

Camillia Smith Barnes
cammie*AT*math.harvard.edu

RESEARCH SUMMARY AND PROPOSAL

Mathematicians have recently studied several notions of ‘shuffling’, including shuffling of a deck of cards (see [Aldous & Diaconis (1986)] [Bayer & Diaconis (1992)] [Diaconis (1988)] [Diaconis (2002)] [Diaconis et al. (1983)] [Trefethen & Trefethen (2002)] [van Zuylen & Schalekamp (2004)]), ‘shuffling’ algorithms, such as the Fisher-Yates shuffle (also known as the Knuth shuffle) that generate random permutations of a finite set (see [Fisher & Yates (1948)] [Knuth (1973)] [Knuth (1998)]), and the perfect shuffle permutation (see [Diaconis et al. (1983)] [Ellis et al. (2000)] [Mevedoff & Morrison (1987)]).

We shall be interested in *shuffles of words*, where a *word* is defined to be a finite string of elements (known as *letters*) of a given set (known as an *alphabet*); in general repetitions of letters are allowed. We define the *length* of a word $u = a_1 \dots a_m$ to be $\mathbf{l}(u) = m$ and the *support* of u to be $\text{supp}(u) = \{a_1, \dots, a_m\}$. A *subword* x of a word u is defined to be a word obtained by crossing out a (possibly empty) subset of the letters of u .

For example, for the alphabet $\mathcal{A} = \{1, 2, 3, 5, 7\}$, the words $u = 25372$ and $v = 123$ have support $\text{supp}(u) = \{2, 3, 5, 7\}$ and $\text{supp}(v) = \{1, 2, 3\}$, and length $\mathbf{l}(u) = 5$ and $\mathbf{l}(v) = 3$. Two subwords of u are 232 and 537.

SHUFFLES OF WORDS

Given two words $u = a_1 a_2 \dots a_m$ and $v = b_1 b_2 \dots b_n$ in some alphabet \mathcal{A} , we obtain a *shuffle* of u and v by concatenating u and v to get

$$c_1 c_2 \dots c_{m+n} = a_1 a_2 \dots a_m b_1 b_2 \dots b_n$$

and then permuting letters in such a way to achieve

$$w = c_{\rho(1)} c_{\rho(2)} \dots c_{\rho(m+n)},$$

for some permutation $\rho \in \mathfrak{S}_{m+n}$ on $m+n$ letters satisfying the order-preserving conditions

$$(1) \quad \rho^{-1}(1) < \rho^{-1}(2) < \dots < \rho^{-1}(m)$$

and

$$(2) \quad \rho^{-1}(m+1) < \rho^{-1}(m+2) < \dots < \rho^{-1}(m+n).$$

In other words, we intersperse the letters of u with those of v to get w in such a way that the subword obtained by restricting w to the letters that came from u is simply u itself (and similarly for the subword obtained by restriction to the letters of v). Two different shuffles of the words 1234 and 5678 are, for instance, 15236784 and 51236748.

In the literature, the shuffle w is sometimes denoted by $u \sqcup\sqcup v$ (see [Hersh (2002)]). Since $\sqcup\sqcup$ depends on a choice of ρ , however, and since $u \sqcup\sqcup v$ sometimes denotes instead the *shuffle product* of u and v in the *shuffle algebra* (see [Reutenauer (1993)], page 24), we will use the notation $\sqcup\sqcup_\rho$ to avoid ambiguity. We define

$$\mathfrak{sh}(u, v) = \{u \sqcup\sqcup_\rho v \mid \rho \in \mathfrak{S}_{m+n} \text{ satisfies (1) and (2)}\}$$

to be the set of all shuffles of u with v . For ease of reference, we shall also set

$$\mathfrak{S}_{m,n} = \{\rho \in \mathfrak{S}_{m+n} \mid \rho \text{ satisfies (1) and (2)}\}.$$

The *shuffle algebra* \mathcal{A} (see [Crossley (2006)] [Ehrenborg (1996)] [Reutenauer (1993)]), a commutative Hopf algebra structure on the free \mathbb{Z} -module generated by finite words in a given alphabet \mathcal{A} , has as multiplication the *shuffle product* Δ , which is given by

$$(3) \quad \Delta(u \otimes v) = \sum_{w \in \mathfrak{sh}(u,v)} \mu_w w$$

for words u and v , where

$$\mu_w = \#\{\rho \in \mathfrak{S}_{\mathbf{l}(u), \mathbf{l}(v)} \mid u \sqcup\sqcup_\rho v = w\}$$

is the multiplicity of w . The shuffle algebra has applications, for instance, in number theory: the multiplication of two multiple zeta values can be expressed as the sum of other multiple zeta values via a shuffle relation or a quasi-shuffle (stuffle) relation (see [Guo & Xie (2008)] [Ihara et al. (2006)]).

