SETs and Anti-SETs: The Math Behind the Game of SET

Charlotte Chan

19 July 2010

In this paper, we will use the above change in font to distinguish between the English word "set" and a SET in the sense of the game. This way, we will be able to easily distinguish between, say SET theory and set theory, where the former means the math behind the game SET and the latter means something I will not discuss here. We will the game, ask some questions, and ultimately try to reformulate one main question into something that we can work with using mathematical tools. Here, no prerequisites are assumed. In fact, knowledge of how to play the game SET is not even assumed. To prove to you that I will stay true to my word, we begin with an introduction of this (awesomely fun) game.

1 The Game of SET

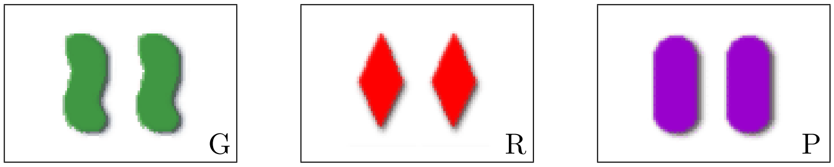
In 1974, a population geneticist by the name of Marsha Jean Falco was studying epilepsy in German shepherds. I know almost nothing about the details of the story, but her work in looking for patterns (somehow) inspired her to invent the game SET. We now transition abruptly into the rules of the game itself.

In SET, we use a special deck of cards. Each card is determined (uniquely!) by four characteristics: number, color, shape, and shading. There are three possibilities in each characteristic: 1, 2, or 3 for number; red, green, or purple for color; ovals, squiggles, or diamonds for shape; and solid, stripe, or open for shading. Because there is a card for each and every combination of possibilities of characteristics, we have a total of $3^4 = 81$ cards. To play, we start with a shuffled deck and the dealer deals 12 cards face up on the table. Your goal, as a player, is to find a SET. Once you find one, you call it out, collect the SET (assuming that it is a legitimate one), and the dealer deals 3 more cards. The object of the game is to collect the most SETs.

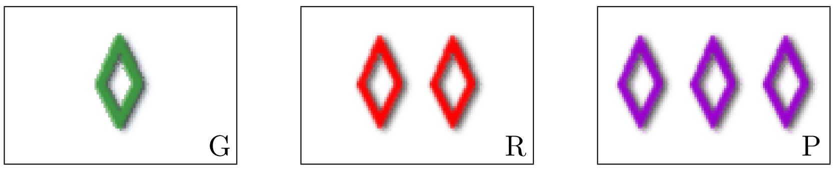
This raises several questions. Firstly, what is a SET, anyway? We answer this with the following definition:

Definition 1. We call a collection of three cards a *SET* if, within each characteristic, either all three possibilities are exhibited or exactly one of the possibilities is exhibited.

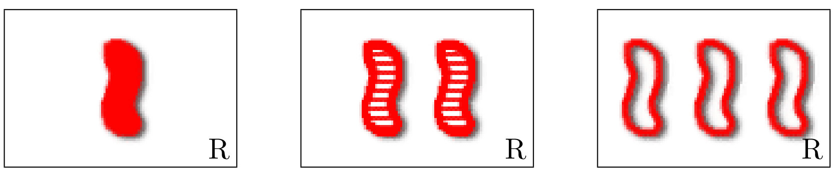
For example, the following three-card collections are SETs, where, because of the limitations in color printing, the letters denote the colors. That is, G denotes a green card, R denotes a red card, and P denotes a purple card.[1]



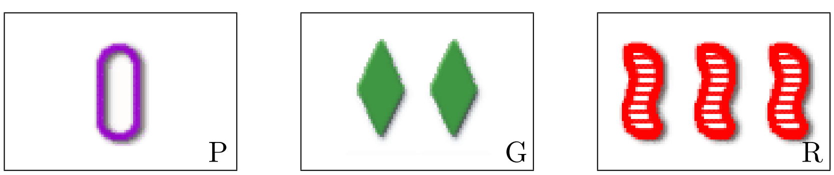
The above cards form a SET because they are all 2's, all different colors, all different shapes, and all solids.



These three are all the different numbers, all different colors, all diamonds, and all empty. Hence they form a SET.



