The Secret Santa Problem

Matthew J. White
Tandem Friends School

Abstract
In this paper, we will investigate the Secret Santa problem, a combinatorics problem involving derangements with at least one two-cycle. We will first consider the probability that a permutation in a set of derangements has at least one two-cycle, and then generalize the result for derangements with at least one cycle of size q or smaller and derangements with at least one q-cycle. We will first solve for the probabilities by using recurrence relations, and will then provide them in non-recursive form. Next, we will reexamine the eight-year-old solution to the Secret Santa problem, demonstrating an error in the original authors' approach. We will solve for the error term, and generalize the results. Finally, we will provide secondary results, including an enumeration of the properties of a class of recurrence relations to which derangements and $n!$ belong.

1. Introduction

During the winter holidays, many organizations have gift-giving traditions. For instance, in my school the student body and faculty participate in a custom that we call "Secret Santa," in which each person randomly picks a name out of a box and then gives his or her chosen recipient small gifts. One year, a mathematics teacher announced that she had selected the same person who had selected her, and speculated on the probability of such an event. Given that the probability that among n individuals a certain individual will select a certain other individual is $\frac{1}{n-1}$, and that the probability that the other individual will select the first individual is also $\frac{1}{n-1}$, the teacher conjectured that the probability of such an event is $\frac{1}{(n-1)^2}$. However, this proved incorrect, and so the Secret Santa problem, a combination of the three following questions, remained open at my school:

1. What is the probability for n individuals that a certain individual selects a certain other individual?
2. What is the probability that a certain individual is a member of a two-cycle; what is the probability that a certain individual selects the same individual who selects him or her?
3. What is the probability that at least one two-cycle exists?

First, consider the teacher's proposed solution of $\frac{1}{(n-1)^2}$ to the first question. The solution seems reasonable, so why does it fail to resolve the problem? Suppose only three individuals, *A*, *B*, and *C*, participate in Secret Santa one year; what would be the

probability that year that *A* would pick *B*'s name and *B* would pick *A*'s name? According to the teacher's proposed solution, the answer should be $1/4$, but in fact such an event simply cannot occur. If *A* picks *B*'s name and *B* picks *A*'s name, *C* would inevitably pick his or her own name, which of course cannot occur, since the rules of Secret Santa state that each individual must select the name of another individual. The teacher erroneously assumed independence between *A* picking *B*'s name and *B* picking *A*'s name, when actually each selection is dependent on all previous selections.

Rather than concentrate on the first and second questions, which we will solve in section five, we will focus primarily on the more difficult third question in this paper. In their 1998 paper on the Secret Santa problem[i], Kelly M. McGuire, George Mackiw and Christopher H. Morrell asked the question in its negative, and restated it as the following: "What fraction of the number of derangements of *n* objects contains no two-cycle?"

This paper has five remaining sections: section two will review the results of the 1998 paper; section three will generalize the Secret Santa problem for derangements with minimal cycle size > *q* and derangements with at least one *q*-cycle; section four will introduce an error in the eight-year-old solution, solve for the error term, and generalize the results; section five will provide secondary results, including some involving recurrence relations of the form $a_n = (n-1)(a_{n-1} + a_{n-2}) + k$; finally, section six will conclude the paper and provide questions for future research.

2. Background

In this section we will consider the methods that McGuire et al. used to solve the third question. Let d_n be the number of different ways that *n* individuals can select each other. Since the set of selections may be considered a permutation, and no individual can select him or herself, it follows that d_n is the number of permutations of *n* individuals with no one-cycles, or derangements. Let T_n be the number of derangements with minimal cycle size of three; then $\frac{T_n}{d_n}$ should provide a solution to the third question in its negative, from which the actual solution quickly follows.

It is well-known that $d_n = \left[\frac{n!}{e}\right]$ for $n > 0$[ii], where $[x]$ is the nearest integer to *x*.

However, to determine the probability $\frac{T_n}{d_n}$, we must ascertain the slightly more elusive T_n. It is not difficult to determine through inspection the initial values $T_1 = 0$, $T_2 = 0$, and $T_3 = 2$, but it becomes more arduous to find T_n as *n* increases. Therefore, consider a derangement of *n* individuals with minimal cycle size of three. Suppose P_1, the first individual to choose, selects some individual P_2; then, to avoid forming a two-cycle, P_2 may not select P_1, but rather must select some other individual P_3. Suppose then that P_3 selects P_1. Since there are $n-1$ individuals who could be P_2, and $n-2$ individuals who could be P_3, the number of different derangements in which P_1 selects P_2, P_2 selects P_3,

and P_3 selects P_1 is $(n-1)(n-2)T_{n-3}$, as there are T_{n-3} ways for the remaining $n-3$ individuals to select each other. However, P_3 might not select P_1; P_3 might select some other individual P_4. The following outlines the different possible selections, where $P_1 \to P_2$ reads "P_1 selects P_2," $P_1 \to P_2 \to P_3$ reads "P_1 selects P_2, and P_2 selects P_3," and so forth:

Selection	Number of derangements
$P_1 \to P_2 \to P_3 \to P_1 \quad P_4 \to P_5 \to \ldots$	$(n-1)(n-2)T_{n-3}$
$P_1 \to P_2 \to P_3 \to P_4 \to P_1 \quad P_5 \to P_6 \to \ldots$	$(n-1)(n-2)(n-3)T_{n-4}$
…	
$P_1 \to P_2 \to P_3 \to P_4 \to \ldots \to P_1$	$(n-1)!$

Thus,

$$T_n = (n-1)(n-2)T_{n-3} + (n-1)(n-2)(n-3)T_{n-4} + \ldots + (n-1)!.$$

Factoring $n-1$ gives:

$$T_n = (n-1)[(n-2)T_{n-3} + (n-2)(n-3)T_{n-4} + (n-2)(n-3)(n-4)T_{n-5} + \ldots + (n-2)!]$$
$$T_n = (n-1)[(n-2)T_{n-3} + T_{n-1}]$$
$$T_n = (n-1)(n-2)T_{n-3} + (n-1)T_{n-1}.$$

Now that we have a recurrence relation with which to define T_n, we can specify $\frac{T_n}{d_n}$ for some constant *n*. Then, $1 - \frac{T_n}{d_n}$ should give the solution to the third question.

The problem that then logically follows is to find the value of $\frac{T_n}{d_n}$ as $n \to \infty$. McGuire et al. also present the following result: as $n \to \infty$, $\frac{T_n}{d_n} \to e^{-1/2}$.

3. Generalization for derangements with at least one cycle of size *q* or smaller and derangements with at least one *q*-cycle[iii]

Let $T_{n,q}$ be the number of derangements of *n* individuals with minimal cycle size > *q*, having initial values $T_{0,q} = 1$, $T_{n,q} = 0$ for $0 < n \le q$, and $T_{q+1,q} = q!$. Then, $\frac{T_{n,q}}{d_n}$ should satisfy the question, "What fraction of the number of derangements of *n* objects contains no cycle of size *q* or smaller?" Furthermore, $1 - \frac{T_{n,q}}{d_n}$ should give the probability that at least one cycle of size *q* or smaller exists in Secret Santa.

If no cycle of size q or smaller exists, then $P_1 \to P_2 \to P_3 \to \cdots \to P_{q+1}$. Then, P_{q+1} may either select P_1 or some other individual P_{q+2}. If P_{q+1} selects P_1, there are $T_{n-q-1,q}$ ways for the remaining $n-q-1$ individuals to select each other, and therefore $(n-1)(n-2)\cdots(n-q)T_{n-q-1,q}$ derangements in which $P_1 \to P_2 \to P_3 \to \cdots \to P_{q+1} \to P_1$. Extending the process produces the following identity:

$$T_{n,q} = (n-1)\cdots(n-q)T_{n-q-1,q} + (n-1)\cdots(n-q)(n-q-1)T_{n-q-2,q} + \ldots + (n-1)!.$$

Using a similar method as in section two gives the following:

$$T_{n,q} = (n-1)\cdots(n-q)T_{n-q-1,q} + (n-1)T_{n-1,q} = \frac{(n-1)!}{(n-q-1)!}T_{n-q-1,q} + (n-1)T_{n-1,q}.$$

With the recurrence relation, we can determine the fraction $\frac{T_{n,q}}{d_n}$ for some fixed n and q. However, $T_{n,q}$ can also be evaluated non-recursively:

Proposition 3.1: $T_{n,q} = n!\sum_{k=0}^{n}\left(\frac{(-1)^k}{k!}\sum_{j=0}^{\left\lfloor\frac{n-k}{q}\right\rfloor}\frac{(1-H_q)^j}{j!}\right)$, where H_q is the q^{th} harmonic number.