We can define, analogously, a shuffle of k words (or k -shuffle) to be a permutation of the concatenation of k words (with lengths n_1, n_2, \dots, n_k) in such a way that the inverse permutation preserves order when restricted to the index subsets $[n_1], [n_1 + 1, n_1 + n_2], \dots, [n_1 + n_2 + \dots + n_{k-1} + 1, n_1 + n_2 + \dots + n_k]$, where the interval notation $[n_1 + 1, n_1 + n_2]$ denotes the set of integers from $n_1 + 1$ to $n_1 + n_2$. A k -shuffle is also sometimes referred to as an α -shuffle, where $\alpha = (n_1, n_2, \dots, n_k) \in \mathbb{P}^k$ is any k -tuple of positive integers. (But we reserve the notation $\sqcup\!\!\!\sqcup_\rho$ for 2-shuffles, as they are the main focus of my research.)

Shuffles of words arise in several contexts. For instance, given a subset

$$T = \{s_1, s_2, \dots, s_{k-1}\} \subseteq [n - 1],$$

it can be seen that a permutation $\tau \in \mathfrak{S}_n$ is a k -shuffle of the sets $[s_1], [s_1 + 1, s_2], \dots, [s_{k-2} + 1, s_{k-1}], [s_{k-1} + 1, n]$ if and only if the descent set $D(\tau^{-1})$ of the inverse permutation is a subset of T (see [Stanley (1997)], page 70). Shuffles appear in the representation theory of finite groups; the left cosets of the Young Subgroup $\mathfrak{S}_{\alpha_1} \times \mathfrak{S}_{\alpha_2} \times \dots \times \mathfrak{S}_{\alpha_k}$ in the Symmetric Group \mathfrak{S}_n (where $n = \sum_{j=1}^k \alpha_j$) correspond exactly to the unique α -shuffles associated with $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_k)$ (see [Stanley (1999)], page 351).

Shuffles play a role in the multiplication of fundamental quasisymmetric functions L_γ ; in fact, if $u \in \mathfrak{S}_m$ and $v \in \mathfrak{S}_{[m+1, m+n]}$, then

$$L_{\text{co}(u)} L_{\text{co}(v)} = \sum_{w \in \text{sh}(u, v)} L_{\text{co}(w)},$$

where $\text{co}(u)$ denotes the composition associated with the descent set $D(u)$ (see [Stanley (1999)], page 482, exercise 7.93). Moreover, shuffle posets on the words u and v can be defined by considering the set of subwords of all possible shuffles of u with v , taking u as the minimal element, v as the maximal element, and defining the cover relation to be $x < y$ if y can be obtained from x either by deleting one letter of u or inserting one letter of v . Greene [Greene (1988)] introduced shuffle posets, and Doran [Doran (2002)] and Hersh [Hersh (2002)] generalized them (see also [Ehrenborg (1996)] [Simion & Stanley (1999)]).

THE MAIN QUESTION

A natural question to ask is how to enumerate the *distinct* shuffles of words.

Question 1. *Given words u and v , how many distinct shuffles are there of u with v ?*

Assuming m and n to be the lengths of u and v , respectively, note that if $\text{supp}(u) \cap \text{supp}(v) = \emptyset$, then there are $\binom{m+n}{m}$ distinct shuffles (all shuffles are distinct).

Observation 2. *For any given words u and v , we can define an equivalence relation on $\mathfrak{S}_{\mathbf{I}(u), \mathbf{I}(v)}$ by $\rho \sim \tau$ if $u \sqcup\!\!\!\sqcup_\rho v = u \sqcup\!\!\!\sqcup_\tau v$.*

The equivalence relation is nontrivial only when $\text{supp}(u) \cap \text{supp}(v) \neq \emptyset$. So one could reformulate Question 1 to ask how many different equivalence classes are induced on $\mathfrak{S}_{\mathbf{I}(u), \mathbf{I}(v)}$ by shuffling a given u with a given v .

