Who would have thought card shuffling was so involved?

Leah Cousins

August 26, 2019

Abstract

In this paper we present some interesting mathematical results on card shuffling for two types of shuffling: the famous riffle shuffle and the random-to-top shuffle. A natural question is how long it takes for a deck to be randomized under a particular shuffling technique. Mathematically, this is the mixing time of a Markov chain. In this paper we present these results for both the riffle shuffle and the random-to-top shuffle. For the same two shuffles, we also include natural results for card shuffling which directly relate to well-known combinatorial objects such as Bell numbers and Young tableaux.

Contents

1 Introduction

Humans have been shuffling decks of cards since playing cards were first invented in 1000AD in Eastern Asia [25]. It is unclear when exactly card shuffling theory was first studied, but since card shuffling theory began with magicians' card tricks, it is fair to think that it was around the time that magicians were studying card tricks. The earliest known publication of card tricks was from 1726 in a book titled "Whole Art and Mystery of Modern Gaming" [10]. This is also the first known documentation of a "perfect shuffle": a shuffle where the deck is split equally in half and the cards are exactly interleaved.

One of the most common shuffling techniques is the overhand shuffle. To perform such a shuffle, one begins by holding the entire deck in one's left hand, then, using one's right hand, grabs a small pack of cards from the back of the deck using their thumb and fingers and inserting them at the front of the deck. The packets that are inserted at the front of the deck however are not released in the same order as they were grabbed; the shuffler will usually only drop one or two cards at a time from their right hand until the entire pack has been released. In 1989, Robin Permantle showed that the mixing time of the overhand shuffle with respect to variation distance is between order n^2 and order $n^2 \log n$ [21]. In 2006, Johan Jonasson showed, with the help of Wilson [29], that the mixing time of the overhand shuffle is indeed $n^2 \log n$. For a standard deck of 52 cards, that is well over four thousand shuffles.

The faro shuffle is a technique which was derived from the gambling card game "Faro" which was played in the late 17^{th} century [24]. To perform a faro shuffle, one splits the deck in half, preferably equally. Holding each half in either hand, the cards are pushed together from the corners, forcing them to interleave in between each other. There is no requirement for the cards to be perfectly interleaved, however some believe that this is necessary for a faro shuffle. A faro shuffle which leaves the original top card on the top and the original bottom card on the bottom of the shuffled deck is called an *out*-shuffle. Similarly, a shuffle which leaves the original top card as the second card and the original bottom card second last is called an *in*-shuffle. Since the faro shuffle is a controlled shuffle, there is no mixing time. However, many results have been proven regarding *in* and *out* shuffles. Scottish magician Alex Elmsley derived a method which moves the original top card to any desired position in the deck. The idea is to express the card's new position in binary, and perform an *in*-shuffle for every 1 and an *out*-shuffle for every 0 from left to right [18]. In 1983, Persi Diaconis, Ronald Graham and William Kantor showed, using number theoretical results, that the order of a perfect *in*-shuffle and *out*-shuffle is 2 mod $(2n \pm 1)$, respectively. *In* and *out*-shuffles also appear in computer science as a way of connecting processors in parallel processing machines [10].

The top-to-random shuffle is performed by removing the top card of the deck and inserting it back in the deck at random. This shuffle was first studied in the 1980s by Persi Diaconis and David Aldous [1], who showed that the mixing time is $O(n \log n)$. Later, in 1992, Diaconis, Fill and Pitman [8] gave a proof for the mixing time of a more general version called the top-m-to-random shuffle, where the first m cards are removed from the top and placed randomly in the deck one by one. The mixing time of this shuffle is $\frac{n}{m} \log n + c$ [8].

In 2016, Amy Pang [20] showed that a particular Markov chain modeling a scenario named "The relative time on a to-do list", consequently models the distribution of a deck under the top-to-random shuffle. Pang also describes a descent operator on a Hopf algebra that corresponds to the top-to-random shuffle [20]. The random-to-top shuffle is the inverse of top-to-random shuffle, and therefore, what results are valid for the top-to-random are also valid for the random-to-top, however they may have a slightly different proof.

The riffle shuffle was first studied by academics in the 1920's at Bell Laboratories. To perform a single riffle shuffle on a deck of cards, we first divide the deck into two parts. This is called a *cut*. Then, interleave the two parts together in any way such that each part remains in it's relative order in the deck. In 1992, Dr. Diaconis and Dr. David Bayer published their famous work "Trailing the Dovetail Shuffle to it's Lair", which gives a rigorous proof of the number of riffle shuffles it takes to randomize a deck of cards.

This paper will show that card shuffling is actually remarkably complicated mathematically. In particular, this paper analyzes two different shuffling techniques: the riffle shuffle and the random-to-top shuffle. We include the result of the natural question: how many shuffles it takes to randomize a deck of cards, as well as a few other card shuffling results which are directly related to well-known combinatorial objects and other unexpected mathematical phenomena such as the sequence of numbers which results from carrying over columns while adding large numbers by hand.

2 What is Card Shuffling, Really?

In plain English, card shuffling is defined as the act of mixing up the order of cards in a pack. Mathematically, card shuffling has more than one definition.

2.1 Shuffles are Bijective Functions

A **permutation** is a bijection from a finite set to itself, defined as follows. Consider a finite set of n objects, labeled 1 to n. A permutation σ defined on the set $\{1, \ldots, n\}$ takes, as input, an ordering of the elements in the set $\{1, \ldots, n\}$ and rearranges the order in a particular manner. It is important to emphasize that a permutation is *applied* to a particular ordering, and is not an ordering itself. In this paper, we write all permutations in cycle notation.

When applying a permutation to a set, the resulting ordering is written as a product of the original ordering and the permutation. In this paper, we will always multiply permutations on the left-hand-side. Let $\sigma = (15)(2)(3)(4) = (15)$, and suppose we apply σ to the ordering 12345. Then, we would write: $\sigma 12345 = (15)12345 = 52341$.

Since permutations are functions, the composition of two permutations σ and τ is also a permutation, which we will denote by $\tau\sigma$. Since we will be multiplying permutations on the left, we apply them from right-to-left. For example, suppose σ is the permutation described above and τ is the permutation $(1523)(4)$. The composition $\tau\sigma$ applied to 12345 is obtained by first applying σ to 12345 to obtain 52341, then applying τ to 52341 to obtain

23145. Henceforth, we refer to a composition of permutations as a *product* of permutations. Observe that we can also multiply $\tau\sigma$ first to obtain $(1523)(4) \cdot (15) = (123)(4)(5) = \alpha$ and then write $\alpha 12345 = 23145$.

Mathematically speaking, shuffling a deck of cards is equivalent to applying a permutation to an ordering of the deck. Moreover, repeatedly shuffling a deck is equivalent to applying a product of permutations to an ordering of the deck.

The Group of "Shuffles": S_n

The symmetric group defined on a set of n elements, denoted by S_n, consists of *all* bijective functions from the set to itself. Since each element in S_n is a function, the group operation is function composition. In this paper, we only consider S_n defined on a finite set of objects, therefore, the elements of S_n are all permutation operations which can be applied to a set of size n.

Since we can think of a single permutation as a single shuffle, then we can think of the group S_n as the group of *all* possible ways that we can shuffle a deck of cards, thus we can think of the group S_n as the foundation of the mathematical analysis of card shuffling.

It can also be thought of as a probability density on S_n [16]. When applied to an ordering, no two permutations give the same rearrangement. Therefore, each permutation in S_n can be thought of as a different way of rearranging the deck. Therefore, each permutation will have a particular probability of occurring, where the sum of the probabilities of all permutations is one.

2.2 Random Walks and Markov Chains

A **random walk** is a random process describing the path of an object through a succession of random steps on some mathematical space (for example the set of integers). In this paper we are only concerned about random walks defined on finite sets. In general, a random walk on a finite set is defined as follows:

Let G be a finite group. Let Q be a probability measure on G, where $Q(g)$ is the probability that $g \in G$ is chosen. The set of probabilities $\{Q(g) : g \in G\}$ is a *probability distribution* on G where $Q(g) \geq 0$ and $\sum Q(g) = 1$. Let $\xi_1, \xi_2, \ldots, \xi_n$ be G-valued random elements. We can think of ξ_i as the i^{th} step in the walk, where ξ_i takes on a value of some element in G. We define the following products:

$$\begin{aligned} X_0 &= \text{identity} \\ X_1 &= \xi_1 \\ &\vdots \\ X_k &= \xi_k X_{k-1} = \xi_k \xi_{k-1} \ldots \xi_1 \end{aligned}$$

The random variable X_i is the position of a randomly-moving object after i steps (with step distribution Q) and is what we call a *random walk* in the set G. For example, consider the random variable X_2, which is the position of a randomly-moving object after two steps. Suppose $X_2 = g$. The probability distribution of X_2 is given by the sum of $Q(h)Q(j)$ for all h, j such that $g = j \circ h$. This particular operation on Q is called *convolution* and we write $Q * Q = Q^{*2}$, therefore:

$$P(X_2 = g) = Q * Q(g) = Q^{*2}(G) = \sum_{h \in G} Q(h)Q(gh^{-1})$$

where $*$ is the convolution operation. In general, for X_k the probability distribution is given by a series of convolutions $Q * Q * Q * \cdots * Q = Q^{*k}$:

$$Q^{*k} = Q^* Q^{*k-1} = \sum_{h \in G} Q(h)Q^{(k-1)}(gh^{-1})$$

In 1906, Andrey Markov [17] proved that repeated convolutions converge to the uniform distribution U. That is,

$$Q^{*k}(g) \to U(g) = 1/|G| \quad \text{as} \quad k \to \infty \tag{1}$$

A random walk on a deck of n cards

We can model a sequence of shuffles as a random walk on the group S_n. Each step ξ_i is a randomly chosen permutation $\pi \in S_n$, X_i is the state of the deck after i shuffles, and $Q^{*i}(\pi)$ is the probability that the state after i shuffles is π.

A card shuffling random walk is, in fact, a Markov chain since the probability that the next step is $g \in G$ is only dependent on the current state. In particular, Markov chains which model repeated card shuffles are irreducible and aperiodic, since it is possible to go from any state to any other state, and, no state g requires a particular number of steps to return to state g. Most often, for card shuffling Markov chains, X_0 is the identity permutation with probability 1, and the transition probabilities are given by Q (as defined above). It is also important to note that most card shuffling Markov chains are regular. That is, $P(X_t = g) \geq 0$ for all times t and all $g \in G$. Therefore, the transition matrix for these types of Markov chains will always have positive entries.

2.3 What is Randomness?

One of the most important questions when studying the mixing time of a particular shuffle is what does it mean for a deck to be *randomized*? Intuitively, for a deck to be randomized, every configuration of the deck should be equally likely, but, it turns out most card shuffling techniques do not produce complete randomness in this sense [16]. Therefore, when studying randomness of a card shuffling technique, it is a matter of how *close* to random it can become and how long this takes.

Markov chains are commonly used to model repeated card shuffling [1] [7]. Moreover, since card shuffling Markov chains are aperiodic and irreducible, they have a limiting stationary distribution, which is unique regardless of the initial density (density of X_0). For most card shuffling methods the limiting stationary distribution is the uniform distribution $U_G = \frac{1}{n!}$ [7]. Knowing how long the chain takes to reach a stationary distribution is very important, but depending on which method is used to measure how "close to stationary" the chain gets, different results are obtained.

One method that is commonly used to measure this is called *total variation distance*. For card shuffling, the total variation distance is defined as

$$||P - U|| = \max_{A \subset G} |P(A) - U(A)| = \frac{1}{2} \sum_{g \in G} |P(g) - U(g)|$$

where G is the set of all possible permutations on a deck of size n, U is the limiting stationary distribution (denoted by U since it is usually the uniform distribution on G) and P is the probability distribution on G for that particular shuffle. By definition, $||P - U||$ will always be between 0 and 1, therefore when $||P - U||$ is small, we say that the chain is close to stationary and when $||P - U||$ is closer to 1, we say that it is far from stationary. It is important to note that total variation distance can still be used to measure the randomness of card shuffling techniques without using Markov chains. For instance, in Bayer and Diaconis' famous work [3], they model a sequence of riffle shuffles using what they call an *ab*-shuffle (see Theorem 3.3 below), and still use total variation distance to measure randomness.

