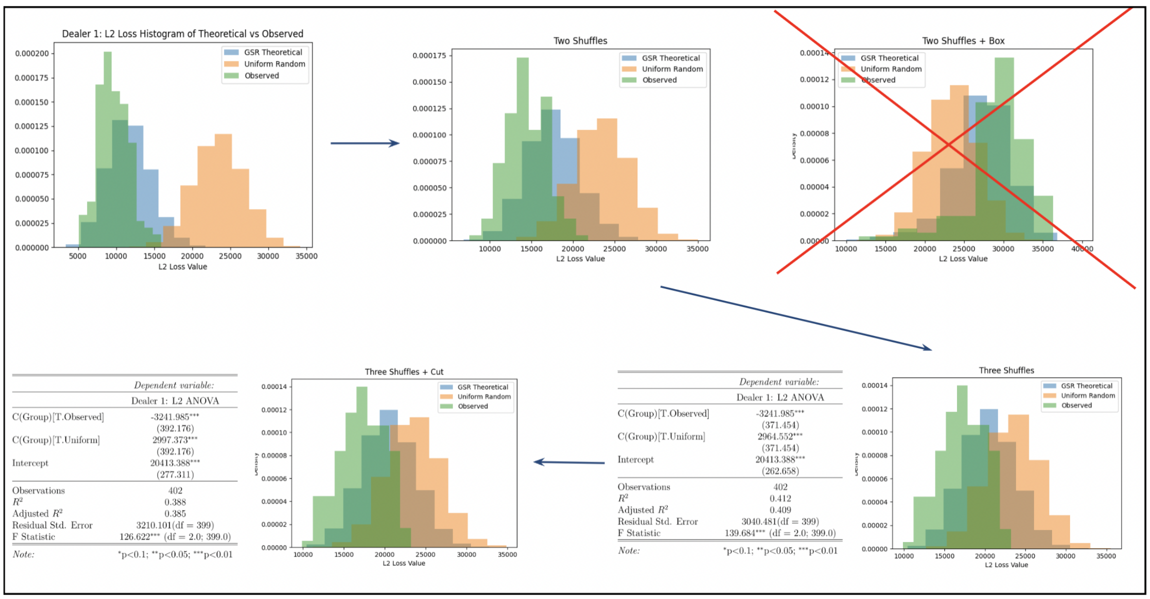


Fig. 10. Heatmap of the distribution of various cards after a standard shuffle sequence by Dealer 1



	Dependent variable:
	Dealer 1: L2 ANOVA
C(Group)[T.Observed]	121.075
	(431.860)
C(Group)[T.Uniform]	576.388
	(431.860)
Intercept	22379.507***
	(305.371)
Observations	402
R^2	0.005
Adjusted R^2	-0.000
Residual Std. Error	3534.926(df = 399)
F Statistic	0.990 (df = 2.0; 399.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

Fig. 11. Distributions and statistical analyses of the standard shuffle sequence by Dealer 1



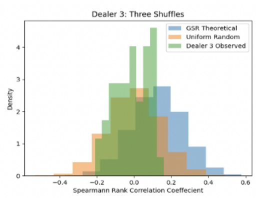
	Dependent variable:
	Dealer 1: L2 ANOVA
C(Group)[T.Observed]	-3241.985***
	(392.176)
C(Group)[T.Uniform]	2997.373***
	(392.176)
Intercept	20413.388***
	(277.311)
Observations	402
R^2	0.388
Adjusted R^2	0.385
Residual Std. Error	3210.101(df = 399)
F Statistic	126.622*** (df = 2.0; 399.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

	Dependent variable:
	Dealer 1: L2 ANOVA
C(Group)[T.Observed]	-3241.985***
	(371.454)
C(Group)[T.Uniform]	2964.552***
	(371.454)
Intercept	20413.388***
	(262.658)
Observations	402
R^2	0.412
Adjusted R^2	0.409
Residual Std. Error	3040.481(df = 399)
F Statistic	139.684*** (df = 2.0; 399.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

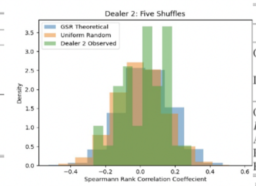
Fig. 12. Distributions and statistical analyses of the standard shuffle sequence by Dealer 1 without the Box/Strip.

unchanged by this maneuver. This result has great implications for low-stakes home games, the conclusions of which are discussed in the following section.

