The Game of Set

Malaika Aiyar, Johnny Sammon, Nico Smith
Mentor: Alex Dam

August 2022

Abstract

The game of Set is a card game, especially popular among mathematicians for its underlying mathematical structure. In Set, each card has four attributes with three values for each one, and a 'SET' is three cards that have either all the same or all different values for each.

In this paper we explore the geometric structure of the game, modelling it as a finite 4-dimensional space, somewhat like points on a hypercube. Through this we answered combinatorial questions about the maximum number of SETs in n cards, probabilities relating to the number of expected SETs.

We set out to find the maximum number of cards with no SET. While this could be theoretically tackled by brute force computation, the number of comparisons was too large (over 10^{21} co-linearity checks).

To overcome this, we invented a notion of linear equivalence mappings which allowed us to show that certain configurations of cards were the same mathematically, despite containing different cards. This drastically reduced the number of cases to check, without losing any precision. We were able to answer how many cards there can be with no SET in the classic game, as well as in variations of the game with different numbers of attributes.

A fascinating remark is that this equivalence relationship is somewhat arbitrary; while we used translations and linear transformations, there are many other equivalence mappings that could be explored in the future. Additionally, many of the results from geometric modelling generalise to finite n-dimensional space.

1 Introduction

The game of Set[1] is a popular card game. However, Set has also attracted attention from the mathematical community due to its underlying mathematical structure. When the game is abstracted from the specifics of its game play, a rich architecture emerges. In order to analyse and better understand its mechanics, we constructed a formal model of the game, which would allow us to combine geometrical and combinatorial approaches to motivate insights that we could then develop computer programs to verify. This allowed us to not only examine the theory of Set, but also generate meaningful numerical results using computational methods.

2 The game, terms, and notation

2.1 The game

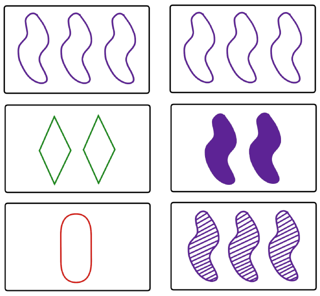
The game of Set is played with a deck of *cards*. Each card has 4 *attributes*: shape, fill, colour, and number. There are 3 possible *values* for each property:

- Shape: oval, squiggle, diamond
- Colour: red, green, purple
- Fill: full, shaded, empty
- Number: one, two, three

Every card is unique, meaning no two cards have the same value for all four properties. Every possible combination of values is in the Set deck. This explains why there are $3^4 = 81$ cards in a Set deck.

A *SET* is a group of 3 cards such that for each property, the values of the 3 cards are all different or all the same. For example, in Figure 1, the first column is a SET, as all three have the same Fill, different Colour, different Shape, and different Number. The second column is not a SET because the Number attribute is not all different or all the same.

Figure 1



The game of Set is played by laying 9 randomly selected cards from the deck in a 3×3 grid, and players call "Set" if they see a SET in the 9 cards. If players think there are no SETs in the 9 cards, 3 new cards are added. This process repeats until a SET is finally found. When a SET is found, the player who found it takes those cards away, and three new cards are added to the grid. The aim of the game is to find the most SETs.

2.2 Terms

Definition (Dimension)**.** The number of attributes in a game of Set. n is often the letter used to denote the dimension of a game of Set. For example, the standard Set game has 4 dimensions.

[1]Set is a trademark of SET Enterprises, Inc. It was invented in 1974 by Marsha Jean Falco. Set game play is protected intellectual property.

Definition (Configuration)**.** A collection of any number of cards.

Notation. If a, b, and c are a configuration of cards, we write $\{a, b, c\}$ to represent the three cards, even if they do not make a SET. Any number of cards can be in a configuration. Examples of configurations include $\{a, b, c, d\}$, $\{a, b, d, e, f, c\}$, $\{a\}$.

If a, b, and c are three cards, and they form a SET, we write $[abc]$ to denote the SET. Only 3 cards can be in a SET, by definition.

Definition (n-optimal SET-free configuration)**.** The configuration in n-dimensional Set with the maximum number of cards with no complete SET.

Definition (SET-full configuration)**.** Given a fixed number of cards, the SET-full configuration for that number of cards is a configuration of cards with the maximum number of SETs possible.

Definition (Completion card)**.** Given a configuration of two cards $\{a, b\}$, a completion card c is a card such that $[abc]$ is a SET.

3 Modelling and methodology

3.1 Generalising attributes

Although the usual game of Set has 4 attributes with very specific values, we can reduce each attribute to a values of $0, 1$, or 2. For example, instead of 'colour' having values 'red', 'green' and 'purple', we can just say that 'red' is 0, 'green' is 1, and 'purple' is 2, which allows us to investigate cards more mathematically.

We also consider different numbers of attributes. This does not change the fundamental principles of the game, as you can still find SETs of three cards where all the attributes are the same or different, as long as each attribute can have exactly three values. Reducing the dimension to 2 or 3 allows us to experiment with smaller numbers, and see if we can generalise to higher dimensions.

3.2 Points in $\mathbb{Z}_3^n$

Cards can be represented as a point in n-dimensional space (for n attributes), mod 3 (three values).

Notation: If a card a has values $a_i \in \mathbb{Z}_3, 1 \leq i \leq n$ for each of its attributes, we write $(a_1, a_2, ..., a_n)$ to denote the point that represents that card in $\mathbb{Z}_3^n$.

For example, $(1, 2, 1, 0)$, $(1, 1, 0, 0)$, and $(1, 0, 2, 0)$ are cards in 4-dimensional Set. Note that they form a SET.

Lemma 3.1. *The vector sum of three cards in a SET (mod* 3*) is* 0*.*

Proof. Let a, b, and c be cards with components a_i, b_i and c_i, with $a_i, b_i, c_i \in \mathbb{Z}_3$, and $1 \leq i \leq n$, where n is the dimension of the game.

Let i be an attribute, arbitrary but fixed. If 3 cards are a SET, a_i, b_i, and c_i are the same or all different. If two cards have the same value for one attribute are the third card has a different value, then the cards are not a SET.

If $a_i = b_i = c_i$, then $a_i + b_i + c_i = a_i + a_i + a_i = 3a_i \equiv 0 \mod 3$.

If all three are different, then $a_i + b_i + c_i = 0 + 1 + 2 = 3 \equiv 0 \mod 3$.

If two have the same value and one has a different value (which means the cards cannot make a SET), without loss of generality, suppose that $a_i = b_i$ and $b_i \neq c_i$. Then $c_i = a_i \pm 1$. So, the sum of the three cards is $a_i + b_i + c_i = 2a_i + (a_i \pm 1) = 3a_i \pm 1$, which is congruent to either 1 or $-1 \mod 3$.

So, if all the attributes sum to $0 \mod 3$, then the cards make a SET. □

Definition (direction vector)**.** We say the *direction vector* between two cards is the difference in each attribute between one card to the next.

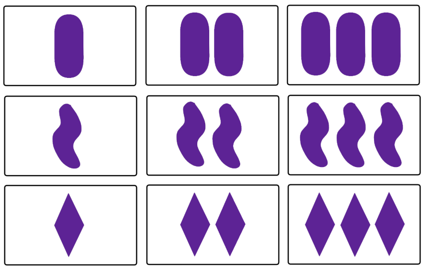
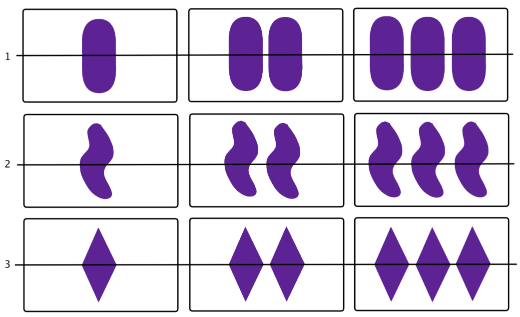
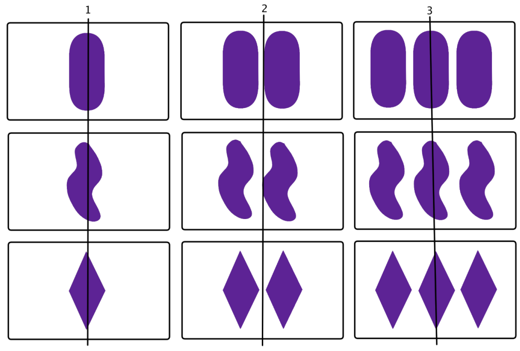