Another way to measure the distance between probability measures is *separation distance* and the l_∞ metric. For card shuffling these measures are defined as

$$\text{sep}(m) = |G| \max_g \{U(g) - P(g)\} \qquad\qquad l_\infty(m) = \max_g \left| 1 - \frac{P(g)}{U(g)} \right|$$

The separation distance is an upper bound on the variation distance [1]. Note that separation distance can also be written as

$$\text{sep}(m) = \max_g 1 - \frac{P(g)}{U(g)}$$

Markov chains modelling natural processes, such as a sequence of card shuffles, show a sharp cutoff in their long term behaviour [7]. The "cut-off phenomenon" occurs when $||Q_1 - Q_2||$ stays close to its maximum value 1, and then suddenly drops to smaller and smaller values exponentially fast. For card shuffling Markov chains, the cut-off is what is used to determine the number of shuffles required to randomize the deck.

2.4 There are even shuffle *algebras* ?

An *alphabet* $\mathcal{A}$ is a set of elements which are called *letters*. A letter, in mathematical terms, is a symbol, whether it is a number or a letter from what the general public knows as the

alphabet. In this paper we will only be concerned with finite alphabets.

A *word* over an alphabet $\mathcal{A}$ is a sequence of letters from $\mathcal{A}$. We denote a word as a concatenated product of letters in a sequence. For example: $a_1a_2a_3 \dots a_m$ where $a_i \in \mathcal{A}$. The *length* of a word is the number of letters it contains, and is denoted by $|w|$. From any alphabet we can obtain the empty word ε which contains no letters, and hence $|\varepsilon| = 0$.

A *shuffle algebra* is defined over the set of all words over a particular alphabet, along with an operation defined on words which we call the *shuffle product*. Let w_1 and w_2 be two words from the same alphabet. We denote the shuffle product of w_1 and w_2 by $w_1 \sqcup\!\sqcup w_2$. The shuffle product between two words $w_1 = a_1a_2 \dots a_n$ and $w_2 = b_1b_2 \dots b_m$ is the sum of all permutations of $a_1, a_2, \dots, a_n, b_1, b_2, \dots b_m$ such that $a_1, a_2, \dots, a_n$ and $b_1, b_2, \dots, b_m$ stay in their relative order. For example, let $\mathcal{A}$ be an alphabet and suppose $w_1 = ab$ and $w_2 = 12$ are words over $\mathcal{A}$. Then, we have

$$w_1 \sqcup\!\sqcup w_1 = ab12 + a1b2 + a12b + 1a2b + 12ab + 1ab2$$

The shuffle product is, in fact, commutative and associative, which can be defined inductively by:

- $w \sqcup\!\sqcup \varepsilon = \varepsilon \sqcup\!\sqcup w = w$
- $ua \sqcup\!\sqcup vb = (u \sqcup\!\sqcup vb)a + (ua \sqcup\!\sqcup v)b$

where u and v are both words and a and b are letters [15].

In relation to cards, since words are sequences, they could also represent a sequence of cards in a deck. We can define an alphabet $\mathcal{D}$ to be a deck of cards, and thus the shuffle product between two words over $\mathcal{D}$ is exactly the sum of the possible ways of riffling them together.

3 The Riffle Shuffle

The first mathematical model of the riffle shuffle was developed by Bell Laboratories' mathematicians Edgar Gilbert and Claude Shannon in the 1950s. Shannon showed in some of his earlier work that if n cards are riffled together (2-shuffled) m times with $m \leq \log_2 n$, then all arrangements with 2^m rising sequences have the same probability [3]. In 1955, Bell Laboratories mathematicians abandoned their research with thoughts that there were too many ways to rearrange the deck [14]. Many years later, in 1981, Bell Labs scientist Jim Reeds independently came up with the same model as Gilbert and Shannon in an unpublished manuscript [22] which he later shared with Persi Diaconis [14]. Today, we call this model the Gilbert-Shannon-Reeds (GSR) model: cut the deck into two piles: one with c cards and the other with $n-c$ with probability $\binom{n}{c}/2^n$. Drop the cards from each pile such that: if the left pile has A cards and the right pile has B cards, then the probability that the next card is dropped from the left pile is $A/(A+B)$ and the probability that the next card is dropped

from the right pile is $B/(B+A)$. [9]

Consider the following example of a riffle shuffle on a deck of eight cards labeled 12345678. Assuming that the deck begins with the cards in the order 12345678, suppose the deck is cut into two parts: 123 and 45678. When interleaving these two parts, they must stay in the same order relative to the other cards in their original part. Consider one possible way these two parts can be interleaved: 14526378. Observe that the cards in the first part remain in the same relative order, namely 123 and the cards in the second part also remain in their relative order, namely, 45678.

3.1 The Mixing Time

The mixing time of the Riffle shuffle is one of the most famous results regarding card shuffling. In 1986, Aldous and Diaconis [1] showed that $2\log_2 n$ riffle shuffles are required to thoroughly mix a deck of n cards. Using a slightly different analysis, in 1992, Bayer and Diaconis [3] gave a slightly better bound $\frac{3}{2}\log_2 n$. Shortly after the work by Bayer and Diaconis [3] was published, an article in the *New York Times* [14] was published regarding their surprising result that only 7 riffle shuffles suffices to shuffle a deck of 52 cards. In this section, we explain the proof of the mixing time of the riffle shuffle given by Bayer and Diaconis in [3].

To begin we must first determine the probability of a particular ordering of a deck occurring under the riffle shuffle. To do this, we determine the probability of any cut occurring followed by any interleaving. The probability of a cut occurring after the first k cards is $\binom{n}{k}/2^n$. The probability of a particular interleaving is $1/\binom{n}{k}$. Therefore, the probability of any cut followed by any interleaving is

$$\binom{n}{k}/2^n \cdot 1/\binom{n}{k} = 1/2^n$$

Observe that this probability density is not dependent on k, which means that every combination of cut and interleaving is equally likely to occur.

Diaconis and Bayer [3] measure randomness using variation distance. As mentioned in section 2.3, variation distance has two definitions. The definition used in [3] is;

$$||R^m - U|| = \max_{A \subset S_n} |R^n(A) - U(A)| \tag{2}$$

where R^m is the GSR probability distribution after m shuffles, U is the uniform density on S_n and $U(A) = \sum_{\pi \in A} U(\pi)$. Observe that $U(\pi) = 1/n!$ for all $\pi \in S_n$.

A **rising sequence** is a maximal consecutively increasing subset of a larger set [16]. No two rising sequences in an ordering of a set intersect. Therefore, each ordering of a deck of cards is uniquely the disjoint union of its rising sequences [3].

A permutation π has a **descent** at card $k \in \{1, \ldots, n-1\}$ if $\pi(k) > \pi(k+1)$. For example, consider the permutation $\pi(12345) = 51324$. There are two descents in this ordering,

namely, at card 5 and at card 3. Typically, we would say the positions at which the descents occur in a permutation and not the card at which it occurs. Thus, for the example above, we would say that π has descents at positions 1 and 3. We can also class permutations by their descents: two permutations π and σ in S_n belong to the same class if $\pi(123...n)$ has descents in the same positions as $\sigma(123...n)$.

It is important to observe that the number of rising sequences and the number of descents are closely related. Consider the following corollary:

Corollary 3.1. *: A permutation π has r rising sequences if and only if π^{-1} has $r-1$ descents*

Proof. [3] Consider an arbitrary permutation π. Observe that the k^{th} entry of π determines the position of the letter k in π^{-1}. It follows that the k^{th} entry of π is a descent in π whenever the letter $i+1$ begins a new rising sequences in π. □

Using the example above, consider $\pi^{-1} = 24351$. Since 5 is in the first position in π, then there will be a rising sequence in π^{-1} beginning from the letter 2, namely (2,3)

An ***a*-shuffle** is a probability density on S_n defined as follows: Let $a \in \mathbb{N}$. Cut the deck into a sets with non-negative sizes $p_1, p_2, \ldots, p_a$, where $p_1+p_2+\cdots+p_a = n$. Now, interleave the cards such that each set stays in their relative order. The probability of an a-shuffle is $1/a^n$ (see section 6 in [16]), therefore, an a-shuffle is also a uniform distribution on S_n.

Theorem 3.2. *[3]: The probability that an a-shuffle will result in a permutation π is*

$$\binom{a+n-r}{n}/a^n \tag{3}$$

or equivalently,

$$\binom{a+n-d-1}{n}/a^n$$

where d is the number of descents in π^{-1}

Note that summing (3) over all permutations π gives 1 [3]. The number of permutations π with exactly r rising sequences is the Eulerian number $A_{n,r}$, where n is the number of cards in the deck. Since the number of a-shuffles that result in the permutation π with r rising sequences is $\binom{a+n-r}{n}$, and there are $A_{n,r}$ permutations with r rising sequences, then we have that

$$\sum_{r=1}^{n} A_{n,r}\binom{a+n-r}{n} = a^n \tag{4}$$

where a^n is the number of possible a-shuffles for an n-card deck. It is important to note that Eulerian numbers are symmetric. That is, $A_{n,r} = A_{n,n-r-1}$ (this can be proven using

an inductive argument on r).

In order to prove the mixing time of the riffle shuffle, we first need a way of representing repeated riffle shuffles. Diaconis and Bayer [3] represent repeated riffle shuffles as repeated 2-shuffles (a-shuffles for $a = 2$), and from there, determine R^m.

Theorem 3.3. *[3]: An a-shuffle followed by a b-shuffle, is equivalent to a single ab-shuffle.*

Proof. We prove this result using inverse a-shuffles: to perform an inverse a-shuffle, label each card in the deck with $1, \ldots, a$ uniformly and independently. Then, from top to bottom, sort the cards into piles according to their label. Once each card is placed into a pile, place the piles on top of each other in descending order such that pile 1 is on top of the deck and pile a is at the bottom.

Suppose we perform an inverse b-shuffle by labelling the cards in the top right corner with $1, \ldots, b$ independently at random, then sorting the 1's before the 2's, before the 3's and so on before the b's. Now perform an inverse a-shuffle, labeling the cards in the top left corner with $1, \ldots, a$ independently at random, then sorting the 1's before the 2's and so on before the a's. Observe that these two processes are equivalent to the following single process: Label the cards independently at random with $\{x, y\}$, where $x \in \{1, \ldots, a\}$ and $y \in \{1, \ldots, b\}$. First, order the cards according to the right-most label (from the b-shuffle). From here, re-order according to the left-most label. This is equivalent to sorting the cards labeled $\{1, 1\}$ before the cards labeled $\{1, 2\}$, the $\{1, 2\}$'s before the $\{1, 3\}$'s, ..., the $\{1, a\}$'s before $\{2, 1\}$'s, the $\{2, 1\}$'s before the $\{3, 1\}$'s and so on before the cards labeled $\{a, b\}$. This single process is the same as labeling each card $1, \ldots, ab$ independently at random and then sorting the 1's before the 2's,..., before the ab's, which is exactly an inverse ab-shuffle. Therefore, an inverse b-shuffle followed by an inverse a-shuffle is equivalent to an inverse ab-shuffle. It follows that an a-shuffle followed by a b-shuffle is equivalent to an ab-shuffle. □

Since a riffle shuffle is an a-shuffle for $a = 2$, then, by the theorem above, m riffle shuffles is the same thing as a 2^m-shuffle and that R^m is the probability density for a 2^m-shuffle. The probability of a permutation π occurring with r rising sequences after m shuffles from the GSR distribution is therefore

$$R^m(\pi) = \binom{2^m + n - r}{n} / 2^{mn}$$

Theorem 3.4. *[3] For $m = \log_2(n^{3/2}c)$ with $0 < c < \infty$ fixed, as n tends to ∞,*

$$||R^m - U|| = 1 - 2\Phi\left(\frac{-2^{-\theta}}{4c\sqrt{3}}\right) + O_c\left(\frac{1}{n^{1/4}}\right)$$

where $\Phi(x) = \int_{-\infty}^{x} e^{-t^2/2} dt/\sqrt{2\pi}$

Note that instead of writing $m = \frac{3}{2}\log_2 n + \theta$, Diaconis and Bayer write $m = \log_2(n^{3/2}c)$ so that $c = 2^\theta$ satisfies $0 < c < \infty$. The proof of Theorem 3.4 uses the asymptotics of

the Eulerian numbers. To see how Eulerian numbers are involved, we use the equivalent definition of variation distance:

$$||R^m - U|| = \frac{1}{2}\sum_{\pi \in S_n} |R^m(\pi) - U(\pi)|$$

Directly substituting the definition of $R_m(\pi)$ and $U(\pi)$, then for a permutation π with r rising sequences, we have:

$$|R^m(\pi) - U(\pi)| = \left| \binom{2^m + n - r}{n} / 2^{mn} - 1/n! \right|$$

Hence, from (4) we have:

$$||R^m - U|| = \frac{1}{2}\sum_{r=1}^{n} A_{n,r} \left| \binom{2^m + n - r}{n} / 2^{mn} - 1/n! \right|$$

Although we don't include a detailed proof of theorem 3.4, it is important to observe that the function $1 - 2\Phi\left(\frac{-1}{4c\sqrt{3}}\right)$ has the following asymptotic behaviour [3]:

$$1 - 2\Phi\left(\frac{-1}{4c\sqrt{3}}\right) \to \frac{1}{2c\sqrt{6\pi}} \quad \text{as } c \to \infty$$

$$1 - 2\Phi\left(\frac{-1}{4c\sqrt{3}}\right) \to 1 - \frac{4c\sqrt{3}}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{-1}{4c\sqrt{3}}\right)^2 \right] \quad \text{as } c \to 0$$

Recall that $m = \log_2(n^{3/2}c)$ and $c = 2^\theta$. Let $\theta = j$, be the number of shuffles performed after the first $\frac{3}{2}\log_2(n)$ shuffles.