In various applications of shuffles, the supports of the words are usually assumed to be disjoint, but I investigate the consequences of discarding this assumption while seeking an answer to Question 1.

I resolve this question for the important case where the words u and v are assumed to be permutations on the letters $\{1, 2, 3, \dots, m\}$ and $\{1, 2, 3, \dots, n\}$, respectively. My answer is given by the following theorem, for which we shall give details in the section entitled ‘‘The Main Theorem’’.

Theorem 3. *The number of distinct shuffles of a permutation $\alpha \in \mathfrak{S}_m$ with a permutation $\beta \in \mathfrak{S}_n$, with $m \leq n$, is given by the following formula:*

$$(4) \quad \#\text{sh}(\alpha, \beta) = \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \sum_{\mathbf{a}=\{0=a_0 < a_1 < \dots < a_{2k} < a_{2k+1}=m+1\}} (-1)^{h(\mathbf{a})} F_\sigma(\mathbf{a}),$$

where $\sigma = \bar{\alpha}^{-1} \circ \beta$, $\bar{\alpha} \in \mathfrak{S}_n$ is the natural extension of α , and $F_\sigma(\mathbf{a})$ is a product of determinants which enumerate the shuffles on a ‘local’ level.

For an explanation of the notation used and a description of the determinants involved, see the section below on the Main Theorem.

ENUMERATION OF THE DISTINCT SHUFFLES OF PERMUTATIONS

We shall start by enumerating shuffles of the identity permutation with itself.

Proposition 4. *The number of distinct shuffles of the identity permutation on n letters with itself is the n^{th} Catalan number C_n , that is*

$$(5) \quad \#\mathfrak{sh}(\text{id}_n, \text{id}_n) = \frac{1}{n+1} \binom{2n}{n}.$$

Proof. A straightforward proof entails showing that set of shuffles of id_n with itself corresponds bijectively with the set of ballot sequences of length $2n$ (which is known to have cardinality C_n). For a given $w \in \mathfrak{sh}(\text{id}_n, \text{id}_n)$, simply substitute a 1 for the first occurrence of each integer between 1 and n , and a -1 for the second occurrence to get a ballot sequence of length $2n$ (that is, a sequence of n ones and n minus ones whose partial sums are all nonnegative). \square

It is possible to show the following formula for the number of distinct shuffles of the identity in two different lengths.

Proposition 5. *For $m \neq n$, the number of distinct shuffles of the identity permutation on m letters with the identity permutation on n letters is given by*

$$(6) \quad \#\mathfrak{sh}(\text{id}_m, \text{id}_n) = \sum_{r=0}^{\lfloor \frac{n-m}{2} \rfloor} (-1)^r \binom{n-m-r}{r} C_{n-r}.$$

More generally, for $m \leq n$ and any $\alpha \in \mathfrak{S}_m$ and $\beta \in \mathfrak{S}_n$, it can be assumed without loss of generality that $\alpha = \text{id}_m$, due to the following fact.

Fact 6. *For any $m \leq n$ and any $\alpha \in \mathfrak{S}_m$, $\beta \in \mathfrak{S}_n$, we have*

$$(7) \quad \#\mathfrak{sh}(\alpha, \beta) = \#\mathfrak{sh}(\text{id}_m, (\bar{\alpha})^{-1} \circ \beta),$$

where $\bar{\alpha} \in \mathfrak{S}_n$ is the natural extension of α to a permutation on n letters.

Here we are simply reordering the alphabet $\mathcal{A} = [m]$ so that α now behaves like the identity permutation id_m on the reordered alphabet. It is also easy to note that $\#\mathfrak{sh}$ is symmetric: $\#\mathfrak{sh}(u, v) = \#\mathfrak{sh}(v, u)$ is true for any words u and v (they need not be permutations).

Now let the reverse permutation word $n, n-1, \dots, 2, 1$ be denoted by rev_n . The following result can be shown via a bijective proof.

Proposition 7.

$$(8) \quad \#\mathfrak{sh}(\text{id}_m, \text{rev}_n) = \binom{m+n}{m} - \binom{m+n-2}{m-1}.$$

THE MAIN THEOREM

Let us now enumerate the number of shuffles of the identity on m letters with any permutation $\sigma \in \mathfrak{S}_n$ (throughout, we shall assume without loss of generality that $m \leq n$). We shall first provide some terminology and motivation and then state the Main Theorem.