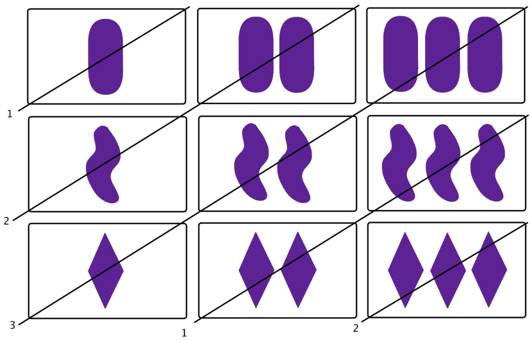
Figure 2: Caption



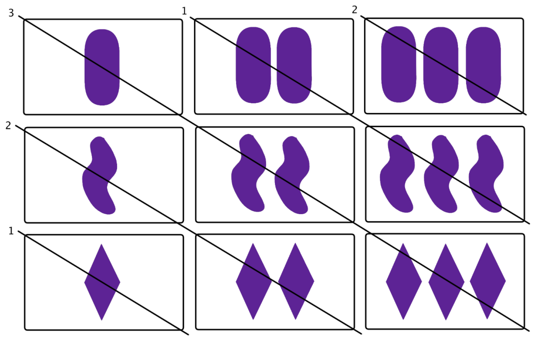
(a) horizontal lines



(b) vertical lines



(c) up-diagonal lines



(d) down-diagonal lines

Figure 3: Types of lines

For example, if a point a has components a_i, and a point b has components b_i, the direction vector $\mathbf{v}$ has values $v_i = b_i - a_i$, with all values in $\mathbb{Z}_3$

Definition (line)**.** If the direction vector between a and b is the same as the direction vector between b and c, then we say that a, b and c lie on a line. In other words, a line is a set of points such that all pairs of consecutive points have the same direction vector. A line is characterised by a point it passes through and a direction vector to get from one point to the next.

A line in our methodology is nearly the same as lines in normal n-dimensional space, except we have more diagonals due to working mod 3. Consider the 2-dimensional game of Set. We can plot every card in the plane as shown in Figure 2.

We use the terms *horizontal*, *vertical*, *up-diagonal*, and *down-diagonal* to denote different types of lines. Figure 3 shows how these look in 2-dimensional Set.

Proposition 3.2. *If* 3 *cards are a SET, they lie on a line.*

Proof. Let a, b, c be three cards with components a_i, b_i, c_i such that $[abc]$ is a SET.

Suppose that the three cards do not lie on a line. In other words, the direction vector between a and b is different to that of b and c.

Let $\mathbf{r}$ be the direction vector between a and b with components $\mathbf{r_i}$, and $\mathbf{s}$ be the direction vector between b and c with components $\mathbf{s_i}$.

WLOG, suppose that $r_1 \neq s_1$.

Then,

$$b_1 = a_1 + r_1$$

$$c_1 = b_1 + s_1 = a_1 + r_1 + s_1$$

So,

$$\begin{aligned} a_1 + b_1 + c_1 &\equiv a_1 + (a_1 + r_1) + (a_1 + r_1 + s_1) \pmod 3 \\ &\equiv 3a_1 + 2r_1 + s_1 \pmod 3 \\ &\equiv 2r_1 + s_1 \pmod 3 \end{aligned}$$

As $r_1 \neq s_1$, we have $s_1 = r_1 \pm 1$. So,

$$\begin{aligned} a_1 + b_1 + c_1 &\equiv 2r_1 + (r_1 \pm 1) \pmod 3 \\ &\equiv 3r_1 \pm 1 \pmod 3 \\ &\equiv \pm 1 \not\equiv 0 \pmod 3 \end{aligned}$$

By Lemma 3.1, since one of the attributes sums to a non-zero value mod 3, the cards cannot be a SET. This is a contradiction, as a, b, and c are a SET. So, our assumption that $\mathbf{r} \neq \mathbf{s}$ must be false, so all three cards lie on a line. □

Theorem 3.3. *In $\mathbb{Z}_3^n$, the number of lines through any point is* $\frac{3^n-1}{2}$

Proof. We can view $\mathbb{Z}_3^n$ as an n-dimensional space of lattice points, with each dimension having three distinct possible values. For example, $\mathbb{Z}_3^2$ can be represented by a 3×3 grid of lattice points, and $\mathbb{Z}_3^3$ can be represented as a $3 \times 3 \times 3$ cube of lattice points. Each lattice point corresponds to a unique Set card.

Suppose the line passes through the point at a, with each component some a_i. Let the next point from a on the line be some b, if we fix an arbitrary forwards direction. For each component b_i of b, b_i could be either $a_i + 1$, a_i, or $a_i - 1$, since in $\mathbb{Z}_3^n$, a change of ± 2 is equivalent to a change of ∓ 1. So there are 3 choices for each dimension, all of which are distinct. However, we have to exclude the case where all $b_i = a_i$, since the next point on the line cannot be a. So there are $3^n - 1$ choices for the next point on the line. However, we must divide by two to account for the arbitrary choice of the forwards direction.

So the total number of lines through any point is $\frac{3^n-1}{2}$ □

Lemma 3.4. *For any line L in $\mathbb{Z}_3^n$, exactly 3 points lie on L.*

Proof. We can characterise a line L in $\mathbb{Z}_3^n$ by one point that it passes through a with components a_i and a direction vector $\mathbf{v}$, some n-tuple of 1, 0, and -1. To move from a to the next point on L, we simply add $\mathbf{v}$ to a. Note that $\mathbf{v}$ must have at least one non-zero component, since the next point on L from a must be different from a. Let some $v_k \neq 0$, so $v_k = \pm 1$.

Let the next point on L from a be some point b. So $b_k = a_k + v_k = a_k \pm 1$. Specifically, $b_k \neq a_k$ in $\mathbb{Z}_3$ so b is not the same point as a in $\mathbb{Z}_3^n$. Let c be the next point on L from b, so $c_k = b_k + v_k = a_k \pm 2$, meaning $c \neq a$ in $\mathbb{Z}_3$ and therefore c is not the same point as a in $\mathbb{Z}_3^n$. Let the next point on L from c be some d. So $d_k = c_k + v_k = a_k \pm 3 = a_k$.

Therefore, the fourth point from a on L has the same k^{th} component for every k where $v_k \neq 0$. For the components d_l where $v_l = 0$, $d_l = a_l$, so every component of d is the same as its corresponding component in a. So d is the same point as a in $\mathbb{Z}_3^n$. So after a, b, c, the points on L cycle back to a and repeat. Therefore, exactly 3 distinct points lie on L. □

Lemma 3.5. *Let $\mathbf{c}$ and $\mathbf{d}$ be two direction vectors. If $\mathbf{c} = -\mathbf{d}$, then the direction of both are the same.*

Proof. Let a_1 be a point in $\mathbb{Z}_3^n$. If we get the three points (there are only three by Lemma 3.4) on the line defined by point a_1 with direction vector $\mathbf{c}$, we get points a_1, a_2 and a_3 on the line such that

$$a_2 = a_1 + \mathbf{c}$$

$$a_3 = a_1 + 2\mathbf{c} = a_2 + \mathbf{c}$$

If we get the points on the line defined by the point a_1 and direction vector $\mathbf{d}$, we get a_1, a_2', a_3' such that

$$a_2' = a_1 + \mathbf{d}$$

$$a_3' = a_1 + 2\mathbf{d} = a_2' + \mathbf{d}$$

As we are working in $\mathbb{Z}_3^n$, we have

$$a_3' = a_1 + 2\mathbf{d} \qquad = a_1 - \mathbf{d} = a_1 + \mathbf{c} \qquad = a_2$$

So $a_3' = a_2$.
We also have

$$a_2' = a_3' - \mathbf{d} \qquad = a_3' + \mathbf{d} = a_2 + \mathbf{c} \qquad = a_3$$

So, $a_2' = a_3$.

The three points on both lines are the same. So, they are the same line. Therefore, the direction of both direction vectors is the same.

□

Theorem 3.6. *There are exactly $3^{n-1} \cdot \frac{3^n-1}{2}$ distinct lines in $\mathbb{Z}_3^n$.*

Proof. Let the number of lines in $\mathbb{Z}_3^n$ be some l_n From Lemma 3.3, we know each point in $\mathbb{Z}_3^n$ has exactly $\frac{3^n-1}{2}$ lines passing through it. There are 3^n points in $\mathbb{Z}_3^n$ and, from Lemma 3.4, each line passes through exactly 3 points in $\mathbb{Z}_3^n$. So:

$$l_n = \frac{1}{3} \cdot 3^n \cdot \frac{3^n-1}{2} = 3^{n-1} \cdot \frac{3^n-1}{2}$$

□

4 Initial combinatorial results

Theorem 4.1. *Every pair of Set cards have exactly one completion card.*

Proof. Let $a = (a_1, a_2, ..., a_n)$ and $b = (b_1, b_2, ..., b_n)$ in $\mathbb{Z}_3^n$ be two Set cards. $c = (-a_1 - b_1, -a_2 - b_2, ..., -a_n - b_n)$. Then, for every component

$$a_i + b_i + c_i = a_i + b_i - a_i - b_i = 0$$

By Lemma 3.1, $[abc]$ is a SET. So every two Set cards have a completion card.