3.2 Riffle Shuffling for Different Card Games

The mixing time of the riffle shuffle changes depending on what type of card game is being played. This is due to the fact that particular attributes such as suit, colour or number/face are irrelevant in some games. In Bayer and Diaconis' work [3], they show that it takes about $\frac{3}{2}\log_2(n)$ riffle shuffles to mix a deck of n cards. This is the mixing time taking every attribute into consideration, that is, suit, number/face and colour. In some card games, for example Blackjack and Baccarat, only certain attributes matter; the 10's and all face cards are considered "indistinguishable". In 2011, Assal et al. [2] expanded on prior research [23] [11] [6] to show the mixing time of the riffle shuffle for different card games. In this section we go through the main result of [2], which gives a general formula for the mixing time of the riffle shuffle with an n-card deck with multiple cards of the same type.

In this section, we use convolution powers to model repeated a-shuffles (these are defined in Section 2.2). Let $Q_a(w_a)$ denote the probability distribution for an a-shuffle where w_a represents any configuration resulting from an a-shuffle. Hence, we have $Q_a = 1/a^n\binom{a+n-r}{n}$.

Lemma 3.5. *Consider a deck of cards with D_i cards labeled i for $1 \leq i \leq v$. Starting with the deck in sorted order with 1's on top down to v's on the bottom, after an a-shuffle the most likely deck configuration is the sorted deck and the least likely deck is the reverse sorted deck w_a^* with v's on the top and 1's at the bottom.*

Proof. Given a sorted n-card deck described above, any cut of the deck into a packets may result in the initial sorted deck if the actual a-shuffle performed is the identity. Furthermore, the identity is at least as likely to occur as any other configuration. In order for the reverse sorted deck to occur after an a-shuffle, it is necessary that each of the a packets only contains one type of card. Every configuration of the deck that is possible under this type of cut is still equally likely to occur. Since w_a^* contributes to $Q_a(w_a)$, it follows that w_a^* will minimize $Q_a(w_a)$ and hence maximizes $1 - \frac{Q_a(w_a)}{U}$. □

In plain English, Lemma 3.5 says that the minimum separation distance is achieved once the original deck configuration has been completely reversed. To give a more mathematical motivation for Lemma 3.5, we present the following lemma from [2], which gives explicit formulas for these probabilities for a deck with two types of cards:

Lemma 3.6. *[2] Suppose an n-card deck contains D_i cards labeled i and D_j labeled j where $D_i + D_j = n$. The probability of an a-shuffle resulting in the sorted deck (i's on top and j s on the bottom) is*

$$\frac{1}{a^{D_i+D_j}}\left(\sum_{k=1}^{a}\left(k^{D_i}-(k-1)^{D_i}\right)(a-k+1)^{D_j}\right) \tag{5}$$

The probability of an a-shuffle resulting in the reverse sorted deck (j's on the top, i's on the bottom) is

$$\frac{1}{a^{D_i+D_j}}\left(\sum_{k=1}^{a-1}\left(k^{D_i}-(k-1)^{D_i}\right)(a-k)^{D_j}\right) \tag{6}$$

Proof. Let w_2 denote any configuration of an n-card deck with D_i cards labelled i and D_j cards labeled j (that is, two distinct labels). Let $A = (A_1, A_2, \ldots, A_a)$ be a non-negative cut of $D_i + D_j$ into a packets, where $A_l \geq 0$, $1 \leq l \leq a$ is the size of the l^{th} packet. Observe that the probability of any configuration occurring for this type of deck is

$$\sum_{A_1+A_2+\cdots+A_a=D_i+D_j}\frac{1}{a^n}\binom{D_i+D_j}{A_1, A_2, \ldots A_a}\text{prob}(w_2|A)$$

where $\text{prob}(w_2|A)$ is the probability that w_2 results from successively dropping cards from the packets A_i. Note, from Lemma 3.5 that in order for the reverse sorted deck to occur from an a-shuffle, every packet must contain the same type of card. Let k be an integer such that $A_1 + A_2 + \cdots + A_k = D_i$. Note that any of the $\binom{D_i+D_j}{D_i}$ configurations are equally likely after a subsequent shuffle under this type of cut A, we have $\text{prob}(w_2|A) = 1/\binom{D_i+D_j}{D_i}$ for each configuration w_2. Summing over these types of cuts A, we have

$$\sum_{\substack{A_1+A_2+\cdots+A_a=D_i+D_j \\ \exists\ k\ s.t.\ A_1+A_2+\cdots+A_k=D_i}} \frac{1}{a^{D_i+D_j}}\binom{D_i+D_j}{A_1,A_2,\ldots,A_a}\frac{1}{\binom{D_i+D_j}{D_i}}$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a-1}\sum_{A_k=1}^{D_i}\sum_{\substack{A_{k+1}+\cdots+A_a=D_j \\ A_1+\cdots+A_{k-1}=D_i-A_k}}\binom{D_i}{A_k}\binom{D_i-A_k}{A_1,\ldots,A_{k-1}}\binom{D_j}{A_{k+1},\ldots,A_a}$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a-1}(a-k)^{D_j}\sum_{A_k=1}^{D_i}\binom{D_i}{A_k}(k-1)^{D_i-A_k}$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a-1}(a-k)^{D_j}\left(k^{D_i}-(k-1)^{D_i}\right)$$

which is exactly (6) above.

Note from Lemma 3.5 that the packets of cards don't necessarily have to contain the same type of card to obtain a sorted deck. Hence, we may assume that there exists a packet with both cards labeled i and j. Therefore, there exists integers k, x, y such that $A_1+\cdots+A_{k-1}=D_i-x$, $A_k=x+y$ and $A_{k+1}+\cdots+A_a=D_j-y$, where $1\le k\le a$, $1\le D_i\le x$ and $1\le D_j\le y$. In order to obtain a sorted deck, we need packet A_k to be placed after all packets $A_1,\ldots,A_{k-1}$ and before all packets $A_{k+1},\ldots,A_a$. Therefore, there is only one way A_k can be placed amongst the rest of the packets. For this type of cut A, we have $\text{prob}(w_2|A)=\binom{D_i}{x}\binom{D_j}{y}/\binom{D_i+D_j}{D_i-x,x+y,D_j-y}$. Therefore, we have

$$\sum_{\substack{A_1+A_2+\cdots+A_a=D_i+D_j \\ \exists\ k\ s.t.\ A_1+A_2+\cdots+A_{k-1}<D_i \\ \text{and } A_{k+1}+\cdots+A_a<D_j}} \frac{1}{a^{D_i+D_j}}\binom{D_i+D_j}{A_1,A_2,\ldots,A_a}\text{prob}(w_2|A)$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a}\sum_{x=1}^{D_i}\sum_{y=1}^{D_j}\binom{D_i}{x}\binom{D_j}{y}\sum_{\substack{A_1+A_2+\cdots+A_{k-1}=D_i-x \\ A_{k+1}+\cdots+A_a=D_j-y}}\binom{D_i-x}{A_1,A_2,\ldots,A_{k-1}}\binom{D_j-y}{A_{k+1},\ldots,A_a}$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a}\sum_{x=1}^{D_i}\sum_{y=1}^{D_j}\binom{D_i}{x}\binom{D_j}{y}(k-1)^{D_i-x}(a-k)^{D_j-y}$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a}\sum_{x=1}^{D_i}\binom{D_i}{x}(k-1)^{D_i-x}\sum_{y=0}^{D_j}\binom{D_j}{y}(a-k)^{D_j-y}$$

$$= \frac{1}{a^{D_i+D_j}}\sum_{k=1}^{a}(a-k+1)^{D_j}\left(k^{D_i}-(k-1)^{D_i}\right)$$

which is exactly (5). □

The next theorem from [2] gives the separation distance

Theorem 3.7. *Consider a deck of n cards with D_i cards labeled i, $1 \leq i \leq v$. The separation distance after an a-shuffle of the sorted deck is given by*

$$1 - \frac{1}{a^n}\binom{n}{D_1, D_2, \ldots, D_v} \sum_{0=k_0<\cdots<k_{v-1}<a} (a-k_{v-1})^{D_v} \prod_{j=1}^{v-1} \left((k_j - k_{j-1})^{D_j} - (k_j - k_{j-1} - 1)^{D_j}\right)$$

Proof. From Lemma 3.5, we know that w_a^* can only result from an a-shuffle if all packets contain the same type of card. We also know that any cut of this type is equally likely to occur. Therefore Q is given by

$$Q_a(w_a^*) = \sum_{\substack{A_1+A_2+\cdots+A_a=n \\ A \text{ partitions } D}} \frac{1}{a^n}\binom{n}{A_1, \ldots, A_a} \frac{1}{\binom{n}{D_1,\ldots,D_v}} \tag{7}$$

where "A partitions D" means there exists indices $k_1, \ldots, k_{v-1}$ such that $A_1 + \cdots + A_{k_1} = D_1$, and for $i = 2, \ldots, v-1$, $A_{k_{i-1}+1} + \cdots + A_{k_i} = D_i$. In order to avoid overcounting compositions with empty packets, we may take the k_i's to be minimal. Hence, (7) simplifies to

$$\frac{1}{a^n} \sum_{0=k_0<\cdots<k_{v-1}<a} (a-k_{v-1})^{D_v} \prod_{j=1}^{v-1} \left((k_j - k_{j-1})^{D_j} - (k_j - k_{j-1} - 1)^{D_j}\right)$$

By Lemma 3.5, the result follows. □

Recall that the original analysis of the mixing time of the riffle shuffle uses total variation distance to measure the distance to uniformity, which is slightly difference from the separation distance measure. However, the results are still highly comparable. The following table gives the results from Bayer and Diaconis' work from [3] and from Assal et al. [2] for a deck of 52 cards after m 2-shuffles:

m	1	2	3	4	5	6	7	8	9
Bayer and Diaconis [3]	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**0.995**	0.928
Blackjack [2]	1.00	1.00	1.00	1.00	**0.999**	0.970	0.834	0.596	0.366

As we can see, since the game of Blackjack uses fewer "distinct" cards, fewer riffle shuffles are required to mix the deck. Since the result in Theorem 3.7 can be used for general deck with different numbers of particular type of cards, in general, fewer distinct cards as well as equal amounts of distinct cards will lead to less riffle shuffles to mix a deck.

3.3 Carries and Card Shuffles

Have you ever considered that the sequence of numbers that resulted from carrying while performing long addition were actually meaningful? In this section we will prove a result by Fulman and Diaconis [9] that relates riffle shuffling to the sequence of numbers carried while performing long addition.

Theorem 3.8. *[9]: The probability that the base-b carries chain goes from 0 to j in r steps is equal to the probability that the permutation in S_n obtained by performing r successive b-shuffles (started at the identity) has j descents*

The proof relies on the following results by Holte [12]:

- Given n integers (base b) whose digits were chosen at random from $\{0, 1, \ldots, b-1\}$, let $\kappa_0 = 0, \kappa_1, \ldots$ be the sequence of carries that results after adding them. The probability that the next carry is j given that the previous carry was i is

$$P_b(i,j) = \frac{1}{b^n} \sum_{l=0}^{j-\lfloor i/b \rfloor} (-1)^l \binom{n+1}{l} \binom{n-1-i+(j+1-l)b}{n} \tag{8}$$

  The matrix P_b whose entries are $P_b(i,j)$ is called an *amazing* matrix in [12].

- The product of two amazing matrices for base-a and base-b number respectively is the amazing matrix for base-ab numbers:

$$P_a P_b = P_{ab} \tag{9}$$

Holte found that the sequence of carries $\kappa_0 = 0, \kappa_1, \ldots$ from adding n base-b integers (chosen at random) forms a Markov chain taking values in $\{0, 1, \ldots, n-1\}$. While studying the Markov chain, he came up with the definition of the *amazing* matrix whose entries $P(i,j)$ are described above. In addition, Holte also showed that the j^{th} entry of the left eigenvector of an amazing matrix with eigenvalue 1 is $A_{n,j}/n!$. So, by the fundamental theorem of Markov chain theory, $A_{n,j}/n!$ is the long term frequency of carries of j when long random numbers are added [9].