Proof:

For fixed q, let $T(x,q)$ be the exponential generating function for $T_{n,q}$.

$T(x,q) = \frac{\exp(-\sum_{k=1}^{q}\frac{x^k}{k})}{1-x}$ [iv], so $T(x,q)$ is the exponential generating function for the inverse binomial transform of the sequence $t_{n,q}$ whose exponential generating function is $\frac{\exp(-\sum_{k=2}^{q}\frac{x^k}{k})}{1-x}$; $T(x,q) = \sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}t_{k,q}$. But $t_{n,q}$ is the exponential convolution of the sequence $a_{n,q}$ whose exponential generating function $A(x,q) = \exp(-\sum_{k=2}^{q}\frac{x^k}{k})$ and the sequence $b_{n,q}$ whose exponential generating function is $B(x,q) = \frac{1}{1-x}$, from which the identity quickly follows.

Corollary 3.1.1: $\lim_{n\to\infty}\frac{T_{n,q}}{d_n} = e^{1-H_q}$.

Proof:

The corollary is clear from the form of $T_{n,q}$: since $\sum_{k=0}^{\infty}\frac{(-1)^k}{k!}=e^{-1}$ and $\sum_{j=0}^{\infty}\frac{(1-H_q)^j}{j!}=e^{1-H_q}$, $\lim_{n\to\infty}\frac{T_{n,q}}{d_n}=\frac{T_{n,q}/n!}{d_n/n!}=\frac{e^{-H_q}}{e^{-1}}=e^{1-H_q}$. For alternative proofs, the reader should turn to the notes[v].

Let $R_{n,q}$ be the number of derangements of *n* individuals with at least one *q*-cycle, having initial values of $R_{n,q}=0$ for $n<q$, and $R_{q,q}=(q-1)!$. Then $\frac{R_{n,q}}{d_n}$ should be the probability that at least one *q*-cycle exists in Secret Santa. Of course, $R_{n,2}=d_n-T_{n,2}$ for all *n*, but the result does not hold for greater *q*.

$R_{n,q}$ is the sum of the number of derangements of *n* individuals with exactly one *q*-cycle, the number with exactly two, etc. In general, there are $\frac{n!}{q^k(n-kq)!k!}$ distinct ways to choose *kq* individuals from *n* to form exactly *k* *q*-cycles, and $d_{n-kq}-R_{n-kq,q}$ ways for the remaining $n-kq$ individuals to select each other. Therefore,

$R_{n,q}=\sum_{k=1}^{\left\lfloor\frac{n}{q}\right\rfloor}\frac{n!}{q^k(n-kq)!k!}(d_{n-kq}-R_{n-kq,q})$, where $\lfloor x\rfloor$ is the floor of *x* and $\left\lfloor\frac{n}{q}\right\rfloor$ is the maximum number of *q*-cycles in a derangement of *n* individuals. Note that if $n\equiv 1(\mathrm{mod}\,q)$, only $\left\lfloor\frac{n}{q}\right\rfloor-1$ *q*-cycles can exist, but $d_{n-\left\lfloor\frac{n}{q}\right\rfloor q}-R_{n-\left\lfloor\frac{n}{q}\right\rfloor q,q}=d_1-R_{1,q}=0$.

Also, for $q=2$ and $q=3$ and $n\equiv 0(\mathrm{mod}\,q)$, $\frac{n}{q}-1$ *q*-cycles implies $\frac{n}{q}$ *q*-cycles, since the remaining individuals could only form a *q*-cycle, but $d_{n-(\frac{n}{q}-1)q}-R_{n-(\frac{n}{q}-1)q,q}=d_q-R_{q,q}=0$.

However, this recursive definition is not very efficient, nor does it easily provide insight into the behavior of $\frac{R_{n,q}}{d_n}$ as $n\to\infty$. Therefore, consider that for $q\neq 2$, there are $(n-1)R_{n-2,q}$ derangements in which $P_1\to P_2\to P_1$; for $q=2$, there are $(n-1)d_{n-2}$, since there is already a *q*-cycle. Continuing the process gives

$$R_{n,q}=(n-1)R_{n-2,q}+(n-1)(n-2)R_{n-3,q}+\ldots+(n-1)\cdots(n-q+1)d_{n-q}$$
$$+(n-1)\cdots(n-q)R_{n-q-1,q}+\ldots.$$

Therefore,

$$R_{n,q}=(n-1)(R_{n-2,q}+R_{n-1,q}+(n-2)\cdots(n-q)R_{n-q-1,q}$$

$$+(n-2)\cdots(n-q+1)(d_{n-q}-(n-q)d_{n-q-1}-R_{n-q,q}))$$

for $q>2$; the $R_{n-2,q}$ and $R_{n-q,q}$ terms disappear when $q=2$. But $d_n = nd_{n-1}+(-1)^n$. (See section five or refer to the notes[vi] for proof.) Thus,

$R_{n,q} = (n-1)(R_{n-1,q}+R_{n-2,q}+(n-2)\cdots(n-q)R_{n-q-1,q}+(n-2)\cdots(n-q+1)((-1)^{n-q+1}-R_{n-q,q}))$ for $q>2$. This can be expressed in terms of factorials as

$$R_{n,q} = (n-1)(R_{n-1,q}+R_{n-2,q}+\frac{(n-2)!}{(n-q-1)!}R_{n-q-1,q}+\frac{(n-2)!}{(n-q)!}((-1)^{n-q+1}-R_{n-q,q})).$$

However, similarly to $T_{n,q}$, $R_{n,q}$ can also be evaluated non-recursively:

Proposition 3.2: $R_{n,q} = n!\sum_{k=0}^{n}(\frac{(-1)^k}{k!}(1-\sum_{j=0}^{\lfloor\frac{n-k}{q}\rfloor}\frac{(-1)^j}{q^j j!})) = \sum_{k=0}^{n}(\frac{(-1)^{k+1}}{k!}\sum_{j=0}^{\lfloor\frac{n-k}{q}\rfloor}\frac{(-1)^j}{q^j j!}))$.

Proof:

For fixed *q*, let $R(x,q)$ be the exponential generating function of $R_{n,q}$.

$R(x,q) = \frac{e^{-x}}{1-x} - \frac{e^{-x-\frac{x^q}{q}}}{1-x}$ [vii], so $R(x,q)$ is the exponential generating function for the inverse binomial transform of the sequence $r_{n,q}$ whose exponential generating function is $\frac{1}{1-x}-\frac{e^{-\frac{x^q}{q}}}{1-x}$; $R(x,q)=\sum_{k=0}^{n}(-1)^{n-k}\binom{n}{k}r_{k,q}$. But from its exponential generating function, which is an exponential convolution, we see that $r_{n,q} = n!-n!\sum_{k=0}^{\lfloor\frac{n}{q}\rfloor}\frac{(-1)^k}{q^k k!}$, so

$$R_{n,q} = n!\sum_{k=0}^{n}(\frac{(-1)^k}{k!}(1-\sum_{j=0}^{\lfloor\frac{n-k}{q}\rfloor}\frac{(-1)^j}{q^j j!})) = \sum_{k=0}^{n}(\frac{(-1)^{k+1}}{k!}\sum_{j=0}^{\lfloor\frac{n-k}{q}\rfloor}\frac{(-1)^j}{q^j j!})).$$

Corollary 3.2.1: $\lim_{n\to\infty}\frac{R_{n,q}}{d_n} = 1-e^{-1/q}$.