Now suppose $[abc]$ and $[abd]$ are both SETs. So for every component, $a_i + b_i + c_i = a_i + b_i + d_i = 0$ in $\mathbb{Z}_3$, so $c_i = d_i$ in $\mathbb{Z}_3$ for every i. Therefore, every component of c is the same as d, so c and d are the same point in $\mathbb{Z}_3^4$, and are therefore the same Set card.

So the completion card for any two cards exists and is unique. □

Corollary 4.1.1. *Every pair of 2 cards always determines exactly one SET, so the number of SETs in the 81 cards is exactly equal to the number of pairs of cards in the deck divided by 3 to account for triple counting, which is* $\frac{\binom{81}{2}}{3} = 1080$.

Lemma 4.2. *The probability of three randomly selected cards being a SET is* $\frac{1}{79}$ *in* 4*-dimensional Set.*

Proof. Suppose two cards a and b are randomly selected from the deck of cards, fixed but arbitrary. By Theorem 4.1, there exists exactly one card left in the deck that is the completion card for $\{a, b\}$. Since 79 cards remain in the deck after a and b have been selected, the probability that the completion card is selected is $\frac{1}{79}$. □

Lemma 4.3. *Each card is in* 40 *distinct SETs.*

Proof. Let us select a card a, fixed but arbitrary. We then have 80 cards remaining in the deck, so there are 80 choices for selecting a second card b. The third completion card c exists and is unique to the $\{a, b\}$ configuration by Theorem 4.1. So there are 80 $\{a, b, c\}$ configurations. However, every SET has been double counted since $\{a, b, c\}$ is the same SET as $\{a, c, b\}$, so we divide by 2 to obtain 40 as the number of distinct SETs containing card a. □

5 SET-full configurations

We can explore the maximum number of SETs that can be found in n cards. The case of 9 cards is particularly interesting, because if you imagine the 3-dimensional cube, it has faces of size 9.

Definition (9-full configuration)**.** A configuration of 9 cards with the maximum number of SETs.

Theorem 5.1. *The number of SETs in* 9 *cards never exceeds* 12*.*

Proof. In this theorem, we consider 9 cards in n-dimensional Set, where $n > 2$. If $n = 1$, there are only 3 cards in the deck so it's impossible to pick 9 cards from 3. If $n = 2$ there are 9 cards in the deck, so picking 9 cards will obviously have the maximum number of SETs.

First, we will show that there exists a configuration with 9 cards that has 12 SETs in it.

Consider the entire 2-dimensional game of Set. By Theorem 3.6, it has 12 SETs in total. To convert this configuration to higher dimension, add an attribute to all the cards, and fix the value of that attribute for each card to be the same. This way, picking any three cards that was a SET in the 2-dimensional game will still give a SET, as the new attribute is the same value in all 3 cards. So, there are still 12 SETs in this higher dimensional configuration.

Now we will show that this is the maximum number of SETs possible in 9 cards.

In the current configuration, all the cards are in 4 SETs (by 3.3). Suppose one card completes 5 SETs instead of 4, so there are a total of 13 SETs in the configuration.

Then, there must be 5 distinct configurations with 2 cards that are completed by that card, as the completion card is unique by Theorem 4.1. So, there are at least $10 + 1 = 11$ cards in the configuration. This is a contradiction, as the configuration has 9 cards. Therefore, there cannot be more than 12 SETs in a 9-full configuration. □

Corollary 5.1.1. *Note that in* 9*-full configuration, if you select any two cards, the completion card is also in the* 9*-full configuration.*

Lemma 5.2. *In a 9-full configuration, there exists a SET where the value of one attribute is the same in all three cards.*

Proof. Suppose there existed 9-full configuration with no SETs with a fixed attribute. This means that there are three SETs that have a 'different' value for one attribute.

So, there are three cards that share an attribute in common. Therefore, if we select two of those cards, their completion card also has that third attribute in common with them. By Corollary 5.1.1, that completion card is in the 9-full configuration as well. So, there exists a SET with an attribute such that the value is the same in all three cards. □

Lemma 5.3. *The cards of a 9-full configuration can be split into 3 disjoint SETs.*

Proof. A 9-full configuration has 12 SETs. Let A be a SET in that configuration.

Remove A from the configuration, so that it now has 6 cards. The cards in that SET complete $\frac{3 \cdot 6}{2} = 9$ SETs with the remaining six cards, as every pair of two cards in a 9-full configuration has a completion card in the 9-full configuration (by Corollary 5.1.1, so there are $3 \cdot 6$ SETs with A and the remaining cards. We must divide by 2 to not double count SETs. A is also a SET by itself. So, after removing A, the configuration of 6 cards has $12 - 10 = 2$ SETs within it.

Let B be one of the SETs in the configuration of 6 cards. Let C be the cards in the 9-full configuration that are not in A and B. The number of SETs completed by B in the 9-full configuration is 10 by the same reasoning used for the SETs completed by A.

We want to show that there is only 1 SET in B and not in both A and C.

Suppose it did contain a SET in B and one of A and C. If the SET is only in A and B, then there must be one card from one configuration and two from the other. However, the completion card is unique (Theorem 4.1, and A and B are both SETs, so if you take two cards from one of A and B, their completion card is in the same configuration. This means that there are no SETs in A and B and not C. This means that there are 9 SETs in B and C and not A, but that is impossible as we know there are only 2. This is a contradiction, so there must be only one SET in B that is not in both A and C, which is B itself.

Therefore, there is one SET in C $(12 - 10 - 1)$, and C has 3 cards, so it must be a SET.

So, we can split the 9-full configuration into 3 disjoint SETs, A, B, and C.

□

Corollary 5.3.1. *This tells us that we can make a 9-full configurations by picking 3 SETs.*

Lemma 5.4. *When constructing a 9-full configuration, if you choose a first SET with direction vector* $\mathbf{d}$*, then the other two SETs chosen must also have direction vector* $\mathbf{d}$.

Proof. By Corollary 5.3.1, we can construct a 9-full configuration by picking 3 SETs.

Let one SET chosen be defined by a card a_1 and a direction vector, $\mathbf{d_a}$, and its cards are $a_i; 1 \leq i \leq 3$.

Let the other SET be defined by a card b_1 and a direction vector $\mathbf{d_b}$, with cards $b_i; 1 \leq i \leq 3$ such that for all i, j, $a_i \neq b_j$, and $\mathbf{d_a} \neq \mathbf{d_b}$. By Corollary 5.1.1, the final SET contains the completion cards of all possible pairings of cards in the first and second SETs.

The completion card of a_i, b_j is a_i plus the difference between a_i and b_j which is $a_i + (a_i - b_i)$ as a formula. This is

$$(a_1 + i\mathbf{d_a}) + ((a_1 + i\mathbf{d_a}) - (b_1 + j\mathbf{d_b}))$$
$$= 2(a_1 + i\mathbf{d_a}) - (b_1 + i\mathbf{d_b})$$

For $\{a_1, b_1\}$, the completion card is $2(a_1 + \mathbf{d_a}) - (b_1 + \mathbf{d_b})$

For a_1, b_2, the completion card is $2(a_1 + \mathbf{d_a}) - (b_1 + 2\mathbf{d_b})$

For a_1, b_3, the completion card is $2(a_1 + \mathbf{d_a}) - (b_1 + 3\mathbf{d_b})$

All three of these cards are different. So, any other completion cards must be equal to one of the three cards.

The completion card of $\{a_2, b_1\}$ is $2(a_1 + 2\mathbf{d_a} - (b_1 + \mathbf{d_b}$.

This is equal to $2(a_1 + \mathbf{d_a}) - (b_1 + k\mathbf{d_b})$, where $k = 1, 2,$ or 3.

$$2(a_1 + 2\mathbf{d_a}) - (b_1 + \mathbf{d_b}) = 2(a_1 + \mathbf{d_a}) - (b_1 + k\mathbf{d_b})$$
$$\mathbf{d_a} - \mathbf{d_b} = -k\mathbf{d_b}$$
$$\mathbf{d_a} = (1 - k)\mathbf{d_b}$$

If $k = 1$, $\mathbf{d_a}$ has values of 0, which is not possible as every card in a SET is distinct.

If $k = 2$, $\mathbf{d_a} = \mathbf{-d_b}$, so by Lemma 3.5, they are the same direction.

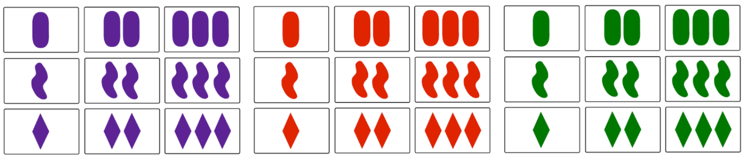
If $k = 3$, $\mathbf{d_a} = \mathbf{d_b}$, which is a contradiction, as $d_a \neq d_b$.

So, our assumption that both SETs have a different direction vector must be false. So, all three SETs in the 9-full configuration have the same direction vector. □

Theorem 5.5. *There are* $\frac{(3^{n-1})^2(3^{n-1}-1)}{4} - 3^{n-1}$ *ways to pick a 9-full configuration in n-dimensional SET.*

Proof. We will prove this by giving a method that can construct any 9-full configuration, and then count the number of ways to construct that configuration.