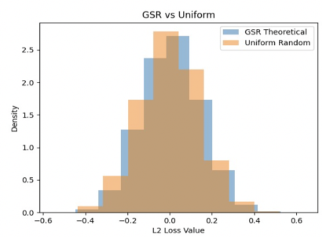
Following up sequence randomization, we then examine at which threshold of simply riffle



	Dependent variable:
	Uniform v D3 Three Shuffles Spearmann ANOVA
C(Group)[T.Uniform]	0.015
	(0.028)
Intercept	-0.005
	(0.020)
Observations	84
R^2	0.004
Adjusted R^2	-0.008
Residual Std. Error	0.127(df = 82)
F Statistic	0.306 (df = 1.0; 82.0)
Note:	*p<0.1; **p<0.05; ***p<0.01



	Dependent variable:
	Uniform v D2 Five Shuffles Spearmann ANOVA
C(Group)[T.Uniform]	0.009
	(0.033)
Intercept	0.017
	(0.023)
Observations	68
R^2	0.001
Adjusted R^2	-0.014
Residual Std. Error	0.136(df = 66)
F Statistic	0.078 (df = 1.0; 66.0)
Note:	*p<0.1; **p<0.05; ***p<0.01



	Dependent variable:
	Uniform v GSR Seven Shuffles Spearmann ANOVA
C(Group)[T.Uniform]	0.016
	(0.034)
Intercept	0.013
	(0.024)
Observations	68
R^2	0.003
Adjusted R^2	-0.012
Residual Std. Error	0.141(df = 66)
F Statistic	0.210 (df = 1.0; 66.0)
Note:	*p<0.1; **p<0.05; ***p<0.01

Fig. 13. Shuffle convergence thresholds for each dealer and the model.

shuffling does non-significance between the uniform distribution occur. Spearmann Rank Correlation is used for the loss function in this case as we are trying to find where clumping disappears. For Dealer 3, Figure 13 depicts that after a mere three shuffles the empirical distribution is insignificantly different from the uniform random distribution. Dealer 2 takes around 5 shuffles to become insignificantly different. As proven in [2], it takes 7 riffle shuffles for the GSR model to become insignificantly different from the uniform distribution. Unfortunately, it took upwards of 20 riffle shuffles for Dealer 1 to become insignificantly different from the uniform distribution with only riffle shuffles, most likely because of running out of cards early (as shown in Figure 4) causing the top cards to move very little. Due to the sample size at this threshold, the figure is not contained.

7 DISCUSSION AND CONCLUSION

Through thorough statistical analysis and the implementation of a novel masking model to augment an original dataset, this paper has shown that shuffling randomization is a skill supported by recent, extensive practice. Dealer 1 revealed that even if one has extensive shuffling experience, not actively dealing for a couple of years can significantly wear down one's skill. Dealer 2 revealed that six months is not enough time to become significantly better at shuffling than the average person GSR model. Dealer 3 revealed that having both extensive shuffling experience and actively keeping up with the practice can make one significantly better at shuffling randomization than the standard model would expect. Heat map distribution figures (Figure 4, Figure 6, Figure 8, Figure 10) reveal how cards in different sections of the deck move after a single riffle shuffle, revealing that the cut point of specific dealers varies and that cards on the top of the deck are oftentimes less well-shuffled than cards near the bottom of the deck. ANOVA tests build thorough statistical evidence behind these conclusions, revealing that not every professional dealer is better than the "average person" GSR model, as previously hypothesized. Instead, the GSR model can vary from being accurate in modeling professional dealers to outperforming some (Figure 5) and underperforming others (Figure 9).

Furthermore, this data illustrates the importance of boxing/stripping the cards, providing randomization equivalent to many riffle shuffles when applied in sequence. The standard shuffle sequence, riffle, riffle, box/strip, riffle, cut, does a great job in randomizing the deck (akin to a uniform randomization), but when the box/strip is removed the empirical shuffle is significantly different from the uniform. In this case, clumping from the previous deck sequence will occur, meaning that cards next to each other in the previous deck state are likely to still be next to each other after the shuffle sequence has occurred. In low-stakes home games, dealers oftentimes do not box/strip the cards, which results in a scenario like this. To exploit this, one must remain aware of the order in which cards are returning to the deck before they are shuffled. From there, one may assume that each card on the top or bottom will remain in a relatively similar position. Cards that begin in the middle will move to the top or bottom, dependent on the dealer's cut position, most likely remaining clumped next to the cards they are originally next to before the shuffle occurs. Using these observations, one may achieve a significant knowledge advantage against other players in any given card game that assumes a uniformly random shuffled card distribution. Further methods of exploitation are further discussed in [21].

Lastly, the simulation figure (Figure 13) confirms the mathematical results of [2], as well as confirming that repeated shuffles move the loss distribution exponentially toward the uniform distribution. The finding that more experienced card mechanics have smaller packet sizes in [5] reveals itself through Spearmann Rank Correlation in our shuffle convergence analysis too. These empirical and simulated (model) distributions confirm the existing literature on the topic.

The main limitation of the data is that only three dealers were studied for this analysis. These three dealers act as proxies for different skill levels (experience and recency) of professional dealers, but a much broader array exists within the real world, and these dealers may not necessarily be fully representative. Furthermore, such exploitations from a lack of randomization can only occur when the cards are poorly manually shuffled, and the use of a shuffle sequence (as described earlier), thorough automatic shuffler, or online play negates these possible exploitations.