The proof also relies on the following proposition from [9]:

Proposition 3.9. *Let π be a permutation with d descents. Let c_{ij}^d be the number of ordered pairs (τ, μ) of permutations in S_n such that τ has i descents and μ has j descents and $\tau\mu = \pi$. Then*

$$\sum_{i,j \geq 0} \frac{c_{ij}^d s^{i+1} t^{j+1}}{(1-s)^{n+1}(1-t)^{n+1}} = \sum_{a,b \geq 0} \binom{n+ab-d-1}{n} s^a t^b \tag{10}$$

Proof of Theorem 3.8. Since an a-shuffle followed by a b-shuffle is an ab-shuffle, then r b-shuffles is the same as a b^r-shuffle. By (2), the number of b^r-shuffles that result in a permutation π with $d(\pi^{-1})$ is $\binom{n+b^r-d(\pi^{-1})-1}{n}$. From the definition of $c_{i,j}^d$ above note that c_{ij}^0 is the number of pairs (τ, μ) such that $d(\tau) = i$, $d(\mu) = j$ and $\tau\mu = \mathrm{id}_{S_n}$. That is, $c_{i,j}^0$ is the number of permutations μ with $d(\mu) = j$ and $d(\mu^{-1}) = i$. For fixed i and j,

$$\binom{n+b^r-i-1}{n} c_{ij}^0$$

is the number of b^r-shuffles (starting at the identity) which results in a permutation μ with j descents. For all i, the number of b^r-shuffles that result in a permutation with j descents is

$$\sum_{i\geq 0}\binom{n+b^r-i-1}{n}c^0_{ij}$$

and therefore, the probability that a b^r-shuffle (starting from the identity) results in a permutation with j descents is

$$\sum_{i\geq 0}\frac{1}{b^{rn}}\binom{n+b^r-i-1}{n}c^0_{ij} \tag{11}$$

From (3), this gives

$$\sum_{i,k\geq 0}\frac{c^0_{i,k}s^{i+1}t^{k+1}}{(1-s)^{n+1}(1-t)^{n+1}}=\sum_{a,h\geq 0}\binom{n+ah-1}{n}s^at^h \tag{12}$$

Taking the coefficient of s^{b^r} of (8) on both sides and multiplying by $(1-t)^{n+1}$ we have

$$\sum_{i,k\geq 0}c^0_{i,k}\binom{n+b^r-i-1}{n}t^{k+1}=\sum_{h\geq 0}(1-t)^{n+1}\binom{n+b^rh-1}{n}t^h$$

Taking the coefficient of t^{j+1} on both sides and multiplying by $\frac{1}{b^{rn}}$, we get

$$\frac{1}{b^{rn}}\sum_{i\geq 0}c^0_{i,j}\binom{n+b^r-i-1}{n}=\frac{1}{b^{rn}}\sum_{l\geq 0}(-1)^l\binom{n+1}{l}\binom{n+b^r(j+1-l)-1}{n} \tag{13}$$

Observe that the left-hand side is exactly the probability that a b^r-shuffle results in a permutation with j descents.

From (4) the entries of the matrix P_{b^r} are

$$P_{b^r}(i,j)=\frac{1}{b^{rn}}\sum_{l=0}^{j-\lfloor i/b^r\rfloor}(-1)^l\binom{n+1}{l}\binom{n-1-i+(j+1-l)b^r}{n}$$

setting $i=0$ we obtain

$$P_{b^r}(0,j)=\frac{1}{b^{rn}}\sum_{l=0}^{j}(-1)^l\binom{n+1}{l}\binom{n-1+(j+1-l)b^r}{n}$$

which is exactly the right-hand side of (10). By (5) have have that $P_{b^r}(0,j)=P^r_b(0,j)$, which is the probability that the carries chain goes from 0 to j in r steps. □

4 The Random-to-top Shuffle

The random-to-top shuffle is the inverse of the top-to-random shuffle (or top-in-at-random shuffle). Consider a deck of n cards labeled 1 through n. To perform a single random-to-top shuffle, one chooses a card uniformly at random from the deck, removes it from the deck, and places it at the top (or front) of the deck, while all other cards remain untouched and in their relative original order. In this section of the paper, we show how random-to-top shuffles are related to two well-known combinatorial objects: Bell numbers and Young tableau. We also give a proof of the mixing time of the random-to-top shuffle.

Corollary 4.1. *There are n^n different sequences of n random-to-top shuffles that can be performed on a deck of size n*

Proof. Consider a deck of n cards. For each random-to-top shuffle, there are n choices for the card that will be chosen to move to the top of the deck. Since each sequence contains n shuffles, there are n^n in total. □

In Section 4.3, we show how many of these sequences return the deck to its original order.

4.1 Lifting Cards

We can think of a random-to-top shuffle as a simple act of "lifting" a card. In this section, we give a detailed explanation of what it means to "lift" a card, as well as some new terminology.

Consider a deck of n distinct cards $1, 2, \ldots, n$. The **state** of the deck is the ordering of its cards, denoted by $\rho = \rho_1\rho_2\cdots\rho_n$, where ρ_i is the card in position i for all $i \in [n]$. That is, the states of the deck are permutations of the elements $\{1, 2, \ldots, n\}$, which are exactly the elements in the symmetric group on n elements, denoted by S_n.

We say that cards ρ_i and ρ_j are in **natural order** when $\rho_i < \rho_j$, for $i < j$, and that cards ρ_i and ρ_j are **out of order** when, $\rho_i > \rho_j$ for $i < j$. When cards $\rho_i < \rho_j$, for $i < j$, for all $i, j \in [n]$, we say that the deck is in **natural state**. That is, the deck is in natural state $\rho_{natural} = 12345...n$ when card $\rho_i = c$ is in position $i = c$ for all $c \in [n]$. This is exactly the identity permutation of S_n, denoted by id_n.

We transform the deck by applying a sequence of random-to-top shuffles, which we call ***lifts***. We refer to the time at which we lift card c as a **stage**. If no such card $c \in [n]$ has been lifted, we say that the deck is in the **initial stage**. In terms of the underlying permutation, we define lifting c as the following: suppose we have a deck in state $\rho = \rho_1\rho_2\cdots\rho_n$ at an arbitrary stage m. To lift $c \in [n]$, identify position i such that $\rho_i = c$. Remove c from its current position, and place it at the top of the deck (in position 1), keeping all other cards in their relative order. After lifting $\rho_i = c$, the deck will be in the state $\rho' = \rho_i\rho_1\ldots\rho_{i-1}\rho_{i+1}\ldots\rho_n$. In terms of the symmetric group, $\rho' = \sigma_{i_m}(\rho)$, where $\sigma_{i_m} = (12\cdots i)$, is a permutation in Sym_n.

4.2 Properties of the Lifting Process

In the previous section we explained what it meant to "lift" a card. For consistency and clarity, we list a few important properties of the lifting process.

Consider a deck of cards $1, 2, \ldots, n$ in natural state. Suppose we apply a sequence of shuffles $\sigma_{a_t} \ldots \sigma_{a_2} \sigma_{a_1}$, where $a_i \in [n]$. Suppose at some stage, cards $c+1, c+2, \ldots, n$ are in natural order, cards $c, c+1$ are out of order, and no card from $c+1, c+2, \ldots, n$ is lifted after that stage. We call this **Property c**, since, at this particular stage, c would be the greatest card out of order with respect to cards $c+1, c+2, \ldots, n$. If the cards are in such an arrangement at a particular stage, we say that the deck *satisfies* Property c.

1. After card c is lifted, then c is in natural order with respect to cards $c+1, c+2, \ldots, n$ and out of order with respect to cards $1, 2, \ldots, c-1$.

2. The only way to change the relative order of cards c and $c+1$ is to either lift c or $c+1$. That is, lifting any other card will not affect the relative order of cards c and $c+1$.

3. If at some stage, the order of the deck satisfies Property c, for some $c \in [n]$, then in order for the deck to be put into natural order, c must be lifted at least once after that stage. Moreover, in the *last* subsequent stage at which c is lifted, no card from $c, c+1, \ldots, n$ will be lifted thereafter and cards $c, c-1$ will now be out of order, hence the deck will satisfy Property $c-1$. That is, at the last subsequent stage that card $c \in [n]$ is lifted, the deck will satisfy property $c-1$.

4. If k is the *greatest* card lifted among all stages, then cards $k, k+1, \ldots, n$ will be in natural order in every state. Moreover, at the last subsequent stage that k is lifted, k will be out of order with cards $k-1, k-2, \ldots, 2, 1$, and the deck will satisfy Property k-1.

4.3 Bell Numbers and the Random-to-top Shuffle

A partition of a set is a collection of nonempty, pairwise disjoint subsets of a set S, whose union is S. Bell numbers count the number of ways a set can be partitioned. We denote by B_t the Bell number for a set of size t.

For $t, n \in \mathbb{N} \cup \{0\}$, let $\mathrm{B}_t(n)$ denote the number of set partitions of $\{1, 2, 3, \ldots, t\}$ into at most n parts. Note that if $n \geq t$, then $\mathrm{B}_t(n)$ is the Bell number B_t. Define $\mathrm{B}'_t(n)$ analogously, such that no part contains both 1 and t or both i and $i+1$ for $i \in \{1, 2, \ldots, t\}$.

For $t, n \in \mathbb{N} \cup \{0\}$, let $\mathrm{S}_t(n)$ be the number of sequences of t random-to-top shuffles whose product is the identity permutation. Define $\mathrm{S}'_t(n)$ analogously, excluding the identity permutation.

Lemma 4.2. *[5]: If $t, n \in \mathbb{N} \cup \{0\}$ then $B_t(n) = S_t(n)$ and $B'_t(n) = S'_t(n)$*

Proof. For $a_i \in [n]$, let $\sigma_{a_t} \cdots \sigma_{a_2}\sigma_{a_1}$, be a sequence of random-to-top shuffles such that $\sigma_{a_t} \cdots \sigma_{a_2}\sigma_{a_1} = \mathrm{id}_n$. Consider a deck of cards $1, 2, \ldots, n$ in the natural state. Apply shuffles, $\sigma_{a_t} \cdots \sigma_{a_2}\sigma_{a_1}$, from right to left. For each card $c \in [n]$, let $\mathbf{A}_c$ denote the set of stages $s \in [t]$ at which card c is lifted.

Now, consider the subsets $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_n$. If k is the greatest card lifted among all stages, then by Property 4, the sets $\mathbf{A}_{k+1}, \mathbf{A}_{k+2}, \ldots, \mathbf{A}_n$, are empty (since there is no stage $s \in [t]$ at which any of $k+1, k+2, \ldots, n$ are lifted). Moreover, there will be exactly $n-k$ empty sets. By Property 3, the sets $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k$ are all non-empty.

For $c \in \{1, 2, \ldots, k\}$, let μ_c denote the greatest element in $\mathbf{A}_c$. The greatest element in $\mathbf{A}_c$ is the last stage in which card c is lifted. Therefore, by Property 3, we have that:

$$\mu_c > \mu_{c+1} > \mu_{c+2} > \cdots > \mu_k$$

Therefore, $\mathbf{A}_c$ contains the greatest element in $\{1, 2, \ldots, t\} \backslash \mathbf{A}_{c-1} \cup \mathbf{A}_{c-2} \cup \cdots \cup \mathbf{A}_1$.

If we discard the empty sets $\mathbf{A}_{k+1}, \mathbf{A}_{k+2}, \ldots, \mathbf{A}_n$, we are left with a set partition of $[t]$, $\{\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_k\}$, in canonically decreasing order of largest element in each block.

We claim that given an arbitrary set partition of $[t]$ into $p \leq n$ parts, we can label each block $\mathbf{A}_1, \mathbf{A}_2, ..., \mathbf{A}_p$, which will determine a sequence of lifts that, when applied to a deck of n cards in natural state, leaves the deck in natural state.

Let Φ be a set partition of $[t]$ into $p \leq n$ blocks. Find the largest element in each part; since each block is non-empty, each one will have a largest element. Order the blocks from left to right in decreasing order of greatest element. Since each block has a unique largest element, there is only one way to arrange the blocks in this order. Label the blocks from left to right with $\mathbf{A}_1, \mathbf{A}_2, \ldots, \mathbf{A}_p$, respectively. That is, label the block with the ℓ^{th} largest element $\mathbf{A}_\ell$. Note that there is only one way to canonically order these blocks into decreasing order of largest element. Therefore, the condition above fixes $\mathbf{A}_1, \ldots, \mathbf{A}_p, \ldots, \mathbf{A}_n$, and determines a unique sequence of lifts, which, when applied to a deck of n cards in the natural state, returns the deck to natural state. Therefore, $\mathrm{B}_t(n) = \mathrm{S}_t(n)$.