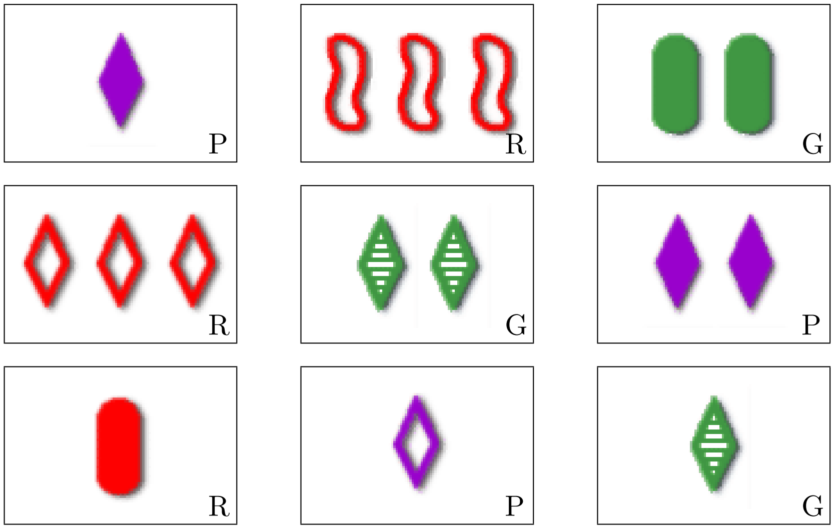
These three are all different numbers, all red, all squiggles, and all different shadings. They hence form a SET. In these first three SETs, there were exactly two characteristics whose variations were all different, and exactly two whose variations were all the same. Please note that this is *not* a necessary condition. Consider the following:



Here we have a collection of three cards between which the number is all different, the color is all different, the shape is all different, and the shading is also all different.

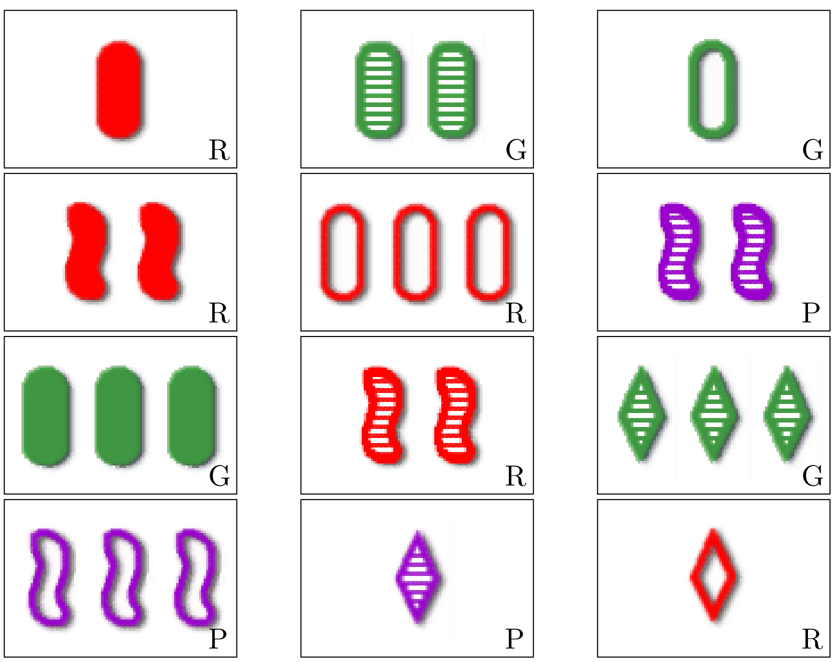
Brief Exercise 1. What about the following three examples? Are they SETs? Why or why not?

[1] SET graphics taken (without permission) from https://www.setgame.com/images/setcards/small/