Possible branches for future work include successfully creating a machine-learning-based card digitizer (most probably through a better OCR implementation), further exploring this dataset beyond our analysis (e.g., comparing the empirical data to other models and results [10] [11] [15] [16] [17] [21] or cross-sectionally analyzing if dealers got 'tired' and decreased randomization over time or as the cards got more warped), or building the mathematical rigor behind some of the new results presented here (e.g., the importance of the Box/Strip in shuffle sequences). Another possible branch is by applying a Neural Network to the dataset, providing prediction feedback to find the validity of [9].

REFERENCES

[1] David Aldous and Persi Diaconis. 1986. Shuffling Cards and Stopping Times. *The American Mathematical Monthly* 93, 5 (1986), 333–348. https://doi.org/10.2307/2323590 Accessed 21 Feb. 2024.

[2] Dave Bayer and Persi Diaconis. 1992. Trailing the Dovetail Shuffle to Its Lair. *The Annals of Applied Probability* 2, 2 (1992), 294–313. http://www.jstor.org/stable/2959752 Accessed 21 Feb. 2024.

[3] P. Berger. 1973. On the Distribution of Hand Patterns in Bridge: Man-Dealt Versus Computer-Dealt. *Canadian J. Statist.* 1 (1973), 261–266.

[4] E. Borel and A. Chéron. 1955. *Théorie Mathématique du Bridge à la Portée de Tous* (2 ed.). Gauthier-Villars, Paris. MR0070896.

[5] Persi Diaconis. 1988. *Group representations in probability and statistics.* Institute of Mathematical Statistics Lecture Notes—Monograph Series, Vol. 11. Institute of Mathematical Statistics, Hayward, CA. Section 4D.(e), page 82.

[6] Persi Diaconis. 2002. Mathematical developments from the analysis of riffle shuffling. http://statistics.stanford.edu/~ckirby/techreports/GEN/2002/2002-16.pdf. (2002).

[7] P. Diaconis and J. Fulman. 2023. *The Mathematics of Shuffling Cards.* American Mathematical Society. AMS.

[8] Persi Diaconis, Jason Fulman, and Susan Holmes. 2013. Analysis of casino shelf shuffling machines. *The Annals of Applied Probability* 23, 4 (2013), 1692 – 1720. https://doi.org/10.1214/12-AAP884

[9] P. Diaconis, R. Graham, X. He, and S. Spiro. 2022. Card guessing with partial feedback. *Combinatorics, Probability and Computing* 31, 1 (2022), 1–20. https://doi.org/10.1017/S0963548321000134

[10] P. Diaconis, M. M. McGrath, and J. Pitman. 1995. Riffle Shuffles, Cycles, and Descents. *Combinatorica* 15 (1995), 11–29. https://doi.org/10.1007/BF01294457

[11] Janet R. Doner and V. R. R. Uppuluri. 1970. A Markov Chain Structure for Riff Shuffling. *Siam Journal on Applied Mathematics* 18 (1970), 191–209. https://api.semanticscholar.org/CorpusID:121483871

[12] E. Gilbert. 1955. *Theory of Shuffling.* Technical Memorandum. Bell Laboratories.

[13] AJ Grover. 2024. *Github Repo.* Technical Report. https://github.com/ajgrover/HonorsThesis

[14] J. Hadamard. 1906. Note de lecture sur J. Gibbs, "Elementary principles in statistical mechanics". *Bull. Amer. Math. Soc* 12 (1906), 194–210.

[15] Johan Jonasson and Benjamin Morris. 2015. Rapid mixing of dealer shuffles and clumpy shuffles. *Electron. Commun. Probab.* 20 (2015), 1–11. https://doi.org/10.1214/ECP.v20-3682

[16] D. D. Kosambi and U. V. Ramamohana Rao. 1958. The Efficiency of Randomization by Card Shuffling. *Royal Statistical Society. Journal. Series A: General* 121, 2 (3 1958), 223–233. https://doi.org/10.2307/2343362 arXiv:https://academic.oup.com/jrsssa/article-pdf/121/2/223/49762054/jrsssa_121_2_223.pdf

[17] Ben Morris. 2009. Improved Mixing Time Bounds for the Thorp Shuffle and L-Reversal Chain. *The Annals of Probability* 37, 2 (2009), 453–477. http://www.jstor.org/stable/30244289

[18] H. Poincaré. 1912. *Calcul des probabilités.* Georges Carré, Paris.

[19] J. Reeds. 1981. Theory of Riffle Shuffling. (1981). Unpublished manuscript.

[20] D. Silverman. 2019. Progressive Randomization of a Deck of Playing Cards: Experimental Tests and Statistical Analysis of the Riffle Shuffle. *Open Journal of Statistics* (2019). https://doi.org/10.4236/ojs.2019.92020

[21] Edward O Thorp. 1973. Nonrandom Shuffling with Applications to the Game of Faro. *J. Amer. Statist. Assoc.* 68, 344 (1973), 842–847. https://doi.org/10.1080/01621459.1973.10481434

[22] C. O. Williams. 1912. A Card Reading. *The Magician Monthly* (1912), 867.