A set partition corresponds to a sequence of non-identity shuffles if and only if the same card is never lifted in consecutive stages, i.e. $\mathbf{A}_c$ cannot contain both s and $s+1$ for all $s \in [t]$, and $1 \notin \mathbf{A}_1$. Observe that this is just a restriction of the above bijection, and we get that $\mathrm{B}'_t(n) = \mathrm{S}'_t(n)$.

□

Consider the following example. Let $n = 9$ and $t = 8$, and consider the sequence of shuffles $\sigma_5\sigma_6\sigma_6\sigma_4\sigma_6\sigma_3\sigma_6\sigma_4$.

Card Lifted	State of the deck	Stage	Shuffle	Position lifted
-	1 2 3 4 5 6 7 8 9	0	id_n	-
lift 4	**4** 1 2 3 5 6 7 8 9	1	σ_4	4
lift 6	**6** 4 1 2 3 5 7 8 9	2	σ_6	6
lift 1	**1** 6 4 2 3 5 7 8 9	3	σ_3	3
lift 5	**5** 1 6 4 2 3 7 8 9	4	σ_6	6
lift 4	**4** 5 1 6 2 3 7 8 9	5	σ_4	4
lift 3	**3** 4 5 1 6 2 7 8 9	6	σ_6	6
lift 2	**2** 3 4 5 1 6 7 8 9	7	σ_6	6
lift 1	**1** 2 3 4 5 6 7 8 9	8	σ_5 (id_n)	5

By definition, $\mathbf{A}_c$ contains stages $s \in \{1, 2, \ldots, t\}$ in which card c is lifted, which means, we only need to look at the first card in each state, and the corresponding stage to determine the sets:

$$\begin{aligned}
\mathbf{A}_1 &= \{3, 8\} \\
\mathbf{A}_2 &= \{7\} \\
\mathbf{A}_3 &= \{6\} \\
\mathbf{A}_4 &= \{1, 5\} \\
\mathbf{A}_5 &= \{4\} \\
\mathbf{A}_6 &= \{2\} \\
\mathbf{A}_7 &= \emptyset \\
\mathbf{A}_8 &= \emptyset \\
\mathbf{A}_9 &= \emptyset
\end{aligned}$$

Now suppose we discard the empty sets $\mathbf{A}_7, \mathbf{A}_8$ and $\mathbf{A}_9$. Then we are left with

$$\begin{aligned}
\mathbf{A}_1 &= \{3, 8\} \\
\mathbf{A}_2 &= \{7\} \\
\mathbf{A}_3 &= \{6\} \\
\mathbf{A}_4 &= \{1, 5\} \\
\mathbf{A}_5 &= \{4\} \\
\mathbf{A}_6 &= \{2\}
\end{aligned}$$

in canonical decreasing order of largest element. Hence the set partition is

$$\{\{1,5\},\{2\},\{3,8\},\{4\},\{6\},\{7\}\}$$

Now consider an example of the reverse process. Suppose we partition a set of $t = 7$, into $p = 3$ parts, and let $n = 8$,

$$\{\{1,3\},\{2,5,7\},\{4,6\}\}$$

Re-order the sets from left to right in decreasing order of largest element, to obtain:

$$\{\{2,5,\mathbf{7}\},\{4,\mathbf{6}\},\{1,\mathbf{3}\}\}$$

Label the sets from left to right with $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ respectively. That is, $\mathbf{A}_1 = \{2,5,7\}, \mathbf{A}_2 = \{4,6\},$ and $\mathbf{A}_3 = \{1,3\}$. From here we can determine the corresponding sequences of lifts:

Stage	Card lifted
1	3
2	1
3	3
4	2
5	1
6	2
7	1

from which we can determine the sequence of shuffles, when we apply them to a deck of $n = 8$ cards in natural state:

Stage	Card lifted	State	Shuffle
-	-	12345678	id_n
1	3	31245678	σ_3
2	1	21345678	σ_3
3	3	32145678	σ_2
4	2	23145678	σ_3
5	1	12345678	σ_2
6	2	21345678	σ_2
7	1	12345678	σ_2

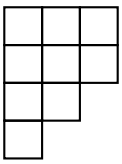
Remark. *Observe that Lemma 4.1 is equivalent to the following: The number of sequences of n random-to-top shuffles that return a deck of size n to its original order is B_n.*

Since there are $n!$ permutations of a deck of size n, then from Lemma 4.2, we get that the probability that a sequence of n random-to-top shuffles returns a deck of size n to it's original order is $B_n/n!$, which is notably larger than $1/n!$.

4.4 Young Diagrams

Young diagrams are most often used to describe and study representations of the symmetric group. For every partition of the elements of a finite group, there is a Young diagram. A Young diagram is an arrangement of boxes (or cells). For each integer in the partition, there is a row of boxes; one box per element. These rows are stacked on top of one another such that the left-most boxes are all aligned. From top-to-bottom, the rows are arranged in non-increasing order with respect to row-length (i.e. the number of boxes in the row).

For example, consider the partition $\lambda = \lambda_1, \lambda_2, \lambda_3, \lambda_4$ of 9, where $\lambda_1 = 3, \lambda_2 = 3, \lambda_3 = 2, \lambda_4 = 1$. The following diagram is the corresponding Young diagram for this partition:



We say that a box in a Young diagram is **removable** if removing it leaves the Young diagram of a partition; a partition is said to be **addable** if, when we add a box, we get a Young diagram of a partition.

A **move** on a partition consists of the removal of a removable box and then addition in an addable position of a single box.

A move is **exceptional** if it consists of the removal and then addition in the same place of the lowest removable box.

Given partitions λ and μ of the same size, let $\mathrm{M}_t(\lambda, \mu)$ denote the number of sequences of t moves that start at λ and finish at μ. Let $\mathrm{M}'_t(\lambda, \mu)$ be defined analogously, considering only non-exceptional moves. For $n \in \mathbb{N} \cup \{0\}$ let $\mathrm{M}_t((n), (n)) = \mathrm{M}_t(n)$, and $\mathrm{M}'_t((n), (n)) = \mathrm{M}'_t(n)$. If the Young diagram of a partition λ has exactly r removable boxes, then $\mathrm{M}_t(\lambda, \mu) = r$ and $\mathrm{M}'_t(\lambda, \mu) = r - 1$.

Let $\mathrm{Cl}(S_n)$ denote the ring of class functions of S_n. That is, $\mathrm{Cl}(S_n)$ consists of all functions that are constant on the conjugacy classes. Let $\pi \in \mathrm{Cl}(S_n)$ be the character of S_n, defined by $\pi(t) =\mid \mathrm{fix}(\tau) \mid$ for $\tau \in S_n$.

Let χ^λ denote the irreducible character of S_n canonically labeled by the partition λ of n. Let $\vartheta = \pi - 1_{S_n}$ where 1_{S_n} is the trivial character of S_n. Then we have that $\vartheta = \chi^{(n-1,1)}$.

Let $\phi \in \mathrm{Cl}(S_n)$. Then, for all $\sigma, \tau \in S_n$ we have $\phi(\sigma) = \phi(\tau^{-1}\sigma\tau)$. Let S_{n-1} be the subgroup of S_n consisting of all permutations on the set $\{1, 2, 3, \ldots, n\}$ in which n is a fixed point. For $i \in [n-1]$, define $t_i = (in) \in S_n$ to be the transposition that interchanges i and n, and let $t_n = (nn)$ denote the identity permutation. Then $\{t_1, t_2, \ldots, t_n\}$ is a set of left coset representatives for S_n with respect to the subgroup S_{n-1}, that is,

$$S_n = t_1 S_{n-1} \cup t_2 S_{n-1} \cup \cdots \cup t_n S_{n-1}.$$

Let Φ be a representation of $S_{n-1} \subset S_n$ of degree d. For every element $s \in S_n$ define a matrix

$$\Psi_s = \begin{bmatrix} \Phi(t_1^{-1}st_1) & \Phi(t_1^{-1}st_2) & \Phi(t_1^{-1}st_3) & \dots & \Phi(t_1^{-1}st_n) \\ \Phi(t_2^{-1}st_1) & \Phi(t_2^{-1}st_2) & \Phi(t_2^{-1}st_3) & \dots & \Phi(t_2^{-1}st_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi(t_n^{-1}st_1) & \Phi(t_n^{-1}st_2) & \Phi(t_n^{-1}st_3) & \dots & \Phi(t_n^{-1}st_n) \end{bmatrix}$$

Note that each Ψ_s is an $n \times n$ array of blocks, each of which has degree d. A fundamental result of Frobenius says that Ψ is a representation of S_n. We say that Ψ is the *induced* representation of S_n and we write $\Psi = \Phi \uparrow_{S_{n-1}}^{S_n}$. Therefore, the induced character of $\Phi \uparrow_{S_{n-1}}^{S_n}$, denoted by $\phi \uparrow_{S_{n-1}}^{S_n}$, is defined as follows:

$$\begin{aligned} \phi \uparrow_{S_{n-1}}^{S_n} &= \mathrm{Tr}(\Phi_s \uparrow_{S_{n-1}}^{S_n}) \\ &= \mathrm{Tr}(\Psi_s) \\ &= \mathrm{Tr} \begin{bmatrix} \Phi(t_1^{-1}st_1) & \Phi(t_1^{-1}st_2) & \Phi(t_1^{-1}st_3) & \dots & \Phi(t_1^{-1}st_n) \\ \Phi(t_2^{-1}st_1) & \Phi(t_2^{-1}st_2) & \Phi(t_2^{-1}st_3) & \dots & \Phi(t_2^{-1}st_n) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \Phi(t_n^{-1}st_1) & \Phi(t_n^{-1}st_2) & \Phi(t_n^{-1}st_3) & \dots & \Phi(t_n^{-1}st_n) \end{bmatrix} \\ &= \sum_{i=1}^{n} \mathrm{Tr}(\Phi(t_i^{-1}st_i)) \\ &= \sum_{i=1}^{n} \phi(t_i^{-1}st_i) \end{aligned}$$

where the summation is over all i such that $t_i^{-1}st_i \in S_{n-1}$. Recall that for all $s \in S_n$, we have $\pi(s) =\mid \mathrm{fix}(s) \mid$. Therefore we have that

$$\begin{aligned} \pi(s) &= \text{number of fixed points in } s \\ &= \text{number of } i \in \{1, 2, \dots, n\} \text{ such that } n \text{ is a fixed point of } t_i^{-1}st_i \\ &= \text{number of } i \in \{1, 2, \dots, n\} \text{ such that } t_i^{-1}st_i \in S_{n-1} \\ &= \sum_{\substack{i=1 \\ t_i^{-1}st_i \in S_{n-1}}} 1 \end{aligned}$$

It follows that, for any character ϕ of S_n, we get

$$\phi(s)\pi(s) = \sum_{i=1}^{n} \phi(t_i^{-1}st_i) \tag{14}$$

and hence,

$$\phi\pi = \phi(1 \uparrow_{S_{n-1}}^{S_n}) = (\phi \downarrow_{S_{n-1}}) \uparrow^{S_n} \tag{15}$$

In any group G, for any complex-valued functions f, g defined on G, we have the group inner product

$$\langle f, g\rangle_G = \tfrac{1}{|G|} \sum_{x\in G} f(x)\overline{g(x)}$$

where $\overline{g(x)}$ is the complex conjugate of $g(x)$.

Frobenius reciprocity: For any subgroup H of G, and any character f of H, and g of G, we have

$$\langle f, g \uparrow_H^G\rangle_G = \langle f \downarrow_H^G, g\rangle_H$$

where $g \downarrow_H^G$ is the restricted character of a representation $\mathfrak{G}$ of G, where, for some $\varepsilon \in G$

$$\begin{aligned} g \downarrow_H^G (\varepsilon) &= \mathrm{Tr}(\mathfrak{G} \downarrow_H^G (\varepsilon)) \\ &= \mathrm{Tr}(\mathfrak{G}(\varepsilon)) \\ &= g(\varepsilon) \end{aligned}$$

In the following Lemma, we show, using Frobenius reciprocity and properties of irreducible characters of the symmetric group, how $M_t(\lambda, \mu)$ can be written as a tensor product of characters of S_n.