Proof:

Similarly to Corollary 3.1.1, $\sum_{k=0}^{\infty}\frac{(-1)^k}{k!} = e^{-1}$ and $\sum_{j=0}^{\infty}\frac{(-1)^j}{q^j j!} = e^{-1/q}$, so

$$\lim_{n\to\infty}\frac{R_{n,q}}{d_n} = \frac{R_{n,q}/n!}{d_n/n!} = \frac{e^{-1}(1-e^{-1/q})}{e^{-1}} = 1-e^{-1/q}.$$

Alternative Proof:

$T_{n,q-1} - T_{n,q}$ is the number of derangements with minimal cycle size > $q-1$ with at least one *q*-cycle. The two events are not independent; the probability that at least one *q*-cycle exists is affected by the condition that the minimal cycle size > $q-1$. However, it is clear that the two events are more independent for large *n*, so that their effect upon each other diminishes to relative insignificance as *n* becomes sufficiently large. Therefore, $\frac{T_{n,q-1} - T_{n,q}}{d_n} \sim \frac{T_{n,q-1}}{d_n}(\frac{R_{n,q}}{d_n})$. But as $n \to \infty$, by Corollary 3.1.1,

$\frac{T_{n,q-1} - T_{n,q}}{d_n} = e^{1-H_{q-1}} - e^{1-H_q} = e^{1-H_{q-1}}(1-e^{-1/q}) = \frac{T_{n-q-1}}{d_n}(1-e^{-1/q})$, so $\frac{R_{n,q}}{d_n} \sim 1-e^{-1/q}$, i.e.

$\lim_{n\to\infty} \frac{R_{n,q}}{d_n} = 1-e^{-1/q}$. For another alternative proof, the reader should turn to the notes[viii].

$R_{n,q}$ is also related to degree-*n* permutations of order *q*, with fixed *q*:

Proposition 3.3: For fixed *q*, the number of degree-*n* even permutations of order *q* minus the number of degree-*n* odd permutations of order *q* equals $(-1)^n(R_{n,q} - nR_{n-1,q})$ if and only if *q* is prime.

Proof:

The exponential generating function for the number of degree-*n* even permutations of order dividing *q* minus the number of degree-*n* odd permutations of order dividing *q* is $\exp\sum_{k|q}\frac{(-1)^{k+1}x^k}{k}$ [ix]. Let $A(x,q)$ be the exponential generating function for the number of degree-*n* even permutations of order *q* minus the number of degree-*n* odd permutations of order *q*. Then $A(x,2) = e^{x-\frac{x^2}{2}} - e^x$, since for prime *q* the number of degree-*n* even permutations of order *q* minus the number of degree-*n* odd permutations of order *q* equals the number of degree-*n* even permutations of order dividing *q* minus the number of degree-*n* odd permutations of order dividing *q* minus one, and $A(x,q) = e^{x+\frac{x^q}{q}} - e^x$ for odd prime *q*.

We know from Proposition 3.2 that $R(x,q) = \frac{e^{-x}}{1-x} - \frac{e^{-x-\frac{x^q}{q}}}{1-x}$. Therefore, $(1-x)R(x,q) = -A(-x,q)$ if and only if *q* is prime, from which the proposition quickly follows.

Also, let a $(1,q)$-permutation be a permutation with only one-cycles and *q*-cycles; then $(-1)^n - (R_{n,q} - nR_{n-1,q})$ is the number of $(1,q)$-permutations with an even number of cycles minus the number of $(1,q)$-permutations with an odd number of cycles.

4. The Error Term

With $\frac{R_{n,q}}{d_n}$, it may appear as if we have succeeded in both solving the third question and generalizing the result. However, throughout this paper we have assumed that the question "What fraction of the number of derangements of *n* objects contains at least one two-cycle?" is mathematically equivalent to "What is the probability that at least one two-cycle exists in Secret Santa?" If a single derangement is chosen randomly to describe the assorted gift-giving, then the two questions are mathematically equivalent. Suppose, however, that *n* individuals select each other one after another in some (presumably random) order, as if they had been assembled in a single file, as is common in Secret Santa. Then we cannot be sure that the questions are identical; it seems to be a reasonable assumption, but it implies that each selection order – henceforth "arrangement" – is equally likely, despite the single file process, or that the average arrangement with at least one two-cycle is as equally likely as the average arrangement. If the assumption proves true, $\frac{R_{n,2}}{d_n}$ will indeed be the solution to the third question; however, if it proves false, we have yet to solve the Secret Santa problem. Suppose four individuals, *A*, *B*, *C*, and *D*, participate in Secret Santa; the following are the only nine different ways (derangements) that the four individuals can select each other, using cycle notation[x]:

(*A*, *B*, *C*, *D*)
(*A*, *B*, *D*, *C*)
(*A*, *C*, *B*, *D*)
(*A*, *C*, *D*, *B*)
(*A*, *D*, *B*, *C*)
(*A*, *D*, *C*, *B*)
(*A*, *B*)(*C*, *D*)
(*A*, *C*)(*B*, *D*)
(*A*, *D*)(*B*, *C*)

Since there are three derangements that have at least one two-cycle, if the assumption is true, then the solution to the third question for $n = 4$ should be $1/3$. When four individuals select each other, they form either one four-cycle or two two-cycles; therefore, each arrangement for $n = 4$ must be equally likely for the assumption to be true.

We should now admit that we have not taken arrangements into account yet; for instance, (*A*, *B*) does not indicate whether *A* or *B* selects first. Since there are four selections for *A*, *B*, *C*, and *D*, there are 4! different arrangements in which the four individuals can choose. For the sake of simplicity, assume that *A* chooses first, *B* second, *C* third, and *D* fourth; note that we can assume a fixed arrangement without compromising the integrity of the mathematics. Then, we can apply elementary probability techniques to determine the probability of each arrangement. For instance, the arrangement (*A*, *B*, *C*, *D*) has a probability of $1/9$, since *A* has three choices (*B*, *C*, or

D), then *B* has three choices (*A*, *C*, or *D*), and then *C* must choose *D* and *D* must choose *A*. Although the probability of this arrangement corresponds to the expected probability, the probability of (*A*, *C*, *B*, *D*) is $1/12$: *A* again has three choices (*B*, *C*, or *D*), but then *C* has only two choices (*A* or *D*), then *B* has two choices (*A* or *C*), and then *D* must choose *A*. As the following table demonstrates, each arrangement is not equally likely:

Arrangement	Probability
(*A*, *B*, *C*, *D*)	$1/9$
(*A*, *B*, *D*, *C*)	$1/9$
(*A*, *C*, *B*, *D*)	$1/12$
(*A*, *C*, *D*, *B*)	$1/6$
(*A*, *D*, *B*, *C*)	$1/12$
(*A*, *D*, *C*, *B*)	$1/6$
(*A*, *B*)(*C*, *D*)	$1/9$
(*A*, *C*)(*B*, *D*)	$1/12$
(*A*, *D*)(*B*, *C*)	$1/12$

We must conclude that the assumption does not hold true for $n = 4$.

It is also now clear that the assumption is false for $n > 5$; we must merely consider the probability of (*A*, *B*, *C*, *D*)(*E*, *F*, …) to that of (*A*, *C*, *B*, *D*)(*E*, *F*, …). The probability of (*E*, *F*, …) is the same in both arrangements, but according to the table above, (*A*, *B*, *C*, *D*) and (*A*, *C*, *B*, *D*) are not equally likely. However, we know neither why each arrangement is not equally likely nor what the actual solution to the third question is, if not $\frac{R_{n,2}}{d_n}$. Therefore, in order to more fully understand arrangements, consider all *n*! different arrangements for each arrangement, rather than just a single fixed arrangement. If each arrangement is not equally likely, there must be something fundamentally different in the way that the probabilities behave for some arrangements than for others.