Let us call a subword x obtained from any word u *consecutive* if the letters of x appear consecutively in u . For instance, 364 is a consecutive subword of 136425. We call a shuffle $w \in \mathfrak{sh}(\text{id}_n, \text{id}_n)$ *indecomposable* if there is no consecutive subword w' of w such that $w' \in \mathfrak{sh}(\text{id}_k, \text{id}_k)$ for some $1 \leq k < n$. For ease of notation, let

$$\text{indc}(x) = \{\text{indecomposable shuffles of } x \text{ with itself}\}.$$

Observe that, when a shuffle w has multiplicity $\mu_w > 1$, this occurs because some consecutive subword x of σ is in fact a string of consecutive elements in the alphabet of w ; we call such a subword of σ an *embedded*

identity subword. On the local level we then have, embedded in w , a shuffle of the identity permutation on a consecutive subset of the intersection of the given alphabets with itself. That is,

$$w = \cdots * (x \sqcup_{\eta} x) * \cdots$$

for some $\eta \in \mathfrak{S}_{\mathbf{I}(x), \mathbf{I}(x)}$, where $*$ denotes concatenation. We shall denote the set of embedded identity subwords of σ as

$$\text{idsub}(\sigma) = \{\text{embedded identity subwords of } \sigma\}.$$

If id_4 is shuffled with 52341, for example, we can obtain the shuffle

$$512342341 \in \{51 * (234 \sqcup_{\eta} 234) * 1 \mid \eta \in \mathfrak{S}_{3,3}\},$$

which has multiplicity 2 because the local shuffle $234234 \in \mathfrak{sh}(234, 234)$ is indecomposable and can be obtained exactly two ways, whereas there are no additional ways of obtaining the global shuffle $512342341 \in \mathfrak{sh}(\text{id}_4, 52341)$.

We say that a set $X = \{x_1, \dots, x_r\}$ of embedded identity subwords of a permutation is *compatible* if the x_i have pairwise disjoint supports and if there exists some shuffle $w \in \mathfrak{sh}(\text{id}_m, \sigma)$ in which each of the x_i is locally shuffled with itself. For instance, $\{23, 45\}$ is a set of compatible embedded identity subwords of 23145 because in the shuffle 1232314455 $\in \mathfrak{sh}(\text{id}_5, 23145)$ both 23 and 45 are locally shuffled with themselves.

Given a permutation word u and a compatible set $X = \{x_1, \dots, x_j\}$ of embedded identity subwords of u , note that u is the concatenation $u = g_0 * x_1 * g_1 * \cdots * x_j * g_j$ for some consecutive subwords g_0, g_1, \dots, g_j of u whose supports are pairwise disjoint. We say that the g_i are the subwords of u *cut out* by the set X .

For instance, for the permutation word 52341, the set $\{23, 4\}$ cuts out the subwords 5, \square , and 1 (where \square denotes the empty word).

Proposition 8. *For $\sigma \in \mathfrak{S}_n$ and any $w \in \mathfrak{sh}(\text{id}_m, \sigma)$, we have $\mu_w = 2^t$ for some integer $t \geq 0$, where t is the maximal number of compatible embedded identity permutation subwords in σ that are locally shuffled with themselves in w .*

To illustrate this statement, we can see that for $311223 \in \mathfrak{sh}(\text{id}_3, 312)$, we have $\mu(311223) = 4$ and $t = 2$. The embedded identity subwords that are locally shuffled with themselves in 311223 are 1, 2, and 12; but $\{1, 2\}$ is the largest set of such subwords that is compatible. In general, we shall call the integer $t = \text{dup}(w)$ the *number of sites of duplication* in w . Moreover, we shall set $N_t^\sigma = \#\{w \in \mathfrak{sh}(\text{id}_m, \sigma) \mid \text{dup}(w) = t\}$.

We can actually enumerate $\#\mathfrak{sh}(\text{id}_m, \sigma)$ by applying the Inclusion-Exclusion principle. First we take the total number of shuffles counted with multiplicity, and then alternately subtract and add the cardinalities of certain subsets counted with multiplicity until we arrive at a count of the total number of shuffles without multiplicity.

Indeed,

$$\#\mathfrak{sh}(\text{id}_m, \sigma) = \binom{m+n}{m} + \sum_{j=1}^m (-1)^j T_j^\sigma,$$

where $T_j^\sigma = \sum_{t=j}^m \binom{t}{j} 2^{t-j} N_t^\sigma$.