Lemma 4.3. *[5]: Let $t \in \mathbb{N} \cup \{0\}$. If λ and μ are partitions of $n \in \mathbb{N}$ then $M_t(\lambda, \mu) = \langle \chi^\lambda \pi^t, \chi^\mu\rangle$ and $M_t' = \langle \chi^\lambda \vartheta^t, \chi^\mu\rangle$.*

Proof. Let χ^λ denote the irreducible character of S_n indexed by the partition λ of n. A standard fact about irreducible characters is the orthogonality relation

$$\langle \chi^\lambda, \chi^\mu\rangle = \begin{cases} 1 \text{ if } \lambda = \mu \\ 0 \text{ otherwise} \end{cases} \tag{16}$$

For a partition α of n and β of $n-1$, we write $\alpha = \beta^+$ when the Young diagram of α can be obtained by adding an addable box to the Young diagram of β. Similarly, we write $\beta = \alpha^-$ when the Young diagram of β can be obtained by removing a removable box from the Young diagram of α. By the Branching Rule, we have that

$$\chi_\alpha = \sum_{\gamma=\alpha^-} \chi^\gamma \tag{17}$$

For the base case, let $t = 0$. We have

$$\langle \chi^\lambda \pi^0, \chi^\mu \rangle_{S_n} = \langle \chi^\lambda, \chi^\mu \rangle_{S_n} = \begin{cases} 1 \text{ if } \lambda = \mu \\ 0 \text{ otherwise} \end{cases}$$

which is exactly $\mathrm{M}_0(\lambda, \mu)$, since the number of sequences of 0 moves that start at λ and end at μ is 1 when $\lambda = \mu$ and 0 otherwise.

For the inductive hypothesis, assume that the result is true for $m = k$, for some arbitrary integer $k \geq 0$, hence

$$\langle \chi^\lambda \pi^k, \chi^\mu \rangle_{S_n} = \mathrm{M}_k(\lambda, \mu) \tag{18}$$

for any partitions λ and μ of n.

Since $\chi^\lambda \pi^k$ is a character of S_n, we can express this as a linear combination of irreducible characters of S_n. That is,

$$\chi^\lambda \pi^k = \sum_\alpha \mathrm{C}_\alpha \chi^\alpha$$

where $\mathrm{C}_\alpha \in \mathbb{N} \cup \{0\}$, and the summation is over all partitions α of n. Therefore, we have

$$\begin{aligned} \langle \chi^\lambda \pi^k, \chi^\mu \rangle_{S_n} &= \langle \sum_\alpha \mathrm{C}_\alpha \chi^\alpha, \chi^\mu \rangle_{S_n} \\ &= \sum_\alpha \mathrm{C}_\alpha \langle \chi^\alpha, \chi^\mu \rangle_{S_n} \\ &= \mathrm{C}_\mu \end{aligned}$$

where C_μ is the coefficient for any partition μ of n. Therefore, we have

$$\chi^\lambda \pi^k = \sum_\alpha \mathrm{M}_k(\lambda, \alpha) \chi^\alpha \tag{19}$$

To prove the result for $m = k + 1$ we have

$$\begin{aligned}
\langle \chi^\lambda \pi^{k+1}, \chi^\mu \rangle_{S_n} &= \langle \left(\chi^\lambda \pi^k\right) \pi, \chi^\mu \rangle_{S_n} \\
&= \langle \left(\left(\chi^\lambda \pi^k\right) \downarrow_{S_{n-1}}\right) \uparrow^{S_n}, \chi^\mu \rangle_{S_n} && \text{from (15)} \\
&= \langle \left(\left(\sum_\alpha \mathrm{M}_k(\lambda, \alpha) \chi^\alpha\right) \downarrow_{S_{n-1}}\right) \uparrow^{S_n}, \chi^\mu \rangle_{S_n} && \text{from (19)} \\
&= \sum_\alpha \mathrm{M}_k(\lambda, \alpha) \langle \left((\chi^\alpha) \downarrow_{S_{n-1}}\right) \uparrow^{S_n}, \chi^\mu \rangle_{S_n} \\
&= \sum_\alpha \mathrm{M}_k(\lambda, \alpha) \langle (\chi^\alpha) \downarrow_{S_{n-1}}, \chi^\mu \downarrow_{S_{n-1}} \rangle_{S_{n-1}} && \text{by Frobenius reciprocity} \\
&= \sum_\alpha \mathrm{M}_k(\lambda, \alpha) \langle \sum_{\gamma = \alpha^-} \chi^\gamma, \sum_{\nu = \mu^-} \chi^\nu \rangle_{S_{n-1}} && \text{from (17)} \\
&= \sum_\alpha \mathrm{M}_k(\lambda, \alpha) \sum_{\gamma = \alpha^-} \sum_{\nu = \mu^-} \langle \chi^\gamma, \chi^\nu \rangle_{S_{n-1}} \\
&= \sum_\alpha \mathrm{M}_k(\lambda, \alpha) \sum_{\alpha^- = \mu^-} 1 && \text{from (16)}
\end{aligned}$$

Since $\sum_{\alpha^- = \mu^-} 1$ is the number of ways that a single move can start with α and end with μ, we have

$$\textstyle\sum_\alpha \mathrm{M}_k(\lambda, \alpha) \sum_{\alpha^- = \mu^-} 1 = \mathrm{M}_{k+1}(\lambda, \mu)$$

which proves the result for $m = k + 1$ and hence finishes the proof by induction on $k \geq 1$.

Now, we give the result for the second statement in the Lemma.Let $\vartheta = \pi - 1_{S_n}$, where 1_{S_n} is the trivial character of S_n. Similar to the case above, when $t = 1$ we have

$$\langle \chi^\lambda \vartheta^0, \chi^\mu \rangle_{S_n} = \langle \chi^\lambda, \chi^\mu \rangle_{S_n} = \begin{cases} 1 \text{ if } \lambda = \mu \\ 0 \text{ otherwise} \end{cases}$$

which is equal to $\mathrm{M}_0(\lambda, \mu)$, hence $\mathrm{M}_0(\lambda, \mu) = \mathrm{M'}_0(\lambda, \mu)$.

For the induction hypothesis, assume the result is true for $m = k$ for some arbitrary integer $k \geq 0$, hence

$$\langle \chi^\lambda \vartheta^k, \chi^\mu \rangle_{S_n} = \mathrm{M'}_k(\lambda, \mu) \tag{20}$$

for any partitions λ and μ of n. Similarly to the case above we also have the following summation result

$$\chi^\lambda \vartheta^k = \sum_\alpha \mathrm{M'}_k(\lambda, \alpha) \chi^\alpha \tag{21}$$

so we obtain

$$\begin{aligned}
\langle \chi^\lambda \vartheta^{k+1}, \chi^\mu \rangle &= \langle (\chi^\lambda \vartheta^k)\vartheta, \chi^\mu \rangle \\
&= \langle (\sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\chi^\alpha)\vartheta, \chi^\mu \rangle \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\langle \chi^\alpha \vartheta, \chi^\mu \rangle \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\langle \chi^\alpha(\pi - 1), \chi^\mu \rangle \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\langle \chi^\alpha \pi, \chi^\mu \rangle - \langle \chi^\alpha, \chi^\mu \rangle\big] \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\langle (\chi^\alpha \downarrow_{\S_{n-1}}, \chi^\mu \rangle - \langle \chi^\alpha, \chi^\mu \rangle\big] \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\langle \chi^\alpha \downarrow_{S_{n-1}}, \chi^\mu \downarrow_{S_{n-1}} \rangle_{S_{n-1}} - \langle \chi^\alpha, \chi^\mu \rangle\big] \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\langle \sum_{\gamma=\alpha^-} \chi^\gamma, \sum_{\nu=\mu^-} \rangle_{S_{n-1}} - \langle \chi^\alpha, \chi^\mu \rangle\big] \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\sum_{\gamma=\alpha^-}\sum_{\nu=\mu^-} \langle \chi^\gamma, \chi^\nu \rangle_{S_{n-1}} - \langle \chi^\alpha, \chi^\mu \rangle\big] \\
&= \sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\sum_{\alpha^-=\mu^-} 1 - 1\big]
\end{aligned}$$

Recall that $\mathrm{M'}_k(\lambda,\mu)$ is the number of sequences of k moves that start at λ and end at μ, only considering non-exceptional moves (moves which do not consist of removing and adding the lowest removable box). Observe that in order to start with λ and end with μ by performing an exceptional move, we must have that $\lambda = \mu$. Therefore in order to be able to calculate $\mathrm{M'}_k(\lambda,\mu)$ we need to count the number of exceptional moves, and hence we need $\lambda = \mu$.

In our calculation above, in order to take away the number of exceptional moves, we require that $\alpha = \mu$, and analogously that $\alpha^- = \mu^-$. By the orthogonality relation, we have that $\langle \chi^\alpha, \chi^\mu \rangle = 1$ and $\langle \chi^\gamma, \chi^\nu \rangle$=1. Hence, we obtain $\sum_{\alpha^-=\mu^-} 1 - 1$, which is the number of ways that a single move can start at α and end with μ, only considering non-exceptional moves. Therefore, we have that

$$\sum_\alpha \mathrm{M'}_k(\lambda,\alpha)\big[\sum_{\alpha^-=\mu^-} 1 - 1\big] = \mathrm{M'}_{k+1}(\lambda,\mu)$$

□

We now use Lemma 4.3 to show that $S_t(n) = M_t(n)$. Let $\mathrm{Des}(S_n)$ be the descent algebra of the symmetric group on n objects. The elements in $\mathrm{Des}(S_n)$ are D_α for $\alpha \subseteq \{1, 2, \ldots, n-1\}$, where $D_\alpha = \sum_{\sigma \in S_n} \sigma$ is the sum over all permutations σ in S_n with

descents in positions α. For example, let $n = 4$ and consider $\alpha = \{1, 3\}$. Then $D_{\{1,3\}}$ is the sum over all permutations of 4 letters with descents in positions 1 and 3. Namely,

$$D_{\{1,3\}} = \mathbf{4231} + \mathbf{4132} + \mathbf{3241} + \mathbf{3142} + \mathbf{2143}$$

In this paper, every element in $\mathrm{Des}(S_n)$ has coefficients belonging to $\mathbb{Q}$ and hence $\mathrm{Des}(S_n)$ can be defined as a subalgebra of the group algebra $\mathbb{Q}S_n$.

Recall that $\sigma_m = (123 \ldots m)$ is the permutation in S_n that *lifts* the card in position m. Let $\Xi = \sum_{m=1}^{n} \sigma_m^{-1}$ and let $\Delta = \Xi - \mathrm{id}_n$. From the definition of D_α above we have that $\Delta = D_{\{1\}}$ and $\Xi = D_{\{1\}} + D_\emptyset$ and hence $\Delta, \Xi \in \mathrm{Des}(S_n)$.

Recall that $\mathrm{Cl}(S_n)$ is the class algebra of S_n. By [4] (Theorem 1.2) or [26] (Theorem 1) let $f : \mathrm{Des}(S_n) \to \mathrm{Cl}(S_n)$ be the epimorphism defined as follows:

$$\Xi \mapsto \pi$$
$$\mathrm{id}_{S_n} \mapsto 1_{S_n}$$
$$\Delta \mapsto \vartheta$$

Define the bilinear form $(-,-)$ on $\mathbb{Q}S_n$ by

$$(g, h) = \begin{cases} 1 & \text{if } g = h^{-1} \\ 0 & \text{otherwise} \end{cases}$$

By [4] the epimorphism defined above is an isometry with respect to the bilinear form on $\mathrm{Des}(S_n)$ and hence we have that $(g, h) = \langle f(g), f(h) \rangle$.

Lemma 4.4. *[5]: If t, $n \in \mathbb{N} \cup \{0\}$ then $S_t(n) = M_t(n)$ and $S'_t(n) = M'_t(n)$*

Recall that in Lemma 4.2 we prove that $B_t(n) = S_t(n)$, hence $B_t(n) = M_t(n)$ follows immediately from Lemma 4.4.

Proof. For $n = 0$, $S_t(0)$ is the number of t sequences of shuffles of zero cards whose product is the identity and $M_t(0)$ is the number of sequences of t moves that start at (0) and finish at (0), therefore for $t \in \mathbb{N}$ we have $S_t(0) = 0 = M_t(0)$ and for $t = 0$ we have $S_0(0) = 1 = M_0(0)$.

Now consider $S_t(n)$ and $M_t(n)$ for $n \geq 1$. From Lemma 4.3, we have $M_t(n) = \langle \chi^{(n)} \pi^t, \chi^{(n)} \rangle$, where $\chi^{(n)}$ is the irreducible character of S_n labeled by the single-part partition (n), hence $\chi^{(n)} = 1_{S_n}$ and we have that $M_t(n) = \langle \chi^{(n)} \pi^t, \chi^{(n)} \rangle = \langle \pi^t, 1_{S_n} \rangle$.