First, however, consider a new notation. Let $P_1|A \to B|P_2(k)$ read "*A*, the k^{th} individual to choose, selects *B*. The probability that *A* selected when he or she did (e.g. *A* was first to choose, or second to choose, etc.) is P_1 and the probability that *A* selected *B* is P_2, given previous selections (i.e. P_1 and P_2 are conditional probabilities)." For example, suppose *A* was the first to choose among *n* individuals and he or she selected *B*; we write $\frac{1}{n}|A \to B|\frac{1}{n-1}(1)$. Furthermore, if we denote multiple selections using this notation, we write the selections next to each other with a '/' in-between, chronologically, so that the notation for the individual first to choose is the leftmost, followed by the

second to choose, etc. For instance, suppose in the previous example that *B* was second to choose and selected *A*; then we write the two selections as

$$\frac{1}{n}\left|A\rightarrow B\right|\frac{1}{n-1}(1)\ /\ \frac{1}{n-1}\left|B\rightarrow A\right|\frac{1}{n-1}(2).$$

We can now use our new notation to evaluate the actual probability that at least one two-cycle exists in Secret Santa for four individuals *A*, *B*, *C*, and *D*. Instead of evaluating 4! arrangements, we will assume, without loss of generality, that *A* chooses first, and evaluate those 3! arrangements. Let (*A*, *B*)(*C*, *D*); the following are the derangement's 3! distinct arrangements:

$$\frac{1}{4}\left|A\rightarrow B\right|\frac{1}{3}(1)\ /\ \frac{1}{3}\left|B\rightarrow A\right|\frac{1}{3}(2)\ /\ \frac{1}{2}\left|C\rightarrow D\right|1(3)\ /\ 1\,|\,D\rightarrow C\left|1(4)\right. = 1/432$$

$$\frac{1}{4}\left|A\rightarrow B\right|\frac{1}{3}(1)\ /\ \frac{1}{3}\left|B\rightarrow A\right|\frac{1}{3}(2)\ /\ \frac{1}{2}\left|D\rightarrow C\right|1(3)\ /\ 1\,|\,C\rightarrow D\left|1(4)\right. = 1/432$$

$$\frac{1}{4}\left|A\rightarrow B\right|\frac{1}{3}(1)\ /\ \frac{1}{3}\,|\,C\rightarrow D\left|\frac{1}{2}(2)\right.\ /\ \frac{1}{2}\left|B\rightarrow A\right|\frac{1}{2}(3)\ /\ 1\,|\,D\rightarrow C\left|1(4)\right. = 1/288$$

$$\frac{1}{4}\left|A\rightarrow B\right|\frac{1}{3}(1)\ /\ \frac{1}{3}\,|\,C\rightarrow D\left|\frac{1}{2}(2)\right.\ /\ \frac{1}{2}\left|D\rightarrow C\right|\frac{1}{2}(3)\ /\ 1\,|\,B\rightarrow A\left|1(4)\right. = 1/288$$

$$\frac{1}{4}\left|A\rightarrow B\right|\frac{1}{3}(1)\ /\ \frac{1}{3}\left|D\rightarrow C\right|\frac{1}{2}(2)\ /\ \frac{1}{2}\,|\,B\rightarrow A\left|\frac{1}{2}(3)\right.\ /\ 1\,|\,C\rightarrow D\left|1(4)\right. = 1/288$$

$$\frac{1}{4}\left|A\rightarrow B\right|\frac{1}{3}(1)\ /\ \frac{1}{3}\left|D\rightarrow C\right|\frac{1}{2}(2)\ /\ \frac{1}{2}\left|C\rightarrow D\right|\frac{1}{2}(3)\ /\ 1\,|\,B\rightarrow A\left|1(4)\right. = 1/288$$

We take the average probability of the arrangements and multiply it by 4! to determine that the probability of (*A*, *B*)(*C*, *D*) is $5/54$. We multiply $5/54$ by 3, since (*A*, *B*)(*C*, *D*), (*A*, *C*)(*B*, *D*), and (*A*, *D*)(*B*, *C*) are essentially identical derangements, which gives an actual probability of $5/18$, which is $1/18$ less than the expected $1/3$.

Proposition 4.1: Let $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$ be the probability of a derangement of *n* individuals with exactly σ_2 two-cycles, exactly σ_3 three-cycles, etc. Let $q(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$ be defined by a product of a nested sum, where the nest runs through *s*, *r*, and *t*, and the sums run through $k_{s,r,t}$, as

$$q(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n) = \sum_{\substack{1\le k_{2,1,0}\le n-2\\ k_{2,1,0}\ne j_m,\,1\le m\le\sigma-2}} \left[\frac{1}{n-k_{2,1,0}+\left\lceil\left\lfloor k_{2,1,0}/k_{2,1,1}\right\rfloor/k_{2,1,0}\right\rceil}\right.$$

$$\cdot \sum_{\substack{1\le k_{2,1,1}\le n-2\\ k_{2,1,1}\ne j_m,1\le m\le \sigma-2\\ k_{2,1,1}\ne k_{2,1,0}}} [\frac{1}{n-k_{2,1,1}+\left\lceil \left\lfloor k_{2,1,1}/k_{2,1,0}\right\rfloor /k_{2,1,1}\right\rceil}$$

$$\cdots$$

$$\sum_{\substack{1\le k_{3,1,0}\le n-2\\ k_{3,1,0}\ne j_m,1\le m\le \sigma-2\\ k_{3,1,0}\ne k_{2,a,b}}} [\frac{1}{n-k_{3,1,0}+\left\lceil \left\lfloor k_{3,1,0}/k_{3,1,2}\right\rfloor /k_{3,1,0}\right\rceil}$$

$$\cdots$$

$$\sum_{\substack{1\le k_{s,r,t}\le n-2\\ k_{s,r,t}\ne j_m,1\le m\le \sigma-2\\ 2\le s\le n\\ s=\sigma\Rightarrow 1\le r\le \sigma_s-1\\ s\ne\sigma\Rightarrow 1\le r\le \sigma_s\\ 0\le t\le s-1\\ s\ne a\Rightarrow k_{s,r,t}\ne k_{a,b,c}\\ s=a,r\ne b\Rightarrow k_{s,r,t}\ne k_{a,b,c}\\ s=a,r=b,t\ne c\Rightarrow k_{s,r,t}\ne k_{a,b,c}}} [\frac{1}{n-k_{s,r,t}+\left\lceil \left\lfloor k_{s,r,t}/k_{s,r,(t-1)(\mathrm{mod}\, s)}\right\rfloor /k_{s,r,t}\right\rceil}$$

$$\cdots]\cdots]\cdots] .$$

Then

$$p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)=\sum[\cdots \sum_{\substack{1\le k_{s,r,t}\le n\\ k_{s,r,t}=n-1\Rightarrow k_{s,r,(t-1)(\mathrm{mod}\, s)}\ne n\\ k_{s,r,t}=n\Rightarrow k_{s,r,(t-1)(\mathrm{mod}\, s)}\ne n-1\\ 2\le s\le n\\ 1\le r\le \sigma_s\\ 0\le t\le s-1\\ s\ne a\Rightarrow k_{s,r,t}\ne k_{a,b,c}\\ s=a,r\ne b\Rightarrow k_{s,r,t}\ne k_{a,b,c}\\ s=a,r=b,t\ne c\Rightarrow k_{s,r,t}\ne k_{a,b,c}}} [\frac{1}{n-k_{s,r,t}+\left\lceil \left\lfloor k_{s,r,t}/k_{s,r,(t-1)(\mathrm{mod}\, s)}\right\rfloor /k_{s,r,t}\right\rceil}\cdots$$

$$+\frac{2\sigma_2}{n(n-1)}q(n,\sigma_2,\sigma_3,...,\sigma_{n-2},\sigma_n,2)$$

$$+\frac{6\sigma_3}{n(n-1)}\sum_{j_1=1}^{n-2}[\frac{1}{n-j_1}q(n,\sigma_2,\sigma_3,...,\sigma_{n-2},\sigma_n,3)]$$

$$+\ldots$$

$$+\frac{2i\sigma_i}{n(n-1)}\sum_{j_1=1}^{n-2}[\frac{1}{n-j_1}\sum_{\substack{j_2=1\\ j_2\ne j_1}}^{n-2}[\frac{1}{n-j_2+\left\lceil \left\lfloor j_2/j_1\right\rfloor /j_2\right\rceil}\sum_{\substack{j_3=1\\ j_3\ne j_1\\ j_3\ne j_2}}^{n-2}[\frac{1}{n-j_3+\left\lceil \left\lfloor j_3/j_2\right\rfloor /j_3\right\rceil}\cdots$$

$$\sum_{\substack{j_{k-2}=1\\ j_{k-2}\ne j_m,1\le m\le i-1}}^{n-2}[\frac{1}{n-j_{i-2}+\left\lceil \left\lfloor j_{i-2}/j_{i-3}\right\rfloor /j_{i-2}\right\rceil}q(n,\sigma_2,\sigma_3,...,\sigma_{n-2},\sigma_n,i)]\cdots]]]$$

$$+\ldots,$$

where the first term is a product of a nested sum similar to the one in $q(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$.