Observation 9. *T_j^σ is the number of (not necessarily distinct shuffles) in $\mathfrak{sh}(\text{id}_m, \sigma)$ with j or more sites of duplication, enumerated by choosing a j -element subset $X = \{x_1, \dots, x_j\}$ of compatible embedded identity permutation subwords of σ and assuming, in turn, that each element $x_i \in X$ is shuffled locally and indecomposably with itself, then counting with multiplicity all local shuffles of each subword of id_m cut out by X with the corresponding subword of σ also cut out by X .*

That is, T_j^σ can be computed as

$$(9) \quad T_j^\sigma = \sum_{\text{compatible } \{x_1, \dots, x_j\} \subseteq \text{idsub}(\sigma)} \binom{\mathbf{I}(f_0) + \mathbf{I}(g_0)}{\mathbf{I}(f_0)} \cdot \#\text{indc}(x_1) \cdot \binom{\mathbf{I}(f_1) + \mathbf{I}(g_1)}{\mathbf{I}(f_1)} \cdots \cdot \#\text{indc}(x_j) \cdot \binom{\mathbf{I}(f_j) + \mathbf{I}(g_j)}{\mathbf{I}(f_j)},$$

where the f_i and g_i are the subwords of id_m and of σ , respectively, that are cut out by the set $\{x_1, \dots, x_j\}$. Recall that the number of local shuffles of f_i with g_i counted with multiplicity is $\binom{\mathbf{I}(f_i) + \mathbf{I}(g_i)}{\mathbf{I}(f_i)}$.

In the example of $\mathfrak{sh}(\text{id}_3, 312)$, we can compute

$$T_1^{312} = \binom{1}{0} \cdot C_0 \cdot \binom{3}{2} + \binom{3}{1} \cdot C_0 \cdot \binom{1}{1} + \binom{2}{2} \cdot C_0 \cdot \binom{2}{0} + \binom{1}{0} \cdot C_1 \cdot \binom{1}{1} = 8$$

because we can fix first double 1's to count shuffles of the form $(\square \sqcup_{\rho_1} 3) * 11 * (23 \sqcup_{\rho_2} 2)$, then fix double 2's to count those of the form $(1 \sqcup_{\rho_3} 31) * 22 * (3 \sqcup_{\rho_4} \square)$, next, fix double 3's to count shuffles of the form $(12 \sqcup_{\rho_5} \square) * 33 * (\square \sqcup_{\rho_6} 12)$, and lastly, fix the unique indecomposable shuffle of 12 with itself to count those of the form $(\square \sqcup_{\rho_7} 3) * 1212 * (3 \sqcup_{\rho_8} \square)$. Similarly,

$$T_2^{312} = \binom{1}{0} \cdot C_0 \cdot \binom{0}{0} \cdot C_0 \cdot \binom{1}{1} = 1,$$

as we can see by counting shuffles of the form $(\square \sqcup_{\rho_9} 3) * 11 * (\square \sqcup_{\rho_{10}} \square) * 22 * (3 \sqcup_{\rho_{11}} \square)$, whereas $T_3^{312} = 0$ because there is no compatible 3-subset of embedded identity permutation subwords, and so

$$\#\mathfrak{sh}(\text{id}_3, 312) = \binom{3+3}{3} - 8 + 1 - 0 = 13.$$

We will use the notation $z_{i,j}^\sigma$ to denote the number of local shuffles counted with multiplicity of the subword a occurring between (and not including) the letters $i < j$ in id_m with the subword b occurring between the letters $i < j$ in σ . That is, if such words a and b exist, then we have $z_{i,j}^\sigma = \binom{\mathbf{1}(a)+\mathbf{1}(b)}{\mathbf{1}(a)}$; otherwise, $z_{i,j}^\sigma = 0$. For example, $z_{0,2}^{312} = \binom{3}{1} = 3$, $z_{1,2}^{312} = \binom{0}{0} = 1$, and $z_{1,3}^{312} = 0$.

We use $z_{i,j}^\sigma$ to construct a square matrix with all ones on the subdiagonal and all zeros below the subdiagonal. For entries on or above the diagonal, $z_{i,j}^\sigma$ keeps track of whether or not i and j are inverted in σ , and if they are not inverted, $z_{i,j}^\sigma$ takes on the value of the total number of possible ways of shuffling the letters between paired occurrences of i and j , including any repeated shuffles.