From the remark before the lemma we have that

$$\langle \pi^t, 1_{s_n} \rangle = \langle f(\Xi^t), f(\mathrm{id}_{S_n}) \rangle = (\Xi^t, \mathrm{id}_{S_n})$$

.

and from the definition of $(-,-)$ we have that

$$\begin{aligned}(\Xi^t, \mathrm{id}_{S_n}) &= (\sum_{\sigma\in S_n} g_\sigma \sigma^{-t}, \mathrm{id}_{S_n}) \\ &= \sum_{\sigma\in S_n} g_\sigma(\sigma^{-t}, \mathrm{id}_{S_n}) \\ &= \sum_{\sigma=\mathrm{id}_{S_n}} g_\sigma(\sigma^{-t}, \mathrm{id}_{S_n}) \\ &= [\mathrm{id}_{S_n}]\Xi^t\end{aligned}$$

By definition, we have

$$\begin{aligned}S_t(n) &= |\{1 \le a_1, a_2, \ldots, a_t \le n : \sigma_{a_t} \cdots \sigma_{a_1} = \mathrm{id}_{S_n}\}| \\ &= |\{1 \le a_1, a_2, \ldots, a_t \le n : \sigma_{a_1}^{-1} \cdots \sigma_{a_t}^{-1} = \mathrm{id}_{S_n}\}| \\ &= [\mathrm{id}_{S_n}]\Xi^t\end{aligned}$$

which proves the result for $S_t(n)$.

Recall that $S'_t(n)$ is the number of sequences of t random-to-top shuffles, none of which are the identity permutation, whose product is the identity permutation. From Lemma 4.3, we have $M'_t(\lambda,\mu) = \langle \chi^\lambda \vartheta^t, \chi^\mu\rangle$ and hence

$$\begin{aligned}M'_t(n) &= \langle \chi^{(n)}\vartheta^t, \chi^{(n)}\rangle \\ &= \langle \vartheta^t, 1_{S_n}\rangle \\ &= \langle (\pi - 1_{S_n})^t, 1_{S_n}\rangle \\ &= \langle f(\Delta^t), f(1_{S_n})\rangle \\ &= (\Delta^t, 1_{S_n})\end{aligned}$$

Similar to the calculation of $[\mathrm{id}_{S_n}]\Xi^t$ we have $(\Delta^t, 1_{S_n}) = [\mathrm{id}_{S_n}]\Delta^t$. By the definition of $S'_t(n)$, we have

$$\begin{aligned}S'_t(n) &= |\{1 \le a_1, a_2, \ldots, a_t \le n : \sigma_{a_t} \cdots \sigma_{a_1} = \mathrm{id}_{S_n}, \sigma_{a_i} \ne \mathrm{id}_{S_n}\}| \\ &= |\{1 \le a_1, a_2, \ldots, a_t \le n : \sigma_{a_1}^{-1} \cdots \sigma_{a_t}^{-1} = \mathrm{id}_{S_n}, \sigma_{a_i} \ne \mathrm{id}_{S_n}\}| \\ &= [\mathrm{id}_{S_n}]\Delta^t\end{aligned}$$

which proves the result for $S'_t(n)$. □

4.5 The Mixing Time

In 1986, Diaconis and Aldous [1] gave a proof of the mixing time of the top-to-random shuffle, which is the inverse of the random-to-top shuffle. Since these shuffles are inverses of each other, they have the same mixing times. However, the proof in [1] cannot be used directly for the random-to-top shuffle. In this section we give a detailed proof of the mixing time of the random-to-top shuffle.

A Few Definitions

Before we give the proof of the mixing time of the random-to-top shuffle, we must first introduce a few definitions.

Consider a deck of n cards. Recall, that when a card is lifted, it is placed at the top of the deck, and that one shuffle consists of a single lift. Let $\mathscr{L}_i$ denote the ordered list of lifted cards, in the order in which they appear in the deck in the i^{th} stage. Recall that after the i^{th} shuffle, the deck is in the i^{th} stage. Let $\mathscr{U}_i$ denote the ordered list of cards that have not yet been lifted, in the order in which they appear in the i^{th} stage. Observe, that at the i^{th} stage, the top $|\mathscr{L}_i| = k$ cards have all been lifted at least once in a previous stage. We also observe that k increases between consecutive shuffles by at most one. That is, if $|\mathscr{L}_i| = |\mathscr{L}_j|$ for $i < j$, then a card already in $\mathscr{L}_i$ was lifted at the j^{th} stage. If $|\mathscr{L}_i| < |\mathscr{L}_j|$, then a card in $\mathscr{U}_i$ was lifted at the j^{th} stage. Once a card belongs to $\mathscr{L}_i$, for any i, it can never be removed.

Since k increases by at most 1 between consecutive shuffles, it follows that $|\mathscr{U}|$ decreases by at most one between consecutive shuffles. The list of values $|\mathscr{L}_i|$ from $i = \{1, \ldots, n\}$ is therefore a weakly increasing sequence, and the list of values of $|\mathscr{U}_i|$ from $i = \{1, \ldots, n\}$ is a weakly decreasing sequence.

Since none of the cards in $\mathscr{U}_i$ have been lifted, they will remain in the same order with respect to the other cards in $\mathscr{U}_i$. Since each card that is lifted is chosen uniformly at random, the list or cards $\mathscr{L}_i$ is random, and the order is also random.

For example, consider the deck 12345.

Stage	k	State	$\mathscr{L}_i$	$\mathscr{U}_i$
0	0	12345	$\{\emptyset\}$	$\{1,2,3,4,5\}$
1	1	**2**1345	$\{2\}$	$\{1,3,4,5\}$
2	1	**2**1345	$\{2\}$	$\{1,3,4,5\}$
3	2	**42**135	$\{4,2\}$	$\{1,3,5\}$
4	3	**542**13	$\{5,4,2\}$	$\{1,3\}$
5	3	**452**13	$\{4,5,2\}$	$\{1,3\}$
6	4	**1452**3	$\{1,4,5,2\}$	$\{3\}$
7	5	**31452**	$\{3,1,4,5,2\}$	$\{\emptyset\}$

Expected number of trials until success

Lemma 4.5. *: Let X be a random variable on a set G. Suppose that $X = g$ for $g \in G$ is considered a success with probability p and that $X \neq g$ is considered a failure, with probability $1 - p$. Then the expected number of independent trials until the first success is $\frac{1}{p}$.*

Proof. Let A denote the event that $X = g$. The probability of the event A occurring for the first time on the m^{th} trial in a series of independent trials is $(1-p)^{m-1}p$. Let $E[A]$ denote the expected number of independent trials before the event A occurs for the first time. We can model $E[A]$ by the following infinite series:

$$E[A] = p + 2(1-p)p + 3(1-p)^2p + 4(1-p)^3p + \cdots + m(1-p)^{m-1}p + ... \qquad (22)$$

Factoring p out, we get:

$$E[A] = p \cdot [1 + 2(1-p) + 3(1-p)^2 + 4(1-p)^3 + \cdots + m(1-p)^{m-1} + ...] \tag{23}$$

Multiply both sides by $(1-p)$:

$$(1-p) \cdot E[A] = p \cdot [(1-p) + 2(1-p)^2 + 3(1-p)^3 + 4(1-p)^4 + \cdots + m(1-p)^m + ...] \tag{24}$$

Subtracting (11) from (10) we get:

$$p \cdot E[A] = p \cdot [1 + (1-p) + (1-p)^2 + (1-p)^3 + (1-p)^4 + \cdots + (1-p)^m + ...] \tag{25}$$

Dividing both sides by p:

$$E[A] = [1 + (1-p) + (1-p)^2 + (1-p)^3 + (1-p)^4 + \cdots + (1-p)^m + ...] \tag{26}$$

This is exactly the infinite geometric series with common ratio $(1-p)$. Hence we have:

$$E[A] = \sum_{m=0}^{\infty} (1-p)^m \tag{27}$$

$$= \frac{1}{1-(1-p)} \tag{28}$$

$$= \frac{1}{p} \tag{29}$$

□

Lemma 4.6. *The expected number of random-to-top shuffles before the deck is considered random is $O(n \log n)$.*

Proof. Consider a deck of n cards. Recall that in a sequence of shuffles, a deck of cards is in the initial stage if no card has been lifted. Thus, in the initial stage, every card is in $\mathscr{U}$ and we have $k = 0$. Observe that $k = 0$ only in the initial stage, hence the expected number of shuffles before the first time $k = 1$ is exactly 1. Henceforth, we denote by T_i, the time at which $k = i$ for the first time.

Suppose $k = 1$, and consider T_2. That is, consider the first time that a second card from $\mathscr{U}$ is lifted. When $k = 1$, the probability of lifting a card in $\mathscr{U}$ is $p = \frac{n-1}{n}$, hence the average number of shuffles before $k = 2$ is $\frac{1}{p} = \frac{n}{n-1}$.

Observe that when $k = i$, we have $|\mathscr{U}| = n - i$. It follows that the probability of lifting a card from $\mathscr{U}$ when $k = i$ is $p = \frac{n-i}{n}$, and that when $k = i$, the average number of shuffles before $k = i + 1$ is $\frac{n}{n-i}$.

Recall that at every stage, the set $\mathscr{L}$ is a random set of cards in random order. Thus, when $k = n - 1$, $n - 1$ of the n cards in the deck are in random order, and since $|\mathscr{U}| = 1$ and $\mathscr{U}$ is also a random set of cards, then we have that the whole deck is in random order.

Let A be the time it takes for every card in the deck to be lifted at least once. By linearity of expected value, $E[A]$ can be expressed as the sum of the expected values of k for all possible values of k:

$$\begin{aligned}\sum_{i=1}^{n-1} \mathrm{E}[k=i] &= 1 + \frac{n}{n-1} + \frac{n}{n-2} + \cdots + \frac{n}{2} + \frac{n}{1} \\ &= n \cdot (\frac{1}{n} + \frac{1}{n-1} + \frac{1}{n-2} + \cdots + 2 + 1) \\ &= n \cdot \Big(\log n + \gamma + \frac{1}{2n} + O\Big(\frac{1}{n^2}\Big)\Big) \\ &= O(n \log n)\end{aligned}$$

□

Consider the example above with a deck of 5 cards: 12345

Stage	k	State	$\mathscr{L}_i$	$\mathscr{U}_i$	T_i
0	0	12345	$\{\emptyset\}$	$\{1,2,3,4,5\}$	
1	1	**2**1345	$\{2\}$	$\{1,3,4,5\}$	T_1
2	1	**2**1345	$\{2\}$	$\{1,3,4,5\}$	
3	2	**42**135	$\{4,2\}$	$\{1,3,5\}$	T_2
4	3	**542**13	$\{5,4,2\}$	$\{1,3\}$	T_3
5	3	**452**13	$\{4,5,2\}$	$\{1,3\}$	
6	4	**1452**3	$\{1,4,5,2\}$	$\{3\}$	T
7	5	**31452**	$\{3,1,4,5,2\}$	$\{\emptyset\}$	

where T is the time at which the deck is randomized and T_i is the first time the i^{th} card is lifted.

Before we give the rigorous proof, we must first define a probability distribution Q for the random-to-top shuffle:

$$Q(\pi) = \begin{cases} 1/n & \text{if } \pi \text{ is the identity or one of the cycles } (1\ 2 \cdots\ i),\ 1 \le i \le n \\ 0 & \text{otherwise} \end{cases} \tag{30}$$

Note that in the proof of the mixing time for the riffle shuffle, they use a very particular result regarding a-shuffles and b-shuffles to obtain the probability of the m repeated riffle shuffles resulting in π. Since there is no such fact for random-to-top shuffles, we must represent repeated random-to-top shuffles using random walks. Hence we look at the probability distribution Q^{*m}.

To measure the difference between two probability distributions Q_1 and Q_2 we use the same total variation distance as in [3]:

$$||Q_1 - Q_2|| = \frac{1}{2} \sum_{\pi} |Q_1(\pi) - Q_2(\pi)|$$

Recall that $Q^{*m}(g) \to U(g) = 1/|G|$ as $m \to \infty$. From this, we have that

$$||Q^{*m} - U|| \to 0 \text{ as } m \to \infty$$

Theorem 4.7. *[1] Let $d(m) = ||Q^{*m} - U||$. For the random-to-top shuffle, we have that:*

$$d(n \log n + cn) \leq e^{-c} \qquad \text{for all } c \geq 0, n \geq 2 \tag{31}$$

$$d(n \log n - c_n n) \to 1 \text{ as } n \to \infty \qquad \text{for all } c_n \to \infty \tag{32}$$

Before we give the proof of Theorem 4.7 we must first give a few definitions and prove some intermediate results.