Proof:

We must find the average probability of the derangement's arrangements. We will therefore consider each factor in the selection process within an arrangement.

We will first consider the effect of P_1 on $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$. Suppose we have two individuals *A* and *B* among *n* individuals in Secret Santa and *A* selects *B*. Then $\frac{1}{n-k+1}\left|A\rightarrow B\right|P_2(k)$ for all *k*; P_1 is completely independent of the objects of all previous selections (i.e. who chooses whom), so P_1 is a variable of *n* and *k* alone, and by elementary probability, $P_1=\frac{1}{n-k+1}$. Therefore, P_1 affects no arrangement differently than others and has a uniform effect on $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$ for all σ_2, σ_3, etc.: P_1 will contribute a factor of $\frac{1}{n!}$ to $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$.

However, note that for every derangement there are *n*! arrangements. We are considering averages, so after deducing the average probability of a derangement's arrangements we must multiply it by *n*!. P_1 contributes a factor of $\frac{1}{n!}$ to $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$, so together the two phenomena have no impact on $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$.

We will now consider the effect of P_2 on $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$. Consider $P_1\left|A\rightarrow B\right|P_2(k)$ for $k\le n-2$. There have already been $k-1$ individuals selected, so if *A* has not been selected yet, then there are *k* individuals that he or she cannot select, including him or herself. Therefore, the probability would be $\frac{1}{n-k}$ that *A* individual would select *B*. However, if *A* has been selected already, then there are only $k-1$ individuals that he or she cannot select. Therefore, in that case the probability would be $\frac{1}{n-k+1}$.

However, whether *A* has already been selected depends on the second-to-last and last-to-choose. We will therefore consider the following two cases:

1. the second-to-last does not select the last and the last does not select the second-to-last;
2. the second-to-last selects the last or the last selects the second-to-last.

The probability of the first case is

$$\sum[\cdots \sum_{\substack{1\le k_{s,r,t}\le n \\ k_{s,r,t}=n-1\Rightarrow k_{s,r,(t-1)(\mathrm{mod}\,s)}\ne n \\ k_{s,r,t}=n\Rightarrow k_{s,r,(t-1)(\mathrm{mod}\,s)}\ne n-1 \\ 2\le s\le n \\ 1\le r\le\sigma_s \\ 0\le t\le s-1 \\ s\ne a\Rightarrow k_{s,r,t}\ne k_{a,b,c} \\ s=a,r\ne b\Rightarrow k_{s,r,t}\ne k_{a,b,c} \\ s=a,r=b,t\ne c\Rightarrow k_{s,r,t}\ne k_{a,b,c}}} [\frac{1}{n-k_{s,r,t}+\left\lceil \left\lfloor k_{s,r,t} \middle/ k_{s,r,(t-1)(\mathrm{mod}\,s)} \right\rfloor \middle/ k_{s,r,t} \right\rceil}\cdots]\cdots].$$

In the nested sum, *s* runs through the cycle sizes, *r* runs through the individual cycles of size *s*, and *t* runs through the individual members of the unique cycle specified by *s* and *r*; $k_{s,r,t}$ denotes the position (i.e. the value of *k* in $P_1|A\rightarrow B|P_2(k)$) held by the unique individual specified by *s*, *r*, and *t* (i.e. P_1 in $P_1|A\rightarrow B|P_2(k)$). $1\le k_{s,r,t}\le n$, but $k_{s,r,t}=n-1\Rightarrow k_{s,r,(t-1)(\mathrm{mod}\,s)}\ne n$ and $k_{s,r,t}=n\Rightarrow k_{s,r,(t-1)(\mathrm{mod}\,s)}\ne n-1$, since the second-to-last does not select the last and the last does not select the second-to-last; $t<s$, so $(t-1)(\mathrm{mod}\,s)=t-1$ for $t\ne 0$ and $(t-1)(\mathrm{mod}\,s)=s-1$ for $t=0$. $2\le s\le n$, $1\le r\le\sigma_s$, and $0\le t\le s-1$ follow from the definitions of *s*, *r*, and *t*, and $s\ne a\Rightarrow k_{s,r,t}\ne k_{a,b,c}$, $s=a,r\ne b\Rightarrow k_{s,r,t}\ne k_{a,b,c}$, and $s=a,r=b,t\ne c\Rightarrow k_{s,r,t}\ne k_{a,b,c}$ follow from the fact that only a single individual can occupy any given position. Since

$\left\lceil \left\lfloor k_{s,r,t} \middle/ k_{s,r,(t-1)(\mathrm{mod}\,s)} \right\rfloor \middle/ k_{s,r,t} \right\rceil = 0 \Leftrightarrow k_{s,r,(t-1)(\mathrm{mod}\,s)} > k_{s,r,t}$ and

$\left\lceil \left\lfloor k_{s,r,t} \middle/ k_{s,r,(t-1)(\mathrm{mod}\,s)} \right\rfloor \middle/ k_{s,r,t} \right\rceil = 1 \Leftrightarrow k_{s,r,t} > k_{s,r,(t-1)(\mathrm{mod}\,s)}$ (note that $k_{s,r,t}\ne k_{s,r,(t-1)(\mathrm{mod}\,s)}$), our argument is complete.

We now consider the probability of the second case. The probability that two members of the same cycle of size *s*, one of whom has selected the other, are second-to-last and last is $\frac{2\sigma_2}{n(n-1)}$ for $s=2$ and $\frac{2s\sigma_s}{n(n-1)}$ for $s>2$, since the probability that a

member of a cycle of size s is second-to-last is $\frac{s\sigma_s}{n}$ and the probability that the individual he or she selected or the individual who selected him or her is last is $\frac{1}{n-1}$ for $s=2$ and $\frac{2}{n-1}$ for $s>2$; for $s=2$, the individual he or she selected is the same individual who selected him or her. Here we can see a bias forming against derangements with two-cycles; a member of a two-cycle who is second-to-last or last is less likely to have his or her object or selector in one of the last two positions. Therefore, a derangement with a greater number of two-cycles is less likely to have the second-to-last individual have a probability of selection of 1, and its average arrangement will therefore be more likely to receive a factor of $1/2$ from the second-to-last individual's selection probability, making it less likely. However, this observation, although it demonstrates bias against derangements with two-cycles, does not conclusively show that derangements with two-cycles are less likely; other biases may exist.

For $s>2$, there will always be an individual – the object of the second-to-last or last – who will select before he or she is selected, and therefore the probability of his or her selection will always be $\frac{1}{n-k}$, from which the nested sum $\sum_{j_1=1}^{n-2}[\frac{1}{n-j_1}\sum_{\substack{j_2=1\\ j_2\neq j_1}}^{n-2}[\frac{1}{n-j_2+\lceil \lfloor j_2/j_1\rfloor / j_2\rceil}\sum_{\substack{j_3=1\\ j_3\neq j_1\\ j_3\neq j_2}}^{n-2}[\frac{1}{n-j_3+\lceil \lfloor j_3/j_2\rfloor / j_3\rceil}\cdots$ follows. Finally, when the cycle is complete, $q(n,\sigma_2,\sigma_3,...,\sigma_{n-2},\sigma_n,i)$ will enumerate all the arrangements of the $n-s$ remaining individuals, by an argument similar to one in the first case.