By defining a matrix $Z_{c,d}^\sigma = [z_{i,j}^\sigma]_{c \leq i \leq d-1, c+1 \leq j \leq d}$ below and taking its determinant, we are taking an alternating sum that systematically looks for compatible sets of letters (that is, compatible length 1 embedded identity subwords of σ) that occur between the letters c and d (not including c and d themselves). When the set of letters, say $\{b_1, b_2, \dots, b_q\}$, is compatible, then we get a nonzero term of absolute value $z_{c,b_1}^\sigma \cdot z_{b_1,b_2}^\sigma \cdots z_{b_q,m+1}^\sigma$.

For $1 \leq e < f \leq m$, if the word $e, e+1, \dots, f$ is a simultaneous consecutive subword for id_m and σ , we will say that $\theta^\sigma(e, f)$ denotes the number of indecomposable local shuffles of the word $e, e+1, \dots, f$ with itself; otherwise we will set $\theta^\sigma(e, f) = 0$. The purpose of the $y_{e,f}^\sigma$ below is to construct this function $\theta^\sigma(e, f)$ by defining a matrix $Y_{e,f}^\sigma = [y_{i,j}^\sigma]_{e \leq i, j \leq f-1}$ in such a way that $\theta^\sigma(e, f) = \det Y_{e,f}^\sigma$.

The subsets $\mathbf{a} = \{0 = a_0 < a_1 < \dots < a_{2k} < a_{2k+1} = m+1\} \subseteq [0, m+1]$ below determine the endpoints of the subwords $a_1 \dots a_2, a_3 \dots a_4$, through $a_{2k-1} \dots a_{2k}$ of id_m , each of which has length greater than one and may possibly be an embedded identity subword for σ . The exponent $h(\mathbf{a})$ ensures the correct sign for purposes of applying the Principle of Inclusion-Exclusion.

We are now ready for the Main Theorem.

Theorem 10. (*Theorem 3, restated in detail*)

$$(10) \quad \#\mathfrak{sh}(\text{id}_m, \sigma) = \sum_{k=0}^{\lfloor \frac{m}{2} \rfloor} \sum_{\mathbf{a}=\{0=a_0 < a_1 < \dots < a_{2k} < a_{2k+1}=m+1\}} (-1)^{h(\mathbf{a})} \prod_{r=0}^k \det Z_{a_{2r}, a_{2r+1}}^\sigma \prod_{s=1}^k \det Y_{a_{2s-1}, a_{2s}}^\sigma,$$

where

$$h(\mathbf{a}) = m - \sum_{t=1}^k (a_{2t} - a_{2t-1}),$$

and we define the matrices

$$Z_{c,d}^\sigma = [z_{i,j}^\sigma]_{c \leq i \leq d-1, c+1 \leq j \leq d},$$

with

$$z_{i,j}^\sigma = \begin{cases} 0, & i > j \\ 1, & i = j \\ 0, & 0 < i < j < m+1 \text{ and } \sigma^{-1}(i) > \sigma^{-1}(j) \\ \binom{j-i-1+\sigma^{-1}(j)-\sigma^{-1}(i)-1}{j-i-1}, & 0 < i < j < m+1 \text{ and } \sigma^{-1}(i) < \sigma^{-1}(j) \\ \binom{j-1+\sigma^{-1}(j)-1}{j-1}, & i = 0, j < m+1 \\ \binom{m-i+n-\sigma^{-1}(i)}{m-i}, & j = m+1, i > 0 \\ \binom{m+n}{m}, & i = 0, j = m+1, \end{cases}$$

and the matrices

$$Y_{e,f}^\sigma = [y_{i,j}^\sigma]_{e \leq i, j \leq f-1},$$

with

$$y_{i,j}^\sigma = \begin{cases} 0, & i - j > 1 \text{ or } \sigma^{-1}(i+1) \neq \sigma^{-1}(i) + 1 \\ -1, & i - j = 1 \text{ and } \sigma^{-1}(i+1) = \sigma^{-1}(i) + 1 \\ C_{j-i}, & i \leq j \text{ and } \sigma^{-1}(i+1) = \sigma^{-1}(i) + 1 \end{cases}$$

and where

$$C_{j-i} = \frac{1}{j-i+1} \binom{2(j-i)}{j-i}, \text{ the } (j-i)^{\text{th}} \text{ Catalan number.}$$