Consider a 3-dimensional Set deck.



We can get a 9-full configuration by methodically selecting three SETs. To pick the first SET, we can choose it from one of the 3×3 monochromatic boxes by fixing the Colour attribute (any arbitrary attribute works) within that SET, as we know by Lemma 5.2 that at least one SET has an attribute that is the same.

By Lemma 3.3, we know there are $\frac{3^n-1}{2}$ direction vectors in the deck. So, we can choose one SET of one direction as the first one. By Lemma 5.4, the other two SETs chosen must also be of that direction.

So we pick a second SET of the same direction vector. The third SET that completes the 9-full configuration is decided by the first two as the completion card between any two cards in the first two SETs is in the third SET, and is unique, so the third SET is also unique.

If we turn this into a formula, we get that the number of 9-full configurations is equal to (the number of directions) $\times$ (the number of SETs in each direction when fixing an attribute) $\times$ (the number of options for the second SET) $\times$ $\frac{1}{2}$ (to account for the double counting of picking SETs).

However, this formula triple counts 9-full configurations where one attribute is fixed throughout all the cards. There are 3^{n-2} such configurations, and they're counted three times. So, we must subtract 3^{n-1} from our formula.

We now have:

- Number of directions $= \frac{3^n-1}{2}$
- Number of SETs in each direction when fixing an attribute $= 3^{n-1}$
- Number of options for the second SET $= 3^{n-1} - 1$
- Number to subtract for triple counted SETs $= 3^{n-1}$

So, the number of ways to get a 9-full configuration in n-dimensional SET is given by:

$$\frac{3^n - 1}{2} \cdot \frac{3^{n-1} \cdot (3^{n-1} - 1)}{2} - 3^{n-1}$$
$$= \frac{(3^{n-1})^2(3^{n-1} - 1)}{4} - 3^{n-1}$$

□

Example: In 3-dimensional Set, this is 4 directions $\times$ 3 SETs in each direction $\times$ 8 second SETs $\times \frac{1}{2}$ − 9 extra configurations, which equals 39.

6 Expected number of SETs

In this section we consider the expected number of SETs in n cards, for $n \geq 3$.

Notation. we will refer to the expected number of complete SETs in n cards as $\mathbb{E}_n[S]$, where S is the number of complete SETs in some group of n cards, and:

$$\mathbb{E}_n[S] = \sum_{i=0}^{M} i \cdot \mathbb{P}(S = i)$$

where m is the maximum possible number of complete SETs in n cards.

6.1 $n = 3$

By Lemma 4.2, the probability that 3 given cards are a SET is $\frac{1}{79}$. Since three cards can only be 0 or 1 SET, $\mathbb{E}_3[S] = \frac{1}{79}$.

6.2 $n = 4$

Let $\{a, b, c, d\}$ be a configuration with 4 cards. There are four potential SETs that these 4 cards could make: $[abc]$, $[abd]$, $[acd]$, $[bcd]$. Since every potential SET shares exactly 2 cards with every other SET, it is not possible to have more than one of these triples as SETs by Theorem 4.1.

For example, if we have $[abc]$ as a SET, this implies that the completion card for $\{a, b\}$ is c, so $[abd]$ cannot be a SET. Similarly, b is be the completion card for $\{a, c\}$ and A is the completion card for $\{b, c\}$ so $[acd]$ and $[bcd]$ cannot be SETs. So with 4 cards, $M = 1$, meaning:

$$\mathbb{E}_4[S] = \sum_{i=0}^{1} i \cdot \mathbb{P}(S = i) = \mathbb{P}(S = 1)$$

Since there can only be one complete SET in the four cards, the probability of there being exactly one complete SET is equal to the probability of there being any SETs in the four cards. So:

$$\begin{aligned} \mathbb{P}(S = 1) &= \mathbb{P}([abc] \cup [abd] \cup [acd] \cup [bcd]) \\ &= \mathbb{P}([abc]) + \mathbb{P}([abd]) + \mathbb{P}([acd]) + \mathbb{P}([bcd]) \end{aligned}$$

since the occurrence of each SET is mutually exclusive to all other potential SETs. The probability of each triple being a SET is equal to $\frac{1}{79}$, by Lemma 4.2. So:

$$\mathbb{P}([abc]) + \mathbb{P}([abd]) + \mathbb{P}([acd]) + \mathbb{P}([bcd]) = \frac{1}{79} + \frac{1}{79} + \frac{1}{79} + \frac{1}{79} = \frac{4}{79}$$

Hence,

$$\mathbb{E}_4[S] = \frac{4}{79}$$

6.3 $n = 5$

Suppose we now have 5 cards: $\{a, b, c, d, e\}$. Here, there could be up to 2 complete SETS. To show this, suppose *WLOG* $[abc]$ is a SET. If there is another SET in the 5 cards, it can only share at most 1 card with $[abc]$, for the same reasoning as in $n = 4$, and therefore must contain d and e. So *WLOG* suppose $[ade]$ is another SET. If a third SET was in the 5 cards, it would have to contain 2 cards from one of the two existing SETs. So there can only be at most 2 SETs, meaning $M = 2$. So:

$$\mathbb{E}_5[S] = \sum_{i=0}^{2} i \cdot \mathbb{P}(S = i) = 2 \cdot \mathbb{P}(S = 2) + \mathbb{P}(S = 1)$$

2 SETs

To find $\mathbb{P}(S = 2)$, we have to consider two separate cases: the case where there is a SET within $\{a, b, c, d\}$, and the case where there is no SET within $\{a, b, c, d\}$. If there is a SET within $\{a, b, c, d\}$, then the addition of e must add 1 SET for the total number of SETs to be 2. If there is no SET within $\{a, b, c, d\}$, then the addition of e must add 2 SETs for the total number to be 2. Note there cannot be more than 1 SET in $\{a, b, c, d\}$, by the same reasoning as outlined in the $n = 4$ section. Therefore:

$$\mathbb{P}(S = 2) = \mathbb{P}(e \text{ adds one SET} \cap \text{SET in } \{a, b, c, d\}) + \mathbb{P}(e \text{ adds two SETs} \cap \text{no SET in } \{a, b, c, d\}$$

And by Bayes' Theorem,

$$\begin{aligned}\mathbb{P}(S = 2) = \mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a, b, c, d\}) \cdot \mathbb{P}(\text{SET in } \{a, b, c, d\}) + \\ \mathbb{P}(e \text{ adds two SETs} \mid \text{no SET in } \{a, b, c, d\}) \cdot \mathbb{P}(\text{no SET in } \{a, b, c, d\})\end{aligned}$$

From the $n = 4$ case, we know that $\mathbb{P}(\text{SET in } \{a, b, c, d\}) = \frac{4}{79}$ and therefore $\mathbb{P}(\text{no SET in } \{a, b, c, d\}) = \frac{75}{79}$. So we are left with two terms.

First, we will calculate $\mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a, b, c, d\})$. Suppose we have $\{a, b, c, d\}$ with one complete SET. Without loss of generality, let $[abc]$ be the complete SET. From Theorem 4.1, there exists exactly one SET containing $\{a, d\}$, $\{b, d\}$, and $\{c, d\}$ respectively, so let $[adx]$, $[bdy]$, and $[cdz]$ be SETs, where x, y, and z are unique. Notice that $x \neq y$, $y \neq z$, and $z \neq x$ since no two SETs can share two cards that are the same. So the probability that e adds one more SET is equal to the probability that e is in $\{x, y, z\}$. Since $a, b, c, d \notin \{x, y, z\}$, 77 cards remain from the 81 in the deck that could possible be one of x, y, or z, so:

$$\mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a, b, c, d\}) = \mathbb{P}(e \in \{x, y, z\}) = \frac{3}{77}$$

Now, we only need to find $\mathbb{P}(e \text{ adds two SETs} \mid \text{no SET in } \{a, b, c, d\})$. Suppose there are not SETs in $\{a, b, c, d\}$. If e adds 2 SETs, then e must be the completion card for 2 pairs of cards from the 4. These pairs are either $\{\{a, b\}, \{c, d\}\}$, $\{\{a, c\}, \{b, d\}\}$, or $\{\{a, d\}, \{b, c\}\}$. We will only consider $\{\{a, b\}, \{c, d\}\}$, and then multiply our result by 3.

$$\begin{aligned}\mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a, b, c, d\}) &= 3 \cdot \mathbb{P}([cde] \cap [abe]) \\ &= 3 \cdot \mathbb{P}([cde] \mid [abe]) \cdot \mathbb{P}([abe])\end{aligned}$$

Since the completion card for $\{a, b\}$ cannot be a, b, c, or d, there are 77 choices for which card it could be, so $\mathbb{P}([abe]) = \frac{1}{77}$.