Definition 4.8. *Let G be a finite group, and let $G^{\mathbb{N}}$ be the set of G-valued infinite sequences $\mathbf{g} = (g_1, g_2, \dots)$. A **stopping rule** $\hat{T}$ is a function $\hat{T} : G^{\infty} \to \{1, 2, 3, \dots; \infty\}$ such that if $\hat{T}(\mathbf{g}) = j$, then $\hat{T}(\mathbf{h}) = j$ for all $\mathbf{h}$ with $h_i = g_i$, $i \leq j$. That is, $\hat{T}$ tells us what to do at time n, given a sequence $\mathbf{g} = (g_1, g_2, \dots)$.*

Definition 4.9. *Let Q be a distribution on G, and let (X_m) be the associated random walk. Given a stopping rule $\hat{T}$, the random time $T = \hat{T}(X_1, X_2, \dots)$ is a **stopping time**.*

That is, the stopping time T is the stopping rule applied to a sequence of random variables $(X_1, X_2, \dots, X_n, \dots)$ and is a numerical value (for example, a particular number of shuffles). The decision to stop at n is only dependent on the first n variables $(X_1, \dots, X_n, \dots)$ in the sequence.

Definition 4.10. *Call T a **strong uniform time** (for U) if for each $m < \infty$, $P(T = m, X_m = g)$ does not depend on g for $g \in G$*

Remark. *(29) is equivalent to the following two statements:*
(a) $P(X_m = g | T = m) = 1/|G|, g \in G$
(b) $P(X_m = g | T \leq m) = 1/|G|, g \in G$

From Lemma 4.6 above, the stopping time $T = T_n$ for the random-to-top shuffle is the first time that $|\mathscr{U}| = 1$. That is, T_n is the first time that the bottom card in the deck is the last (and only) card yet to be lifted. It's important to note that $T_n = T_{n-1}$, since, when $|\mathscr{U}| = 1$, the deck is in random order and when $|\mathscr{U}| = 0$, the deck is also in random order, but in no "more" of a random order. Hereafter we use $T = T_{n-1}$.

Lemma 4.11. *[1] Let Q be a probability distribution on a finite group G. Let T be a strong uniform time for Q. Then*

$$d(m) = ||Q^{*m} - U|| \leq P(T > m) \text{ for all } m \geq 0$$

Proof. For any $A \subset G$

$$\begin{aligned} Q^{*m}(A) &= P(X_m \in A) \\ &= \sum_{j \leq m} P(X_m \in A, T = j) + P(X_m \in A, T > m) \\ &= \sum_{j \leq m} U(A) P(T = j) + P(X_m \in A | T > m) P(T > m) \\ &= U(A) + [P(X_m \in A | T > m) - U(A)] P(T > m) \end{aligned}$$

and so

$$Q^{*m}(A) - U(A) = [P(X_m \in A | T > m) - U(A)]P(T > m)$$

and therefore,

$$|Q^{*m}(A) - U(A)| \leq P(T > m)$$

□

Lemma 4.12. *"The Coupon Collector's Problem"*

Given n coupons, how many coupons do you need to draw, with replacement, before having drawn each coupon at least once?

Let V denote the number of draws required until each coupon has been drawn at least once. Then

$$P(V > n \log n + cn) \leq e^{-c}$$

for all $c \geq 0$, $n \geq 1$.

Proof. Let V_i be the number of draws until the i^{th} distinct coupon is drawn. Then $P(V_i) = \frac{(n-(i-1))}{n}$, and we have

$$\begin{aligned}
E(V) &= E(V_1) + \cdots + E(V_n) \\
&= \frac{1}{P(V_1)} + \frac{1}{P(V_2)} + \cdots + \frac{1}{P(V_n)} \\
&= \frac{n}{n} + \frac{n}{n-1} + \cdots + \frac{n}{2} + \frac{n}{1} \\
&= n \cdot \Big(\frac{1}{n} + \frac{1}{n-1} + \cdots + \frac{1}{2} + \frac{1}{1}\Big) \\
&= n \cdot \Big(\log n + \gamma + \frac{1}{2n} + O\Big(\frac{1}{n^2}\Big)\Big) \\
&= n \log n + n\gamma + \frac{1}{2} + O\Big(\frac{1}{n}\Big)
\end{aligned}$$

where γ is the Euler-Mascheroni constant.

Now, let Z_i^m be the event that coupon i is not drawn in the first m draws. Then we have

$$P(Z_i^m) = \Big(1 - \frac{1}{n}\Big)^m \leq e^{\frac{-m}{n}}$$

for each i. From above, setting $m = n \log n + cn$ for $c \geq 0$ we have

$$P(V > m) \leq \sum_i P(Z_i) = n\Big(1 - \frac{1}{n}\Big)^m \leq n \cdot e^{\frac{-m}{n}} = e^{-c}$$

□

Proof of Theorem 4.7. To prove the result from (a), we show that T has the same probability distribution as V from Lemma 4.12. Recall that T_i is the number of shuffles until the i^{th} card is lifted for the first time, and consider the following way we can write T:

$$T = (T_{n-1} - T_{n-2}) + (T_{n-2} - T_{n-3}) + \cdots + (T_2 - T_1) + T_1 \tag{33}$$

When exactly i cards have been lifted at least once, the chance that the next card will be lifted for the first time is $\frac{n-i}{n}$. Therefore, $T_i - T_{i-1}$ has geometric distribution

$$P(T_i - T_{i-1} = j) = \frac{n-i}{n}\Big(1 - \frac{n-i}{n}\Big)^{j-1} \tag{34}$$

for $j \geq 1$. Similarly, we can write V as

$$V = (V - V_{n-1}) + (V_{n-1} - V_{n-2}) + \cdots + (V_2 - V_1) + V_1 \tag{35}$$

where, from before, V_i is the number of draws until i distinct coupons have been drawn at least once. After i distinct coupons have already been drawn at least once, the probability that the next coupon drawn will be drawn for the first time is $\frac{n-i}{n}$, so the probability distribution of $V_i - V_{i-1}$ is

$$P(V_i - V_{i-1} = j) = \frac{n-i}{n}\Big(1 - \frac{n-i}{n}\Big)^{j-1} \tag{36}$$

for $j \geq 1$. Since T and V have the same distribution, part (a) of Theorem 4.7 follows directly from Lemma 4.12.

To establish the lower bound for part (b) of Theorem 4.7, we chose a set A for which $|Q^{*m}(A) - U(A)|$ is large, and use

$$d(m) = ||Q^{*m} - U|| \geq |Q^{*m}(A) - U(A)|$$

Suppose at time $m = n\log n - c_n n$, all but j cards have been lifted at least once. After $n - j$ cards have been lifted, we have $|\mathscr{L}| = n - j$ and $|\mathscr{U}| = j$ and the bottom j cards will remain in their original relative order. Let A_j denote the set of deck configurations of n cards where j cards (chosen uniformly at random) remain in their original relative order in the last j positions of the deck. Then $U(A_j) = 1/(n-j)!$. We claim that

$$Q^{*m}(A_j) \to 1 \text{ as } n \to \infty \tag{37}$$

We first note that $Q^{*m}(A_j) \geq P(T_{n-1} - T_{j-1} > m)$, where $T_{n-1} - T_{j-1}$ is the time that $n - j$ cards have all been lifted for the first time. That is, the first time that $|\mathscr{U}| = j$. If $n - j$ cards have not been lifted for the first time by time m, then the bottom j cards must remain in their original relative order. Therefore, it suffices to show

$$P(T_{n-1} - T_{j-1} \leq m) \to 0 \text{ as } n \to 0 \tag{38}$$

To do this, we use Chebyshev's inequality:

$$P\big(|Z - EZ| \geq a\big) \leq \frac{\text{var}(Z)}{a^2}$$

where Z is a random variable and $a > 0$ is any real number.

From (30) we know $E(T_i - T_{i-1}) = \frac{n}{n-i}$ and $\text{var}(T_i - T_{i-1}) = \left(\frac{n}{n-i}\right)^2\left(1 - \frac{n-1}{n}\right)$. Hence, from (29) we have,

$$E(T_{n-1} - T_j) = \sum_{i=j}^{n-2} \frac{n}{n-i} = n \log n + O(n)$$

$$\text{var}(T_{n-1} - T_j) = \sum_{i=j}^{n-2} \left(\frac{n}{n-i}\right)^2\left(1 - \frac{n-i}{n}\right) = O(n^2)$$

Applying Chebyshev's inequality to $Z = T_{n-1} - T_{j-1}$ we get the desired result. □

5 Concluding Remarks

From the results presented above, we can see how mathematically "involved" just two types of card shuffles are. It is also important to note that this paper does not include every results regarding these two shuffles (see [13], [27], [28]). Furthermore, these are not the only types of shuffles which have interesting results. Many other shuffling techniques, such as the Faro shuffle and the top-to-random, also have many interesting mathematical results [10] [19], and are just as mathematically "involved" as the ones presented in this paper.

References

[1] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986.

[2] Sami Assaf, Persi Diaconis, Kannan Soundararajan, et al. A rule of thumb for riffle shuffling. *The Annals of Applied Probability*, 21(3):843–875, 2011.

[3] Dave Bayer, Persi Diaconis, et al. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313, 1992.

[4] D. Blessenohl and M. Schocker. *Noncommutative Character Theory of the Symmetric Group*. Imperial College Press, 2005.

[5] John R Britnell and Mark Wildon. Bell numbers, partition moves and the eigenvalues of the random-to-top shuffle in dynkin types a, b and d. *Journal of Combinatorial Theory, Series A*, 148:116–144, 2017.

[6] Mark Conger and D Viswanath. Normal approximations for descents and inversions of permutations of multisets. *Journal of Theoretical Probability*, 20(2):309–325, 2007.

[7] Persi Diaconis. The cutoff phenomenon in finite markov chains. *Proceedings of the National Academy of Sciences*, 93(4):1659–1664, 1996.

[8] Persi Diaconis, James Allen Fill, and Jim Pitman. Analysis of top to random shuffles. *Combinatorics, probability and computing*, 1(2):135–155, 1992.

[9] Persi Diaconis and Jason Fulman. Carries, shuffling, and symmetric functions. *Advances in Applied Mathematics*, 43(2):176–196, 2009.

[10] Persi Diaconis, RL Graham, and William M Kantor. The mathematics of perfect shuffles. *Advances in applied mathematics*, 4(2):175–196, 1983.

[11] Jason Fulman. A card shuffling analysis of deformations of the plancherel measure of the symmetric group. *The Electronic Journal of Combinatorics*, 11(1):21, 2004.

[12] John M Holte. Carries, combinatorics, and an amazing matrix. *The American Mathematical Monthly*, 104(2):138–149, 1997.

[13] Johan Jonasson et al. Biased random-to-top shuffling. *The Annals of Applied Probability*, 16(2):1034–1058, 2006.

[14] Gina Kolata. In shuffling cards, 7 is winning number. *The New York Times*, 1990.

[15] M. Lothaire. Combinatorics on words. *Cambridge University Press*, 1997.

[16] Brad Mann. How many times should you shuffle a deck of cards? *Harvard University.*

[17] Andrey Andreyevich Markov. Extension of the law of large numbers to dependent quantities. *Izv. Fiz.-Matem. Obsch. Kazan Univ.(2nd Ser)*, 15:135–156, 1906.

[18] S Brent Morris. The basic mathematics of the faro shuffle. *Pi Mu Epsilon Journal*, 6(2):85–92, 1975.

[19] CY Pang. Card-shuffling via convolutions of projections on combinatorial hopf algebras. *arXiv preprint arXiv:1503.08368*, 2015.

[20] CY Pang. Markov chains from descent operators on combinatorial hopf algebras. *arXiv preprint arXiv:1609.04312*, 2016.

[21] Robin Pemantle. Randomization time for the overhand shuffle. *Journal of Theoretical Probability*, 2(1):37–49, 1989.

[22] Jim Reeds. Unpublished manuscript. 1981.

[23] JU Reyes. Random walk, semi-direct products, and card shuffling. 2003.

[24] Will Roya. The faro: A card shuffle and a card game. *playingcarddecks.com*, March 19, 2019.

[25] Will Roya. The history of playing cards: The evolution of the modern deck. *playingcarddecks.com*, October 16, 2018.

[26] Louis Solomon. A mackey formula in the group ring of a coxeter group. *Journal of Algebra*, 41(2):255–264, 1976.

[27] Richard P Stanley. Generalized riffle shuffles and quasisymmetric functions. *Annals of Combinatorics*, 5(3-4):479–491, 2001.

[28] Dudley Stark, Ayalvadi Ganesh, and Neil O'connell. Information loss in riffle shuffling. *Combinatorics, Probability and Computing*, 11(1):79–95, 2002.

[29] David Bruce Wilson et al. Mixing times of lozenge tiling and card shuffling markov chains. *The Annals of Applied Probability*, 14(1):274–325, 2004.