While equation (10) may look unwieldy, it is relatively easy to write a computer algorithm for Maple that will calculate the number of distinct shuffles of any two permutations. If at least one of the permutations has length bounded by 13, the processor on my laptop can easily handle the calculation. Examples of calculations include $\#\mathfrak{sh}(\text{id}_3, 321) = 14$, $\#\mathfrak{sh}(\text{id}_2, 3421) = 11$, $\#\mathfrak{sh}(2431, 1432) = 44$, $\#\mathfrak{sh}(\text{id}_6, 126354) = 374$, and if $\sigma = 7, 8, 9, 10, 11, 12, 13, 1, 2, 3, 4, 5, 6 \in \mathfrak{S}_{13}$, then $\#\mathfrak{sh}(\text{id}_{13}, \sigma) = 10104590$.

PLANS FOR FUTURE RESEARCH

In the immediate future, I plan to make progress on projects such as the following.

1. Enumerating Distinct Shuffles of Multiset Permutations. Compute the number of distinct shuffles of any two multiset permutations; for example, $\#\mathfrak{sh}(12322, 33214)$. This is a significant generalization of the current problem, because the possible ways that duplications in such shuffles can occur are much more complicated than with ordinary permutations, and multiplicities of shuffles no longer need to be powers of 2. I believe, however, that once I can classify the types of multiplicities that can occur the problem will become tractable, and that the intuitions gained in solving the current problem will help me to reach that point.

2. Enumerating Distinct k -Shuffles of Permutations. Compute the number of distinct k -shuffles of k permutations of any k lengths, where k is any positive integer; for example, $\#\mathfrak{sh}(132, 231, 1324)$. This is another important generalization. Again, multiplicities need not be powers of 2; rather, they appear to be related to products of factorials, but it is not yet clear how exactly to compute them. It seems that making progress on counting shuffles of multiset permutations should give insight into what occurs with k -shuffles of ordinary permutations; observe that $\#\mathfrak{sh}(132, 231, 1324)$ is equal to $\sum_{w \in \mathfrak{sh}(132, 231)} \#\mathfrak{sh}(w, 1324)$ minus a certain number of shuffles y such that $y \in \mathfrak{sh}(w, 1324) \cap \mathfrak{sh}(w', 1324)$ for some $w' \neq w \in \mathfrak{sh}(132, 231)$. Note that w and w' can be thought of as multiset permutations.

3. Deducing Monotonicity Results. Deduce monotonicity results for the number of distinct shuffles on permutation groups. Such results would help to clarify the meaning of the formula given in Theorem 10. For $1 \leq n \leq 6$, the minimal number of distinct shuffles of a permutation with the identity permutation of the same length is achieved by identity permutation, and I conjecture that this is the case for all n . For $n = 4, 5, 6$, the maximal number of distinct shuffles of a permutation with the identity permutation of the same length is achieved by the halfway-shifted permutations 3412, 34512, and 456123, respectively. I would like determine whether this maximality result is true for $n \geq 7$, and also to determine in general a poset structure on \mathfrak{S}_n for which the function $\sigma \mapsto \#\mathfrak{sh}(\text{id}_n, \sigma)$ is monotone increasing. The Bruhat order fails to provide such a structure for $n = 4, 5, 6$, but perhaps a modification of the Bruhat order would provide the desired poset structure.

4. Characterizing Properties of Permutations in a Shuffle Equivalence Class. Investigate the similarities among permutations $\rho, \eta \in \mathfrak{S}_{m,n}$ such that $\rho \sim \eta$ in the sense of Observation 2 above, for a given choice of two words to shuffle. It would be interesting to see if, for certain words, the equivalence classes on $\mathfrak{S}_{m,n}$ induced by shuffling those words can be characterized by criteria intrinsic to the permutations contained in each equivalence class. Candidates for such criteria would be various permutation statistics (descent sets, number of inversions, major index, and so forth).

5. Enumerating Distinct Shuffles according to Permutation Statistics. Enumerate distinct shuffles according to various permutation statistics, such as descent sets, number of inversions, or major index. Enumeration by statistics could yield insights into the above problems and refine my current results.

In the long term, I hope to work on other open problems in the field of permutation enumeration and permutation statistics.