Now suppose $[abe]$ is a SET. Let $[acx]$, $[bcy]$, and $[cez]$ be SETs. We want to find the probability that $d = z$. Note $x \neq y$ since no 2 SETs can share 2 of the same cards. Also, $d \neq x$ and $d \neq y$, since there are no SETs in $\{a, b, c, d\}$ by our assumption. Also, since $[acx]$ is a SET, $[cex]$ cannot be a SET, since no 2 SETs can share 2 cards, so $z \neq x$. Similarly, $[cey]$ cannot be a SET, so $z \neq y$.

From this point, we can partition the deck into two groups: $\{a, b, c, e, x, y\}$ and the 75 remaining cards. d and z are not in the first group, so both are in the 75 remaining cards. So $\mathbb{P}(d = z) = \frac{1}{75}$. Therefore:

$$\begin{aligned}\mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a, b, c, d\}) &= 3 \cdot \mathbb{P}([cde] \mid [abe]) \cdot \mathbb{P}([abe]) \\ &= 3 \cdot \frac{1}{75} \cdot \frac{1}{77} \\ &= \frac{3}{75 \cdot 77}\end{aligned}$$

Returning to the original equation:

$$\begin{aligned}\mathbb{P}(S=2) = &\ \mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a,b,c,d\}) \cdot \mathbb{P}(\text{SET in } \{a,b,c,d\}) + \\ &\ \mathbb{P}(e \text{ adds two SETs} \mid \text{no SET in } \{a,b,c,d\}) \cdot \mathbb{P}(\text{no SET in } \{a,b,c,d\}) \\ &= \left[\frac{3}{77} \cdot \frac{4}{79}\right] + \left[\frac{3}{75 \cdot 77} \cdot \frac{75}{79}\right] \\ &= \frac{12}{77 \cdot 79} + \frac{3}{77 \cdot 79} = \frac{15}{77 \cdot 79}\end{aligned}$$

1 SET

To find $\mathbb{P}(S=1)$, we again have to consider the two cases where there is a SET within $\{a,b,c,d\}$ and where there is no SET within $\{a,b,c,d\}$. If there is a SET within $\{a,b,c,d\}$, then the addition of e must not add a SET to keep the number of SETs to 1. If there is no SET within $\{a,b,c,d\}$, then the addition of e must add 1 SET. Therefore:

$$\begin{aligned}\mathbb{P}(S=1) &= \mathbb{P}(e \text{ adds no SET} \cap \text{SET in } \{a,b,c,d\}) + \mathbb{P}(e \text{ adds a SET} \cap \text{no SET in } \{a,b,c,d\} \\ &= \mathbb{P}(e \text{ adds no SET} \mid \text{SET in } \{a,b,c,d\}) \cdot \mathbb{P}(\text{SET in } \{a,b,c,d\}) + \\ &\quad \mathbb{P}(e \text{ adds a SET} \mid \text{no SET in } \{a,b,c,d\}) \cdot \mathbb{P}(\text{no SET in } \{a,b,c,d\})\end{aligned}$$

As above, $\mathbb{P}(\text{SET in } \{a,b,c,d\}) = \frac{4}{79}$ and $\mathbb{P}(\text{no SET in } \{a,b,c,d\}) = \frac{75}{79}$.
From the $S=2$ case,

$$\begin{aligned}\mathbb{P}(e \text{ adds no SET} \mid \text{SET in } \{a,b,c,d\}) &= 1 - \mathbb{P}(e \text{ adds one SET} \mid \text{SET in } \{a,b,c,d\}) \\ &= 1 - \frac{3}{77} = \frac{74}{77}\end{aligned}$$

The only remaining term is $\mathbb{P}(e \text{ adds a SET} \mid \text{no SET in } \{a,b,c,d\})$.

Suppose there are no SETs in $\{a,b,c,d\}$. If we add e, we could have 6 possible pairs ($= \binom{4}{2}$) that make a SET with e. We will only consider $[abe]$ without loss of generality, and then multiply our result by 6. However, if e adds only one SET, we also require $[cde]$ to not be a SET. So:

$$\begin{aligned}\mathbb{P}(e \text{ adds a SET} \mid \text{no SET in } \{a,b,c,d\}) &= 6 \cdot \mathbb{P}([abe] \cap \neg[cde] \\ &= 6 \cdot [\mathbb{P}([abe]) - \mathbb{P}([abe] \cap [cde])]\end{aligned}$$

The completion card for $\{a,b\}$ cannot be a, b, c, or d, so there are 77 remaining options. So $\mathbb{P}([abe]) = \frac{1}{77}$.
Also, from the $S=2$ case, we know $\mathbb{P}([abe] \cap [cde]) = \frac{1}{75 \cdot 77}$, so:

$$\mathbb{P}(e \text{ adds a SET} \mid \text{no SET in } \{a,b,c,d\}) = 6 \cdot \left[\frac{1}{77} - \frac{1}{75 \cdot 77}\right] = \frac{6 \cdot 74}{75 \cdot 77}$$

Returning to the original equation,

$$\begin{aligned}\mathbb{P}(S=1) = &\ \mathbb{P}(e \text{ adds no SET} \mid \text{SET in } \{a,b,c,d\}) \cdot \mathbb{P}(\text{SET in } \{a,b,c,d\}) + \\ &\ \mathbb{P}(e \text{ adds a SET} \mid \text{no SET in } \{a,b,c,d\}) \cdot \mathbb{P}(\text{no SET in } \{a,b,c,d\}) \\ &= \left[\frac{74}{77} \cdot \frac{4}{79}\right] + \left[\frac{6 \cdot 74}{75 \cdot 77} \cdot \frac{75}{79}\right] \\ &= \frac{10 \cdot 74}{77 \cdot 79}\end{aligned}$$

Therefore, when we combine the two cases:

$$\begin{aligned}\mathbb{E}_5[S] &= 2\cdot\mathbb{P}(S=2)+\mathbb{P}(S=1)\\ &= 2\cdot\frac{15}{77\cdot 79}+\frac{10\cdot 74}{77\cdot 79}\\ &= \frac{770}{77\cdot 79}=\frac{10}{79}\end{aligned}$$

We can also calculate the probability of a SET in the 5 cards by adding $\mathbb{P}(S=1)$ to $\mathbb{P}(S=2)$. So:

$$\mathbb{P}(S\geq 1)=\mathbb{P}(S=1)+\mathbb{P}(S=2)=\frac{15}{77\cdot 79}+\frac{10\cdot 74}{77\cdot 79}=\frac{755}{6083}$$

6.4 Cases of more cards

As n gets larger, the calculations involved in finding $\mathbb{E}_n(S)$ quickly become much more complicated and intricate, with many more subtle conditional implications and more cases to consider. This makes finding exact values very difficult without resorting to brute force computation. However, a useful approximation is to consider all triples within the n cards, of which there are $\binom{n}{3}$, and the probability that each is independently a SET, which is $\frac{1}{79}$. So:

$$\mathbb{E}_n(S)\approx\binom{n}{3}\cdot\frac{1}{79}$$

However, this ignores the overlap between different SETs as it treats the occurrence of each SET as independent events.

7 Computational Methods

7.1 Motivation

The motivation to apply computational methods to SET comes from the complexity of the structure. When calculating expected values for the number of SETs in n cards, it became clear that the systems in question quickly become highly complex, with far too many cases to consider. Therefore, as more cards are introduced to a configuration, it becomes increasingly difficult to make statements about quantitative properties with any degree of certainty. It was necessary to develop efficient computational methods in order to generate meaningful results like expected number of SETs and n-optimal SET-free configurations. Writing an intelligent algorithm would also give us more insight into the structure behind SET than calculating probabilities by hand.

7.2 Finding the Expected Value

We have proved the expected values of SETs in n cards by combinatorics, but we used brute-force programs to calculate expected values to verify our results. This was possible as the numbers we considered were relatively low, so it was computationally easy to check every single configuration of n cards. In reality, the equivalence approach detailed in the next section could be used to optimize this algorithm.

The idea is to have functions which tell you how many SETs exist within a configuration. So we first constructed a function which would tell us if three cards make a SET by summing their components together and checking if the result has each component equal to 0 modulo 3 (works by Lemma 3.1). Then, to find how many SETs exist within a configuration we just have to take all 3-card combinations of the configuration and check each individually. So the algorithm for computing the expected number of SETs in n cards is as follows:

- Generate all n-card configurations from the deck of 81 (this was done recursively).
- Iterate through all 3-card combinations within the n-card configuration and check if each combination is a SET by checking if the sum of all components is 0 modulo 3.

- Add one to a running tally every time a SET appears.
- Finally divide the number of SETs found by the total number of n-card configurations.

As n gets larger, the number of configurations that must be checked grows very quickly, so for larger n the expected value can be approximated by randomly generating n-card configurations instead of checking every possible configuration.