We conclude the proposition with an example:

$p(4,2,0) =$

$$\sum_{\substack{1\le k_{2,1,0}\le 4\\ k_{2,1,0}=3\Rightarrow k_{2,1,1}\neq 4\\ k_{2,1,0}=4\Rightarrow k_{2,1,1}\neq 3}}[\frac{1}{4-k_{2,1,0}+\left\lceil \left\lfloor k_{2,1,0}/k_{2,1,1}\right\rfloor / k_{2,1,0}\right\rceil}\sum_{\substack{1\le k_{2,1,1}\le 4\\ k_{2,1,1}\neq k_{2,1,0}\\ k_{2,1,1}=3\Rightarrow k_{2,1,0}\neq 4\\ k_{2,1,1}=4\Rightarrow k_{2,1,0}\neq 3}}[\frac{1}{4-k_{2,1,1}+\left\lceil \left\lfloor k_{2,1,1}/k_{2,1,0}\right\rfloor / k_{2,1,1}\right\rceil}$$

$$\sum_{\substack{1\le k_{2,2,0}\le 4\\ k_{2,2,0}\neq k_{2,1,a}\\ k_{2,2,0}=3\Rightarrow k_{2,2,1}\neq 4\\ k_{2,2,0}=4\Rightarrow k_{2,2,1}\neq 3}}[\frac{1}{4-k_{2,2,0}+\left\lceil \left\lfloor k_{2,2,0}/k_{2,2,1}\right\rfloor / k_{2,2,0}\right\rceil}\sum_{\substack{1\le k_{2,2,1}\le 4\\ k_{2,2,1}\neq k_{2,1,a}\\ k_{2,2,1}\neq k_{2,2,0}\\ k_{2,2,0}=3\Rightarrow k_{2,2,1}\neq 4\\ k_{2,2,0}=4\Rightarrow k_{2,2,1}\neq 3}}\frac{1}{4-k_{2,2,1}+\left\lceil \left\lfloor k_{2,2,1}/k_{2,2,0}\right\rfloor / k_{2,2,1}\right\rceil}]]]$$

$$+\frac{1}{3}\sum_{1\le k_{2,1,0}\le 2}[\frac{1}{4-k_{2,1,0}+\left\lceil \left\lfloor k_{2,1,0}/k_{2,1,1}\right\rfloor / k_{2,1,0}\right\rceil}\sum_{\substack{1\le k_{2,1,1}\le 2\\ k_{2,1,1}\neq k_{2,1,0}}}\frac{1}{4-k_{2,1,1}+\left\lceil \left\lfloor k_{2,1,1}/k_{2,1,0}\right\rfloor / k_{2,1,1}\right\rceil}]$$

$=\frac{5}{54}$, and the definition holds true.

The example, however, reveals how inefficient this definition of $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$ is; of course, the nested sum is more efficient than generating

all the arrangements and finding their average probability, but the efficiency of the definition is still $O(m^n)$ for all *n*. Also, the definition's form does not lend itself easily to a full analysis of $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$. Therefore, consider that a derangement is simply a union of disjoint cycles; we can then limit our investigation of the selection process to *q*-cycles, rather than more complex derangements. We must now only determine the average number of individuals within a *q*-cycle who select after they have already been selected and then weigh the probability that the individuals will occupy certain positions against the possible probabilities of each position (with respect to the larger *n*-cycle).

Proposition 4.2: Let $u(q,k)$ be the number of arrangements of a *q*-cycle such that exactly *k* individuals select after they have already been selected. Then $u(q,k) = q\left\langle {q-1 \atop k-1} \right\rangle$, where $\left\langle {q-1 \atop k-1} \right\rangle$ is an Eulerian number.

Proof:

Let P_1 be the first individual to choose, and let P_2 be the object of P_1's selection. P_1 will never be selected before he or she selects, and P_2 will always be selected before he or she selects. Therefore, we only need to find the number of arrangements of the $q-1$ individuals not P_1 such that $k-1$ individuals select after they have been selected. However, this is just the number of permutations of order $q-1$ with $k-1$ permutation descents, i.e. $\left\langle {q-1 \atop q-k-1} \right\rangle$, which equals $\left\langle {q-1 \atop k-1} \right\rangle$. But P_1 can be *q* different individuals, so $u(q,k) = q\left\langle {q-1 \atop k-1} \right\rangle$.

Proposition 4.3: $\frac{1}{q\cdot q!}\sum_{k=1}^{q-1} k\cdot u(q,k)$, the average number of individuals who select after they have already been selected divided by (i.e. relative to) *q* equals $1/2$.

Proof:

$u(q,k) = q\left\langle {q-1 \atop k-1} \right\rangle$, so

$$\frac{1}{q\cdot q!}\sum_{k=1}^{q-1} k\cdot u(q,k) = \frac{1}{q!}\sum_{k=1}^{q-1} k\left\langle {q-1 \atop k-1} \right\rangle = \frac{1}{q!}((q-1)(q-1)! - \sum_{k=1}^{q-2}(q-k-1)\left\langle {q-1 \atop k-1} \right\rangle)$$

$$= \frac{1}{q!}((q-1)(q-1)! - (\sum_{k=1}^{q-2}(q-k)\left\langle {q-1 \atop q-k-1} \right\rangle - \sum_{k=1}^{q-2}\left\langle {q-1 \atop q-k-1} \right\rangle)) = \frac{1}{q!}(q! - \sum_{k=1}^{q-1} k\left\langle {q-1 \atop k-1} \right\rangle)$$

$$\Rightarrow \frac{1}{q\cdot q!}\sum_{k=1}^{q-1} k\cdot u(q,k) = 1/2.$$

Remark: Since $\frac{1}{q \cdot q!}\sum_{k=1}^{q-1} k \cdot u(q,k)$ is constant, there is no extra bias against any *q*-cycle because of the average number of such individuals.

If $\frac{1}{q \cdot q!}\sum_{k=1}^{q-1} k \cdot u(q,k)$ had not been constant, some *q*-cycles would have been less likely than others, since relative to their size they would have produced more $\frac{1}{n-k+1}$ factors in the larger derangement. But, since it is constant, one might be tempted to conjecture that only bias against two-cycles exist, since derangements with more two-cycles are more likely to have their second-to-last individual have a selection probability of $1/2$. However, examining $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$ for different derangements reveals, for instance, that $p(6,0,2,0,0) < p(6,0,0,0,1)$, so there is bias against cycles of size greater than two as well.

Even if the average number of individuals within a *q*-cycle who are selected before they select relative to *q* is constant, those individuals may still produce different factors. For example, if the individuals within a three-cycle who are selected before they select on average choose before such individuals within a six-cycle, there will be some bias against three-cycles relative to six-cycles, since the $\frac{1}{n-k+1}$ factors will be smaller on average. Therefore, we must weigh the probability that the individuals within a *q*-cycle will occupy certain positions against the possible probabilities of each position with respect to the larger *n*-cycle. It is clear that some derangements will produce arrangements that are unique to them; for instance, an *n*-cycle is the only derangement of *n* individuals that will produce an arrangement with a probability of $\frac{2}{(n-1)(n-1)!n!}$, when all individuals except the first-to-choose are selected before they select (the least likely arrangement of any derangement), or an arrangement with a probability of $\frac{1}{(n-1)!n!}$, when all individuals are selected after they select (the most likely arrangement of any derangement).

However, as *n* grows, the differences between the arrangements become smaller, suggesting that the derangements will become more equally likely. Also, analysis of $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$ for small *n* suggests that $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n) \sim \frac{1}{d_n}$ for all values of σ_2, σ_3, etc. However, we cannot determine whether this is true until we weigh the probability that the individuals within a *q*-cycle will occupy certain positions against the possible probabilities of each position with respect to the larger *n*-cycle. Nevertheless, we can still solve exactly, if not asymptotically, for the probability of the generalized third question:

Proposition 4.4: Let $P(n,q)$ be the actual probability that at least one *q*-cycle exists in Secret Santa for *n* individuals. Then

$$P(n,q)=\sum_{\substack{1\le\sigma_q\le\lfloor n/q\rfloor\\ 0\le\sigma_i\le\lfloor n/i\rfloor, i\ne q, i\le n-q\\ 2\sigma_2+3\sigma_3+\ldots+(n-q)\sigma_{n-q}=n}} \frac{n!}{2^{\sigma_2}3^{\sigma_3}\cdots(n-q)^{\sigma_{n-q}}\sigma_2!\sigma_3!\cdots\sigma_{n-q}!}p(\sigma_2,\sigma_3,\ldots,\sigma_{n-q},0,\ldots,0),$$

where the sum runs through all derangements with at least one two-cycle.