UNDERGRADUATE RESEARCH PROPOSAL

I believe that Projects 1, 2, 4, and 5 above may, in part, be accessible to undergraduate researchers. For instance, in Project 1, I could direct undergraduates in writing a brute force program in Maple or Mathematica to compute the number of distinct shuffles of small multiset permutations. For Project 2, the number of distinct k -shuffles of permutations on n letters for small k and small n could be similarly computed. I could then help the students to analyze the data given by the programs and to look for plausible conjectures regarding formulae to count the distinct shuffles in various special cases. Once we have an idea of what is happening in the special cases, I will guide them in an effort to generalize and prove the results.

I also expect that Projects 4 and 5 (enumerating, according to various statistics, the permutations that induce shuffles of permutations as well as the shuffles of permutations themselves) may be approachable by undergraduates. Again, I would suggest to students that they start out by working out special cases and writing programs that can generate data for small parameters. I feel that it is likely that I could lead students in making and proving interesting conjectures regarding the similarities among permutations that induce the same shuffle, or else regarding the number of distinct shuffles of permutations classified by descent set or other suitable statistics.

I also plan to seek out further problems in permutation enumeration or other areas of Enumerative Combinatorics that may be suitable for undergraduate research projects.

REFERENCES

- D. Aldous & P. Diaconis, *Shuffling cards and stopping times*, American Mathematical Monthly **93** (5) (1986), 333-348.
- D. Bayer & P. Diaconis, *Trailing the dovetail shuffle to its lair*, Annals of Applied Probability **2** (2) (1992), 295-313.
- M.D. Crossley, *Some Hopf algebras of words*, Glasgow Math. J. **48** (2006), 575-582.
- P. Diaconis, *Group Representations in Probability and Statistics (Lecture Notes Vol 11)*, Institute of Mathematical Statistics, 1988, 77-84.
- P. Diaconis, *Mathematical developments from the analysis of riffle shuffling*, Technical Report 2002-16, Stanford University Department of Statistics, 2002.
- P. Diaconis, R.L. Graham, & W.M. Kantor, *The mathematics of perfect shuffles*, Advances in Applied Mathematics **4** (2) (1983), 175-196.
- W.F. Doran IV, *Shuffling lattices*, J. Combinatorial Theory Ser. A **66** (2002), 1-26.
- R. Ehrenborg, *On posets and Hopf algebras*, Adv. Math. **119** (1996), 1-25.
- J. Ellis, T. Krahn, & H. Fan, *Computing the cycles in the perfect shuffle permutation*, Information Processing Letters **75** (2000), 217-224.
- R.A. Fisher & F. Yates, *Statistical tables for biological, agricultural and medical research*, 3rd ed., Oliver & Boyd, 1948, 26-27.
- C. Greene, *Posets of shuffles*, J. Combinatorial Theory Ser. A **47** (1988), 191-206.

- L. Guo & B. Xie, *Explicit double shuffle relations and a generalization of Euler's decomposition formula*, arXiv:0808.2618v1 [math.NT], 2008.
- P. Hersh, *Two generalizations of posets of shuffles*, J. Combinatorial Theory Ser. A **97** (2002), 1-26.
- K. Ihara, M. Kaneko, & D. Zagier, *Derivation and double shuffle relations for multiple zeta values*, Compos. Math. **142** (2006), 307-338.
- D.E. Knuth, *The Art of Computer Programming 3*, Addison-Wesley, 1973.
- D.E. Knuth, *The Art of Computer Programming 2*, 3rd ed., Addison-Wesley, 1998.
- S. Mevedoff & K. Morrison, *Groups of perfect shuffles*, Math. Magazine **60** (1) (1987), 3-14.
- C. Reutenauer, *Free Lie Algebras*, Clarendon Press, 1993.
- R. Simion & R.P. Stanley, *Flag-symmetry of the poset of shuffles and a local action of the symmetric group*, Discrete Math. **204** (1999), 369-396.
- R.P. Stanley, *Enumerative Combinatorics 1*, Cambridge Univ. Press, 1997.
- R.P. Stanley, *Enumerative Combinatorics 2*, Cambridge Univ. Press, 1999.
- R.P. Stanley, personal communication, 2007.
- L.N. Trefethen & L.M. Trefethen, *How many shuffles to randomize a deck of cards?*, Proc. of the Royal Society Ser. A **456** (2002), 2451-2568.
- A. van Zuylen & F. Schalekamp, *The Achilles' heel of the GSR shuffle: a note on New Age Solitaire*, Probability in the Engineering and Informational Sciences **18** (3) (2004), 315-328.