7.3 Brute-forcing the search for n-optimal SET-free Configurations in Dimension Three and Four

We first note the issue of brute-force in this problem. In three dimensions, to check if there exists a configuration with no SET you must in the worst case check around $\binom{27}{n}$ configurations. In three dimensions this is $\binom{27}{9} = 4686825$ combinations. In addition you must also check for the presence of a SET and this is going to be done in $\binom{9}{3} = 84$ combinations. So in total you must perform 393693300 co-linearity tests. Since a co-linearity test may consist of 9 additions and 3 reductions $\mod 3$, you will have far over a billion operations needing to be done. If this wasn't bad enough, you must also check that within 10 cards there always exists a SET, and this requires even more operations. Now a billion operations is not too bad, and is something a modern computer can handle given a bit of time. But if you raise now the problem to 4 dimensions, and perform the same analysis, the result is that you will have to perform over 5×10^{21} co-linearity tests which is computationally unfeasible.

So it clear that a more efficient procedure will be necessary to determine the 4-optimal SET-free configuration.

7.4 A more efficient procedure to determine the 4-optimal SET-free configuration

The main issue to tackle is finding a method of drastically reducing cases while still examining every possible case. To do this, we established a notion of equivalence between configurations of cards. For example, the fact that $(0, 0, 0, 0)$, $(0, 0, 0, 1)$, and $(0, 0, 0, 2)$ is a SET is implied directly by the fact that $(1, 1, 1, 1)$, $(1, 1, 2, 1)$, and $(1, 1, 0, 1)$ is a SET. This amounts to simply a translation, and all SETs remain SETs under translation. Similarly, all non-SETs remain non-SETs under translation. It is also true that the same holds for all linear transformations. Both facts are proved later in the section, but they allow us to construct equivalence classes of configurations, and then examine only one representative from each, which reduces the number of cases that need to analysed dramatically. The method for computing n-optimal SET-free configurations is outlined below:

- Generate all 3 card configurations and remove all that contain SETs.
- Fix one configuration and search for all its equivalent configurations. Remove all of them and leave only the original configuration, which is now 'unique'.
- Add every possible card to all remaining configurations and repeat the process, removing all configurations with SETs.
- Then remove all equivalent configurations, leaving only one representative for each equivalence class.
- Repeat this process, continually adding every possible card and removing configurations with SETs and all but one configuration in each equivalence class.
- Eventually, the addition of any card to the remaining configuration will create a SET, so we take the size of this final configuration (there may be multiple, in which case we take any of them), and this is the size of the largest SET-free configuration possible. So we have found the n-optimal SET-free configuration.

The details and precise justifications for this process are given in this section. First, a few definitions:

Definition. We define A_n, the set of all configurations of size n, as the set $A_n \subset (\mathbb{Z}_3^d)^n$, such that each element of A_n consists of only distinct elements.

Definition. We define $\overline{A_n}$, the set of all SET-free configurations of size n, as the set $\overline{A_n} \subset A_n$, such that every element of $\overline{A_n}$ contains no SET, and such that every element of $A_n \setminus \overline{A_n}$ contains at least one SET.

Definition. We define two configurations of n cards over d dimensions $C_1, C_2 \in A_n$ to be equivalent iff the number of SETs in C_1 is the same as the number of SETs in C_2 and there exists a bijective mapping $\phi : \mathbb{Z}_3^d \mapsto \mathbb{Z}_3^d$ with the following properties:

- Every element of C_1 is mapped to an element of C_2.
- When a new card a is added to C_1, if we add $\phi(a)$ to C_2, then for $x_1, x_2 \in C_1$, a, x_1, x_2 form a SET iff $\phi(a), \phi(x_1), \phi(x_2)$ form a SET.

Definition. We write $\bar{0}$ to denote the origin in $\mathbb{Z}_3^d$.

Definition. An equivalence mapping ϕ is a mapping such that a is equivalent to $\phi(a)$.

Theorem 7.1. *A translation $T : \mathbb{Z}_3^d \mapsto \mathbb{Z}_3^d$ defined by $T(x) = x + b$ for some fixed $b \in \mathbb{Z}_3^d$ is an equivalence mapping.*

Proof. Firstly, we show that T is a bijective mapping.

Injective: Suppose $x, y \in \mathbb{Z}_3^d$ and let $x + b = y + b$.

$$(x + b) + (-b) = (y + b) + (-b)$$

So $x = y$. So T is injective.

Surjective: Let $y \in \mathbb{Z}_3^d$. Then if we define $x = y - b$, then $x + b = y$ so x maps to y under T. Hence T is also surjective. Hence T is bijective.

Now let us define two n card configurations $X, Y \in A_n$ such that $Y = T(X)$. Then every element of X is mapped to an element of Y by bijectivity. Consider now the addition of a new card a to X, and the addition of a new card $T(a)$ to Y. We now show that for any $x_1, x_2 \in X$, if a, x_1, x_2 form a SET, then so too do $T(a), T(x_1), T(x_2)$. The three cards form a SET if $a + x_1 + x_2 = \bar{0}$ in $\mathbb{Z}_3^d$. Then:

$$\begin{aligned} T(a) + T(x) + T(y) &= (a + b) + (x + b) + \\ &= a + x_1 + x_2 + 3b \\ &= a + x_1 + x_2 = \bar{0} \end{aligned}$$

since $3b = 0b = \bar{0}$ in $\mathbb{Z}_3^d$. Hence X is equivalent to Y so X is equivalent to $T(X)$. □

Definition. We define $\mathcal{M}_d$ to be the set of all $d \times d$ matrices with elements in $\mathbb{Z}_3$ such that $\forall M \in \mathcal{M}_d$ $\det(M) \neq 0 \pmod 3$ (every element of $\mathcal{M}_d$ is invertible.)

Theorem 7.2. *A matrix $M \in \mathcal{M}_d$ can be given as a linear transformation $L : \mathbb{Z}_3^d \mapsto \mathbb{Z}_3^d$. Then L is an equivalence mapping.*

Proof. Firstly we show that L corresponding to $M \in \mathcal{M}_d$ is bijective.

Injective: We have $\det(M) \neq 0 \pmod 3$, hence M^{-1} exists. Therefore L^{-1} exists. Suppose $L(x) = L(y)$. So $L^{-1}L(x) = L^{-1}L(y)$, which means $x = y$. Hence, L is injective.

Therefore, by rank-nullity, injectivity implies $\dim(\text{null}(L)) = 0$, since if we suppose that $v \in \text{null}(L)$, then $L(v) = \bar{0}$. But by definition, $L(\bar{0}) = \bar{0}$, so $L(v) = L(\bar{0})$. By injectivity, $v = \bar{0}$

$$d = \dim(Z_3^d) = \dim(\text{null}(L)) + \dim(\text{range}(L)) = 0 + \dim(\text{range}(L))$$

So $\dim(\text{range}(L)) = d$, and therefore L is surjective. Hence L is bijective.

Now let $X, Y \in A_n$ such that $Y = L(X)$. Then every element of X is mapped to an element of Y by bijectivity. So we need only show that a, x_1, x_2 form a SET iff $L(a), L(x_1), L(x_2)$ form a SET.

Let $a + x_1 + x_2 = \bar{0}$. Then we note by additivity of a linear transformation, if we let basis vectors of L be given by $v_1, ..., v_n$, and $a = a_1 v_1 + a_2 v_2 + ... + a_n v_n$, then $L(a) = a_1 L(v_1) + ... + a_n L(v_n)$.

Hence,

$$L(a) + L(x_1) + L(x_2) = (a_1 + (x_1)_1 + (x_2)_1)L(v_1) + ... + (a_n + (x_1)_n + (x_2)_n)L(v_n)$$

But by $a + x_1 + x_2 = \bar{0}$, we must have $a_i + (x_1)_i + (x_2)_i = 0$, this gives $L(a) + L(x_1) + L(x_2) = \bar{0}$.

Let $L(a)+L(x_1)+L(x_2) = \bar{0}$. By additivity of L (it being a linear transformation), $L(a+x_1+x_2) = \bar{0}$. By $\dim \text{null}(L) = 0$ we conclude $a + x_1 + x_2 = \bar{0}$.

Hence we conclude that a, x_1, x_2 form a SET iff $L(a), L(x_1), L(x_2)$ form a SET, thereby showing that L is an equivalence mapping.

□

Definition. Given $C_n \in (\mathbb{Z}_3^d)^n$ an n card configuration, we define $Eq(C_n)$ to be the set of all possible n card configurations $D_n \in (\mathbb{Z}_3^d)$ such that C_n is equivalent to D_n under translations and linear transformations as outlined above. So $\forall c \in C_n$, $\exists d \in D_n$ such that $d = Mc + b$ for some fixed $M \in \mathcal{M}_d$ and $B \in \mathbb{Z}_3^d$.

Definition. A representative is a card in $Eq(C_n)$ that is fixed, arbitrary, and unique.

Theorem 7.3. *There is only one 3-card configuration without a SET up to equivalence.*

Proof. Begin with all configurations of 3 cards (c_1, c_2, c_3) with no SETs present. Since by Theorem 7.1 all translations are equivalent mappings, *WLOG* we can apply some translation mapping c_1 to the zero element $\bar{0} \in Z_3^d$. For any $M \in \mathcal{M}_d$, $M\bar{0} = \bar{0}$ since M is a linear transformation. So we therefore need only consider in what ways the two vectors $c_2, c_3 \in Z_3^d$ can map under after multiplication by M.