Proof:

$\frac{n!}{2^{\sigma_2}3^{\sigma_3}\cdots(n-q)^{\sigma_{n-q}}\sigma_2!\sigma_3!\cdots\sigma_{n-q}!}$ is the number of derangements with exactly σ_2 two-cycles, exactly σ_3 three-cycles, etc., and $p(\sigma_2,\sigma_3,\ldots,\sigma_{n-q},0,\ldots,0)$ is the probability of such a derangement. We then sum through all derangements with at least one two-cycle, so

$$P(n,q)=\sum_{\substack{1\le\sigma_q\le\lfloor n/q\rfloor\\ 0\le\sigma_i\le\lfloor n/i\rfloor, i\ne q, i\le n-q\\ 2\sigma_2+3\sigma_3+\ldots+(n-q)\sigma_{n-q}=n}} \frac{n!}{2^{\sigma_2}3^{\sigma_3}\cdots(n-q)^{\sigma_{n-q}}\sigma_2!\sigma_3!\cdots\sigma_{n-q}!}p(\sigma_2,\sigma_3,\ldots,\sigma_{n-q},0,\ldots,0).$$

Remark: If $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)\sim\frac{1}{d_n}$, $P(n,q)\sim\frac{R_{n,q}}{d_n}$, since without the factor of $p(\sigma_2,\sigma_3,\ldots,\sigma_{n-q},0,\ldots,0)$, the sum is exactly $R_{n,q}$.

We will further expound upon consequences of the error term in the next section.

5. Secondary results and recurrence relations of the form $a_n=(n-1)(a_{n-1}+a_{n-2})+k$

We will begin this section by returning to the recurrence relations of section three; we will now provide an alternative derivation of $R_{n,q}=(n-1)(R_{n-1,q}+R_{n-2,q}+\frac{(n-2)!}{(n-q-1)!}R_{n-q-1,q}+\frac{(n-2)!}{(n-q)!}((-1)^{n-q+1}-R_{n-q,q}))$ for $q>2$. Let $S_{n,q,k}$ be the number of derangements of *n* individuals with exactly *k* *q*-cycles. Consider that for $q\ne 2$, there are $(n-1)S_{n-2,q,k}$ derangements in which $P_1\to P_2\to P_1$; for $q=2$, there are $(n-1)S_{n-2,q,k-1}$, since there is already a *q*-cycle. Continuing gives

$$S_{n,q,k}=(n-1)S_{n-2,q,k}+(n-1)(n-2)S_{n-3,q,k}+\ldots+(n-1)\cdots(n-q+1)S_{n-q,q,k-1}$$
$$+(n-1)\cdots(n-q)S_{n-q-1,q,k}+\ldots.$$

Therefore,

$$S_{n,q,k}=(n-1)(S_{n-1,q,k}+S_{n-2,q,k}+\frac{(n-2)!}{(n-q-1)!}(S_{n-q-1,q,k}-S_{n-q-1,q,k-1})$$
$$+\frac{(n-2)!}{(n-q)!}(S_{n-q,q,k-1}-S_{n-q,q,k}))$$

for $q > 2$; the $S_{n-2,q,k}$ and $S_{n-q,q,k}$ terms disappear when $q = 2$. But

$S_{n,q,k} = \frac{n!}{q^k (n-qk)!\,k!}(d_{n-qk} - R_{n-qk,q})$, from which

$R_{n,q} = (n-1)(R_{n-1,q} + R_{n-2,q} + \frac{(n-2)!}{(n-q-1)!}R_{n-q-1,q} + \frac{(n-2)!}{(n-q)!}((-1)^{n-q+1} - R_{n-q,q}))$ quickly follows.

We will now turn to the first and second questions from the introduction; now that we have solved the generalized third question, we can also solve the generalized first and second questions:

1. What is the probability for *n* individuals that *q* certain individuals are in the same *q*-cycle?
2. What is the probability that a certain individual is a member of a *q*-cycle?

Proposition 5.1: Let $P(n,q)$ be the probability that *q* certain individuals are in the same *q*-cycle. Then

$$P(n,q) = \sum_{\substack{0 \le \sigma_i \le \lfloor (n-q)/i \rfloor, i \le n-q \\ 2\sigma_2 + 3\sigma_3 + \ldots + (n-q)\sigma_{n-q} = n-q}} \frac{(n-q)!}{2^{\sigma_2} 3^{\sigma_3} \cdots (n-q)^{\sigma_{n-q}} \sigma_2! \sigma_3! \cdots \sigma_{n-q}!} p(n, \sigma_2, \ldots, \sigma_{q-1}, \sigma_q + 1, 0, \ldots)$$

Proof:
We have that *q* certain individuals are in the same *q*-cycle, so the remaining $n-q$ individuals' selections will not involve them. Therefore, the remaining individuals will form among themselves a derangement disjoint from the *q*-cycle.

$\frac{(n-q)!}{2^{\sigma_2} 3^{\sigma_3} \cdots (n-q)^{\sigma_{n-q}} \sigma_2! \sigma_3! \cdots \sigma_{n-q}!}$ is the number of derangements of the $n-q$ remaining individuals with exactly σ_2 two-cycles, exactly σ_3 three-cycles, etc., and $p(n, \sigma_2, \ldots, \sigma_{q-1}, \sigma_q + 1, 0, \ldots)$ is the probability of a derangement of the *n* individuals with one more *q*-cycle than there are within the derangement of the $n-q$ remaining individuals. We then sum through all derangements of the $n-q$ remaining individuals, so

$$P(n,q) = \sum_{\substack{0 \le \sigma_i \le \lfloor (n-q)/i \rfloor, i \le n-q \\ 2\sigma_2 + 3\sigma_3 + \ldots + (n-q)\sigma_{n-q} = n-q}} \frac{(n-q)!}{2^{\sigma_2} 3^{\sigma_3} \cdots (n-q)^{\sigma_{n-q}} \sigma_2! \sigma_3! \cdots \sigma_{n-q}!} p(n, \sigma_2, \ldots, \sigma_{q-1}, \sigma_q + 1, 0, \ldots)$$

Remark: If $p(n, \sigma_2, \sigma_3, \ldots, \sigma_{n-2}, \sigma_n) \sim \frac{1}{d_n}$, $P(n,q) \sim \frac{d_{n-q}}{d_n}$.

Proposition 5.2: Let $P(n,q)$ be the probability that a certain individual is a member of a *q*-cycle. Then

$$P(n,q) = \frac{(n-1)!}{(n-q)!} \sum_{\substack{0 \le \sigma_i \le \lfloor (n-q)/i \rfloor, i \le n-q \\ 2\sigma_2 + 3\sigma_3 + \ldots + (n-q)\sigma_{n-q} = n-q}} \frac{(n-q)!}{2^{\sigma_2} 3^{\sigma_3} \cdots (n-q)^{\sigma_{n-q}} \sigma_2! \sigma_3! \cdots \sigma_{n-q}!} p(n, \sigma_2, \ldots, \sigma_{q-1}, \sigma_q + 1, 0, \ldots)$$

Proof: We have that a *q*-cycle exists, and a certain individual is a member of it. There are $\frac{(n-1)!}{(n-q)!}$ distinct ways for the remaining $n-1$ individuals to also be members of the *q*-cycle; we then multiply this by the probability that *q* certain individuals are in the same *q*-cycle (see Proposition 5.1), so

$$P(n,q)$$
$$= \frac{(n-1)!}{(n-q)!} \sum_{\substack{0\le\sigma_i\le\lfloor (n-q)/i \rfloor, i\le n-q \\ 2\sigma_2+3\sigma_3+\ldots+(n-q)\sigma_{n-q}=n-q}} \frac{(n-q)!}{2^{\sigma_2}3^{\sigma_3}\cdots(n-q)^{\sigma_{n-q}}\sigma_2!\sigma_3!\cdots\sigma_{n-q}!} p(n,\sigma_2,\ldots,\sigma_{q-1},\sigma_q+1,0,\ldots)$$

Remark: If $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n) \sim \frac{1}{d_n}$, $P(n,q) \sim \frac{(n-1)!\,d_{n-q}}{(n-q)!\,d_n}$.

In an attempt to find a closed form definition for $R_{n,q}$, I investigated a number of sequences, and produced a few results for a class of recurrence relations to which d_n and $n!$ belong, and have included them below.

Consider the recurrence relation $d_n = (n-1)(d_{n-1}+d_{n-2})$ [xi]; we now inspect the more general $a_n = (n-1)(a_{n-1}+a_{n-2})+k$ with initial values a_0 and a_1.