We hypothesise that for $d \geq 3$, for any two vectors $v_1, v_2, \in Z_3^d$ where $(\bar{0}, v_1, v_2)$ is not a SET, v_1, v_2 can map to any other two non-zero vectors u_1, u_2 under a suitable linear transformation M, where $(\bar{(0)}, u_1, u_2)$ is also not a SET. This would imply that any 3 card configuration without a SET can be mapped to any other card configuration without a SET with only a translation to fix the first card to the zero element and a linear transformation. That is, all 3 card configurations without a SET are equivalent. This would establish the result.

We perform now an induction on d

Claim: For $d \geq 3$, there is only one 3 card configuration without a SET, up to equivalence.

Base case: $d = 3$. The base case can be verified by computation.

Inductive Assumption: Now assume the hypothesis holds for $d = n$. Consider now an $(n + 1) \times$

$(n+1)$ matrix N, and also the $n \times n$ matrix N' nestled inside N, where $N' \in \mathcal{M}_n$.

$$N = \begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_{n+1} \\ 0 & x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} N' = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix}$$

Since $N' \in \mathcal{M}_n$, $\det(N') \neq 0$ and therefore $\det(N) \neq 0$ as $det(N) = a_1 \cdot \det(N)$ (we take $a_1 \neq 0$). Suppose we have 3 cards with $n+1$ components ($\bar{0}$, c_1, c_2) without a SET and we want to map to some other three cards ($\bar{0}$, d_1, d_2) without a SET. Then N' will map components 2 through to $n+1$ of c_1 and c_2 to the corresponding components of d_1 and d_2, by our inductive assumption. So we just need to show that the first components $(c_1)_1$ and $(c_2)_1$ can be mapped to $(d_1)_1$ and $(d_2)_1$ respectively. Therefore, $a_1, a_2, ..., a_{n+1}$ must be chosen such that:

$$\begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_{n+1} \\ 0 & x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \begin{bmatrix} (c_1)_1 \\ (c_1)_2 \\ \vdots \\ (c_1)_{n+1} \end{bmatrix} = \begin{bmatrix} (d_1)_1 \\ (d_1)_2 \\ \vdots \\ (d_1)_{n+1} \end{bmatrix}$$

And:

$$\begin{bmatrix} a_1 & a_2 & a_3 & \dots & a_{n+1} \\ 0 & x_{11} & x_{12} & \dots & x_{1n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & x_{n1} & x_{n2} & \dots & x_{nn} \end{bmatrix} \begin{bmatrix} (c_2)_1 \\ (c_2)_2 \\ \vdots \\ (c_2)_{n+1} \end{bmatrix} = \begin{bmatrix} (d_2)_1 \\ (d_2)_2 \\ \vdots \\ (d_2)_{n+1} \end{bmatrix}$$

So:

- $a_1 \neq 0$
- $a_1(c_1)_1 + a_2(c_1)_2 + ... + a_{n+1}(c_1)_{n+1} = (d_1)_1$
- $a_1(c_2)_1 + a_2(c_2)_2 + ... + a_{n+1}(c_2)_{n+1} = (d_2)_1$

If there exists some $i > 1$ such that $(c_1)_i \neq (c_2)_i$, then we may set all $a_j = 0$ for $j > 0, j \neq i$, such that the equations become

- $a_1(c_1)_1 + a_i(c_1)_i = (d_1)_1$
- $a_1(c_2)_1 + a_i(c_2)_i = (d_2)_1$

Hence:

$$a_i((c_1)_i - (c_2)_i) = (d_1)_1 - (d_2)_1 + a_1(c_2)_1 - a_1(c_1)_1$$

We can fix $a_1 = 1$ so

$$a_i((c_1)_i - (c_2)_i) = (d_1)_1 - (d_2)_1 + (c_2)_1 - (c_1)_1$$

Since $(c_1)_i \neq (c_2)_i$, $(c_1)_i - (c_2)_i \neq 0$ so this has a multiplicative inverse in $\mathbb{Z}_3$. When we multiply both sides by this inverse, we obtain a value for a_i, so we can always construct some a_i such that N maps ($\bar{(}0)$, c_1, c_2) to ($\bar{(}0)$, d_1, d_2). So the claim holds.

In the case that no such i exists, we instead consider an $(n+1) \times (n+1)$ matrix of the form:

$$\begin{bmatrix} x_{11} & x_{12} & \dots & x_{1n} & 0 \\ x_{21} & x_{22} & \dots & x_{2n} & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{nn} & 0 \\ a_1 & a_2 & a_3 & \dots & a_{n+1} \end{bmatrix}$$

It cannot be that $(c_1)_1 = (c_2)_1$, or otherwise $c_1 = c_2$ but these two cards are distinct. Therefore, in this case there exists an i, namely $i = 1$, such that $(c_1)_1 \neq (c_2)_1$. In this case, the condition becomes that $a_{n+1} \neq 0$ instead of a_1, so by a similar argument we can find suitable values for $a_1, a_2, ...a_n$. So the holds for this case as well.

Therefore, all configurations of 3 cards without a SET are equivalent to each other, and so can be represented by a single configuration. □

Definition. We define the collapse of an n card configuration $C_n \in A_n$ as $\textbf{Col}(C_n) := \{X \in Eq(C_n)\}$, where X is the representative given by the first element in the lexicographic ordering of $Eq(C_n)$.

Definition. We define the addition of an n card configuration $C_n \in A_n$ as $\textbf{Add}(C_n) := \{C_n \cup \{x\} : x \in \mathbb{Z}_3^d, x \notin C_n\}$. The set $\textbf{Add}(C_n)$ contains all sets containing all cards from C_n and one additional different card in $\mathbb{Z}_3^d$. We define $\textbf{Add}(A)$ on a set of configurations A to be the set $\bigcup_{C \in A} \textbf{Add}(C)$

Definition. We define the removal of a set of configurations A as: $\textbf{Rem}(A) := \{C : C \in A, C$ does not contain a SET$\}$.

Definition. We define the prune of a set of configurations A as $\textbf{Prune}(A) := \textbf{Rem}(\textbf{Add}(\textbf{Col}(A)))$. So the process of pruning a set of configurations collapses all equivalent configurations into one representative, then creates every possible configuration with one extra card, and finally removes all configurations containing a SET.

Definition. We define the completion of a set of configurations as $\textbf{Com}(A_n) := \{MC+b : C \in A_n, M \in \mathcal{M}_d, b \in \mathbb{Z}_3^d\}$. Here, $MC + b = \{Mc + b : c \in C\}$. The completion process can be thought of as 'undoing' the collapsing process for each configuration in A_n, generating the entire equivalence class from one representative.

Theorem 7.4. *The set of all SET-free configurations of* $n+1$ *cards* $\overline{A_{n+1}}$ *can be generated from* $\overline{A_n}$ *by pruning and then taking the completion. So* $\overline{A_{n+1}} := \textbf{Com}(\textbf{Prune}(\overline{A_n}))$

Proof. We will show that $\textbf{Com}(\textbf{Prune}(\overline{A_n})) \subset \overline{A_{n+1}}$.

Since **Rem** removes all configurations containing SETs, there can be no SETs within $\textbf{Prune}(\overline{A_n})$. By our Theorems 7.1 and 7.2, **Com** is an equivalence mapping. So, since there are no SETs within $\textbf{Prune}(\overline{A_n})$, there must be no SETs within $\textbf{Com}(\overline{A_n})$.

Since **Add** increases the size of all configurations in $\overline{A_n}$ from n to $n+1$, and **Col** and **Rem** do not change the size of the configurations, $\textbf{Prune}(\overline{A_n})$ must contain configurations of size $n+1$. **Com** conserves the size of SETs size translations and linear transformations are injective from Theorems 7.1 and 7.2. So, as $\overline{A_{n+1}}$ contains every possible SET-free configuration of size $n+1$, $\textbf{Com}(\textbf{Prune}(\overline{A_n})) \subset \overline{A_{n+1}}$.

We will now show that $\overline{A_{n+1}} \subset \textbf{Com}(\textbf{Prune}(\overline{A_n}))$. We must show that if $C \in \overline{A_{n+1}}$ then $C \in \textbf{Com}(\textbf{Prune}(\overline{A_n}))$. To do this, write $C = \{c_1, c_2, ..., c_{n+1}\}$. Since C is SET-free by it being in $\overline{A_{n+1}}$, so is $C' = \{c_1, c_2, ..., c_n\}$, and so $C' \in \overline{A_n}$. Let $D = MC' + b$ where $M \in \mathcal{M}_d$ be the representative of $Eq(C')$ contained within $\textbf{Col}(\overline{A_n})$. Then we show $C' \in \textbf{Com}(\textbf{Prune}((\overline{A_n}))$ by showing that $D \in \textbf{Com}(\textbf{Rem}(\textbf{Add}(D)))$.

During the pruning processing, every possible $(n+1)$ card configuration containing D is generated, so at least one configuration that is equivalent to C will be generated (namely the configuration with $Mc_{n+1} + b$ added). Let this configuration be C_e. Since $C \in \textbf{Com}(\textbf{Prune}(\overline{A_n})) \subset \overline{A_{n+1}}$, C is SET-free. So C_e is SET-free. Therefore, the **Rem** process will not remove C_e.

So we want to show $C \in \textbf{Com}(\textbf{Add}(D))$, so we want:

$$C \in \textbf{Com}((MC' + b) \cup \{\omega\} : \omega \in \mathbb{Z}_3^d, \omega \notin (MC' + b))$$

That is, $\exists \omega \in \mathbb{Z}_3^d, \exists T \in \mathcal{M}_d, \exists d \in \mathbb{Z}_3^d$ such that $T((MC' + b) \cup \{\omega\}) + d = C$. We can choose $T = M^{-1}$, which exists because $M \in \mathcal{M}_d$ so $det(M) \neq 0$.

So:

$$\begin{aligned} T((MC'+b) \cup \{\omega\}) + d &= (M^{-1}(MC'+b) \cup M^{-1}\{\omega\}) + d \\ &= ((C' + M^{-1}b) \cup M^{-1}\omega) + d \end{aligned}$$

Therefore, we choose $d = -M^{-1}b$, which means:

$$\begin{aligned} ((C' + M^{-1}b) \cup M^{-1}\omega) + d &= ((C' + M^{-1}b) \cup M^{-1}\omega) - M^{-1}b \\ &= (C' \cup \{M^{-1}\omega - M^{-1}b\}) \end{aligned}$$

Therefore, we want $M^{-1}(\omega - b) = c_{n+1}$. Hence $\omega - b = Mc_{n+1}$, and $\omega = b + Mc_{n+1}$.

Therefore, we have shown that $C \in \textbf{Com}(\textbf{Rem}(\textbf{Add}(D)))$ and $D \in \textbf{Col}(\overline{A_n})$, which implies that $C \in \textbf{Com}(\textbf{Prune}(\overline{A_n}))$. We have therefore shown that $\overline{A_{n+1}} \subset \textbf{Com}(\textbf{Prune}(\overline{A_n}))$.

$\textbf{Com}(\textbf{Prune}(A_n)) \subset \overline{A_{n+1}}$ and $\overline{A_{n+1}} \subset \textbf{Com}(\textbf{Prune}(A_n))$. So therefore $\overline{A_{n+1}} = \textbf{Com}(\textbf{Prune}(\overline{A_n}))$ □

Theorem 7.5. *Let C be an arbitrary 3 card configuration without any SETs. Denote $\textbf{Prune}^n(S) = \textbf{Prune}(\textbf{Prune}^{n-1}(S))$ for $n \in \mathbb{N}_0$ with $\textbf{Prune}^0(S) = S$. Then $\overline{A_{n+3}} = \textbf{Com}(\textbf{Prune}^n(C))$*

Proof. The proof follows by induction.

Base case: Let $n = 0$. Then by Theorem 7.3, $\overline{A_3} = \textbf{Com}(C)$, since there is only one equivalence class for 3 card SET-free configurations.

Inductive Assumption: Assume the result holds for $n = k$, so $\overline{A_{k+3}} = \textbf{Com}(\textbf{Prune}^k(C))$. Note that for an arbitrary configuration T, $\textbf{Col}(\textbf{Com}(T)) = \textbf{Col}(T)$, since **Com** generates the entire equivalence class of T, and **Col** collapses the equivalence class down to one representative. So $\textbf{Prune}(T) = \textbf{Prune}(\textbf{Com}(T))$, as by definition $\textbf{Prune}(T) = \textbf{Rem}(\textbf{Add}(\textbf{Col}(T))) = \textbf{Rem}(\textbf{Add}(\textbf{Col}(\textbf{Com}(T)))) = \textbf{Prune}(\textbf{Com}(T))$

Then consider the expression:

$$\begin{aligned} \textbf{Com}(\textbf{Prune}^{k+1}(C)) &= \textbf{Com}(\textbf{Prune}(\textbf{Prune}^k(C))) \\ &= \textbf{Com}(\textbf{Prune}(\textbf{Com}(\textbf{Prune}^k(C)))) \\ &= \textbf{Com}(\textbf{Prune}(\overline{A_{k+3}})) \end{aligned}$$

by our inductive hypothesis. But by Theorem 7.4, this is equal to $\overline{A_{k+4}}$. Therefore, the inductive hypothesis is complete.

□

From the above procedure, we were able to show the following interesting results through computation, the specifics of which are detailed in Section 7.5:

- The 3-optimal SET-free configuration is of size 9.
- The 4-optimal SET-free configuration is of size 20.
- Both the 3-optimal and 4-optimal SET-free configurations are unique up to equivalence.

A closing remark is that the choice of equivalence mapping involved in **Col** $MC + b$ is somewhat (but not at all entirely) arbitrary. It is reasonable to suggest that there may exist some set of equivalent mappings other than translations and linear transformations which could be used instead. Our method using linear transformations and translations requires a lot of computational case checking in order to determine whether two configurations are equivalent. All possible matrices and translation vectors have to be exhausted. There may exist a more efficient equivalence mapping with a smaller pool of possibilities that require checking.

7.5 Application to computer programming

We will now give a description of how the Theorems above can be applied into computer programs to generate SET-free configurations.

We give to each card in $\mathbb{Z}_3^d$ a 'size' given by the following function:

```
def size(card):
    sum = 0
    for i in range(0,d):
        sum += (3^i)*card[i]
    return sum
```

Then a configuration is ordered from smallest to largest based on the sizes of its cards, using the function above. This gives each configuration a unique identifier, called a hash.

```
def getHash(configuration)
    hashString = ""
    for cardSize in sorted([size(card) for card in configuration]:
        hashString = hashString + to_string(cardSize) + ":"
    return hashString
```

The implementation of both **Add** and **Rem** is straightforward, and so only the implementation of **Col** shall be included. The array *space* refers to cards, or elements of $\mathbb{Z}_3^d$. By the array *matrices*, we refer to elements of $\mathcal{M}_d$

```
def collapse(configurations)
    hashes = set()
    transforms = set()
    unique = set()
    for configuration in configurations:
        for b in space:
            hsh = getHash(configuration + b)
            if hsh in hashes:
                break
            hashes.add(hsh)
        else:
            transforms.add(configuration)
    hashes = set()
    for configuration in transforms:
        local = set()
        for M in matrices:
            hsh = getHash(matrixMultiply(M, configuration))
            if hsh in hashes:
                break
            local.add(hsh)
```

```
        else:
            unique.add(configuration)
        hashes = hashes.union(local)
    return unique
```

We define a prune as expected, by combining collapse, add, and remove. The final process of complete is defined as:

```
def complete(configurations)
    allConfigurations = set()

    for b in space:
        for M in matrices:
            allConfigurations.add(matrixMultiply(M, configuration) + b)

    return allConfigurations
```

To complete the algorithm outlined at the beginning of Section 7.4, we then simply iterate through the pruning and completing process until no more cards can be added to any configurations without forcing a SET to appear:

```
S = [space[1], space[2], space[3]] # an arbitrary non-set of size 3

while True:
    T = prune(S)
    if len(T) == 0:
        break
    S = T

print(complete(S)) # all configurations of the maximum size in dimension d
```

Further areas of research

When you play the game of Set, there is a nagging sense that there is a deeper mathematical system at play beneath the surface than it seems at first. The intrinsic purpose of this project was to try to unpick the veiled mechanics and elucidate this structure.

One of the most striking initial findings was how quickly systems of Set cards became increasingly complicated, with many cases and intricate relationships between cards to consider. Finding meaningful results in configurations of only five cards presented a substantial challenge. This surprising impenetrability prompted us to try to connect the game with more familiar areas of maths with workable characteristics. Geometrical representations created a setting that re-framed the game so that we could convert our experience into more rigorous formulations. This abstraction to geometry and linear transformations was the key insight that helped us bypass many of the delicate intricacies.

The natural progression of this would would be to link our earlier combinatorial approach with our geometric construction of Set. Furthermore, since the architecture of Set lends itself well to the tools of linear algebra, it seems like a good idea to explore whether our conclusions generalise beyond Set. Once you start looking, Set has strong links to other areas of mathematics, which indicates that there might be other analogous systems or isomorphisms that we could explore. Perhaps, pursuing these similar systems could in turn give us new ways to interpret and understand the inner workings of Set.

Acknowledgements

We would like to sincerely thank the PROMYS Europe Programme, the Clay Mathematics Institute, and the Mathematical Institute of Oxford for the opportunity to work on this project, which would not have been possible without their support. In addition, we would like to express our gratitude to Alex Dam for her guidance in putting this paper together.