Proposition 5.3: a_n can also be defined recursively as $a_n = na_{n-1} + \frac{k}{2} + (a_0 - a_1 + \frac{k}{2})(-1)^n$.

Proof:

We will use induction. First note that $a_1 = a_0 + \frac{k}{2} + (a_0 - a_1 + \frac{k}{2})(-1)^1)$, so the identity is true for $n=1$. Now assume that the identity is true for some $q \in \mathbf{Z}^+$. By the first recurrence relation, $a_{q+1} = q(a_q + a_{q-1}) + k = qa_q + qa_{q-1} + k$. But then

$$a_{q+1} = qa_q + (qa_{q-1}\frac{k}{2} + (a_0 - a_1 + \frac{k}{2})(-1)^q) + k - (\frac{k}{2} + (a_0 - a_1 + \frac{k}{2})(-1)^q))$$
$$\Rightarrow a_{q+1} = (q+1)a_q + \frac{k}{2} + (a_0 - a_1 + \frac{k}{2})(-1)^{q+1}.$$

Thus, if the identity is true for some *q* then it is also true for $q+1$. Since we have already shown that the identity is true for a_1, the statement is true for all *n*.

Remark: From the two recursive definitions, we see that $d_n = (n-1)(d_{n-1}+d_{n-2}) = nd_{n-1} + (-1)^n$. In fact, if the sequence s_n is defined recursively as $s_n = ns_{n-1} + c(-1)^n$ for some real constant *c*, s_n can also be defined as $s_n = (n-1)(s_{n-1}+s_{n-2})$. If $c=0$, $s_n = n!$, so *n*! and d_n can be defined by the same recurrence relation, and the only difference between the two sequences' definitions is that $1!=1$ but $d_1 = 0$.

Proposition 5.2: a_n has the closed form definition of

$$a_n = \left\lfloor \frac{ke^2 - 2e(k-a_1) + k + 2(a_0 - a_1)}{2e} n! \right\rfloor \text{ for } n > 0 \text{ whenever } a_n \in \mathbf{Z}.$$

Proof:

We first determine the exponential generating function of the sequence and then use it to define the sequence in terms of *n* and *k*.

Let $A(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!}$. Then,

$$A(x) = \frac{e^{-x}(ke^{2x} - 2(k-a_1)e^x + k + 2(a_0 - a_1))}{2(1-x)}$$

$$= \frac{1}{2}\sum_{n=0}^{\infty} x^n \cdot (ke^x - 2(k-a_1) + (k + 2(a_0 - a_1))e^{-x})$$

$$= \frac{1}{2}(k\sum_{n=0}^{\infty}(x^n \sum_{i=0}^{n}\frac{1}{i!}) - 2(k-a_1)\sum_{n=0}^{\infty} x^n + (k + 2(a_0 - a_1))\sum_{n=0}^{\infty}(x^n \sum_{i=0}^{n}\frac{(-1)^i}{i!}))$$

Therefore,

$$A(x) = \sum_{n=0}^{\infty} a_n \frac{x^n}{n!} = \frac{1}{2}\sum_{n=0}^{\infty}(k\sum_{i=0}^{n}\frac{1}{i!} - 2(k-a_1) + (k + 2(a_0 - a_1))\sum_{i=0}^{n}\frac{(-1)^i}{i!})n!\frac{x^n}{n!} \text{ and}$$

$$a_n = \frac{1}{2}(k\sum_{i=0}^{n}\frac{1}{i!} - 2(k-a_1) + (k + 2(a_0 - a_1))\sum_{i=0}^{n}\frac{(-1)^i}{i!}n!, \text{ so}$$

$$a_n = \left\lfloor \frac{ke^2 - 2e(k-a_1) + k + 2(a_0 - a_1)}{2e} n! \right\rfloor \text{ for } n > 0 \text{ whenever } a_n \in \mathbf{Z}.$$

6. Conclusion

With $p(n, \sigma_2, \sigma_3, \ldots, \sigma_{n-2}, \sigma_n)$, $R_{n,q}$, and $T_{n,q}$, we should now be able to more easily answer questions related to Secret Santa. However, we need to further understand the behavior of $p(n, \sigma_2, \sigma_3, \ldots, \sigma_{n-2}, \sigma_n)$ in order to determine the probabilities' asymptotic values. Interestingly, if $p(n, \sigma_2, \sigma_3, \ldots, \sigma_{n-2}, \sigma_n) \sim \frac{1}{d_n}$,

$P(n,2) \sim \frac{d_{n-2}}{d_n} = \frac{\left\lfloor \frac{(n-2)!}{e} \right\rfloor}{\left\lfloor \frac{n!}{e} \right\rfloor} \sim \frac{1}{n(n-1)}$, where $P(n,q)$ is the probability of the generalized first question, and the teacher's proposed solution of $\frac{1}{(n-1)^2}$ for $P(n,2)$ will be very accurate for sufficiently large *n*.

This paper should serve at least as a cautionary tale: it is tempting to assume that events are equally likely when in fact they are not. Also, any algorithm that attempts to randomly select permutations with some imposed conditions – such as that no cycles of certain sizes may exist – should take into account that some bias may exist for some permutations; for example, an algorithm that calculates all derangements and then randomly selects one will be completely random, but an algorithm that selects a random derangement by a single-file-like method, although much more efficient, will not be.

This paper should provide a few opportunities for future research. Are there simpler definitions for $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$? What is the exact behavior of $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$? What is the exact correlation between average cycle length and $p(n,\sigma_2,\sigma_3,\ldots,\sigma_{n-2},\sigma_n)$? The nested sum may seem difficult to unravel, but the fact that if $v(q)=\sum_{i=0}^{q-1}\left\lceil \left\lfloor j_i \Big/ j_{(i-1)(\mathrm{mod}\,q)} \right\rfloor \Big/ j_i \right\rceil$,

$u(q,k)=\sum_{1\le j_0\le n}\left[\sum_{\substack{1\le j_1\le n\\ j_1\ne j_0}}\left[\cdots\sum_{\substack{1\le j_{n-1}\le n\\ j_{n-1}\ne j_m,\,m\le n-2}}\left\lceil \left\lfloor v(q)/2 \right\rfloor / v(q)\right\rceil-\left\lceil \left\lfloor v(q)/3 \right\rfloor / v(q)\right\rceil\right.\right.=q\left\langle {q-1 \atop k-1}\right\rangle$ suggests the existence of a simpler definition. Also, it is curious that $\gcd(d_n,R_{n,2})=n-1$ for $1\le n\le 16$, but $\gcd(d_{17},R_{17,2})=908$. What is the behavior of $\gcd(d_n,R_{n,q})$?

Nevertheless, although we cannot evaluate asymptotic probabilities for the generalized three questions, we can still determine their exact solutions.

[i] Kelly McGuire, and others, "The Secret Santa problem." *The Mathematical Gazette*, (November 1998), 467–472.
[ii] Kenneth P. Bogart, *Introductory Combinatorics, Second Edition* (Washington, D.C.: Harcourt Brace & Company, 1990), 165.
[iii] Many thanks to Vladeta Jovovic for the proofs of Proposition 3.2 and Proposition 3.3, and for the observation about $(1,q)$ cycles.
[iv] Herbert S. Wilf, *Generatingfunctionology* (New York: Academic Press, 1994), 176.
[v] See E.A. Bender, "Asymptotic methods in enumeration." *Siam Review*, (Issue 4, 1974), 499 or Herbert S. Wilf, *Generatingfunctionology* (New York: Academic Press, 1994), 176–177.
[vi] Kenneth P. Bogart, *Introductory Combinatorics, Second Edition* (Washington, D.C.: Harcourt Brace & Company, 1990), 167, 573.
[vii] Herbert S. Wilf, *Generatingfunctionology* (New York: Academic Press, 1994).
[viii] E.A. Bender, "Asymptotic methods in enumeration." *Siam Review*, (Issue 4, 1974), 499.
[ix] Herbert S. Wilf, *Generatingfunctionology* (New York: Academic Press, 1994).
[x] Joseph A. Gallian, *Contemporary Abstract Algebra, Fourth Edition* (New York: Houghton Mifflin Company, 1998), 93.
[xi] Kenneth P. Bogart, *Introductory Combinatorics, Second Edition* (Washington, D.C.: Harcourt Brace & Company, 1990), 164.