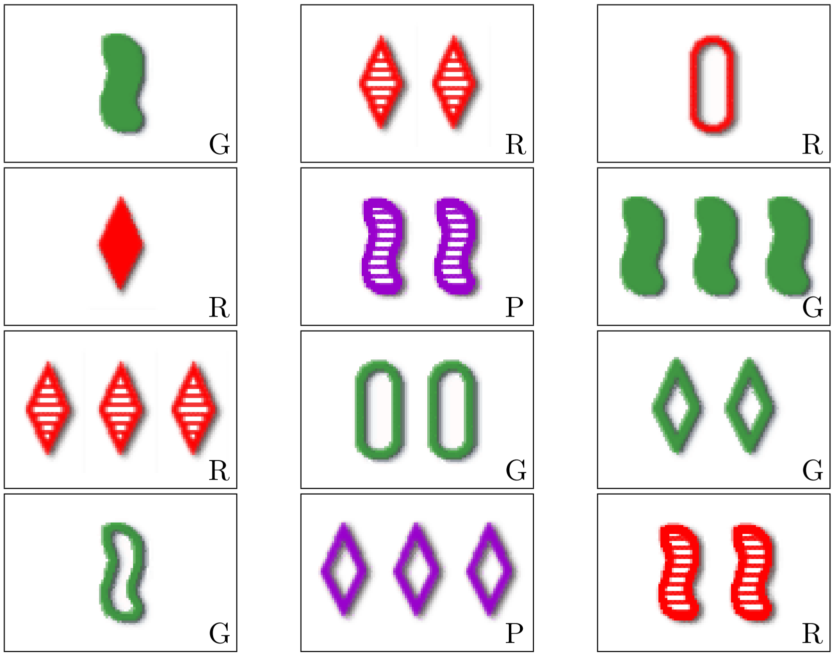


Solution. We can check each category to see if the above three cards satisfy the requirements for a `SET`. We have: all different numbers (!), all different colors (!), all different shapes (!), and... two solids and an empty. Hence the shading is where the above fails to be a `SET`. By the exact same check, we get that the second example is not a `SET` because of a fault in the numbering and the third is not a `SET` because of a fault in the shape.

Brief Exercise 2. For a harder exercise (the solution will not be provided in here), let's work with a simulation of the game `SET`. Try finding all the sets in the following display of cards:



Now that we've gotten our hands dirty and built up some experience with `SET`s, we can start asking questions beyond "`SET` or not?". We saw that we could find `SET`s in the previous display of 12 cards. A natural question, then, is, given any 12 cards, can we always find a `SET`? The answer is no, and here's an example:



We wouldn't want `SET` to come to a standstill if such a situation arises, so in the rulebook, we agree to deal 3 additional cards in the case that there are no sets present. We iterate this process until a `SET` is found. If a `SET` is collected, then the dealer will deal 3 cards if and only if the number of cards displayed drops below 12. This concludes the explanation of the rules of `SET`.

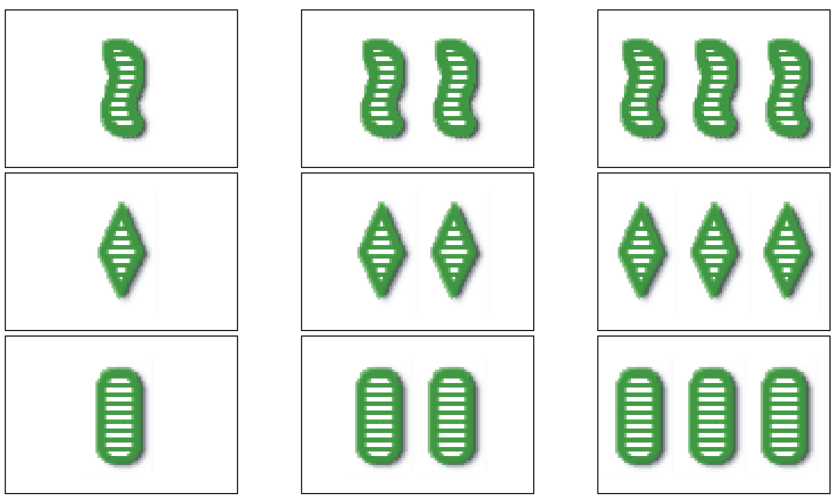
Notice that we have observed that we can find a collection of 12 cards that contains no `SET`s. Clearly, if all 81 cards were dealt, you would be able to find at least one `SET` (you would be able to find quite a few, in fact). This brings us to the first motivating question of today's talk: What is the minimum number of cards we need to guarantee the presence of a `SET`? We will play with this question, rephrase it, reword it, and ultimately try to make it something that we can formalize in a mathematically precise way.

2 Pessimism and the Anti-SET

In some sense, SET is an optimistic game: our goal is to *find* and *collect* SETs. But what if we reversed the goal? Let's be a little pessimistic and say, instead, that we want to avoid SETs like the plague. So instead of the question "What is the minimum number of cards we need to guarantee a SET?", we can coin the term "anti-SET" to mean a collection of cards without a SET, and ask: What is the largest possible anti-SET?

We could try to brute-force this by picking cards out of the deck, but this gets hard quite quickly. Of course, for the first few, it's easy. For instance, any two cards form an anti-SET of size 2. But what if we wanted to find, say, an anti-SET of size 18? Is it even possible? Instead of trying to tackle this question head-on, let's take a look at some simpler cases.

Let's first restrict our search for anti-SETs to only the striped green cards. So instead of looking at all 81 cards, we'll start by only looking at 9 of them. Essentially what we are doing is considering 2 characteristics instead of all 4. (Notationally, we drop the letter subscript as it is not necessary to distinguish between colors. This may not be worth noting, but if this document were to be printed in color, I assure you that the cards below would be green.)



We observe that we can pick at least of 4 cards without collecting a set. In fact, it isn't much harder to see that there are several ways of picking an anti-SET of size 4. The tic-tac-toe boards below demonstrate a few ways of choosing 4 cards from the above array of set cards. Here the entries in the tic-tac-toe board correspond directly to the above ordering.

X		X
X		X

,

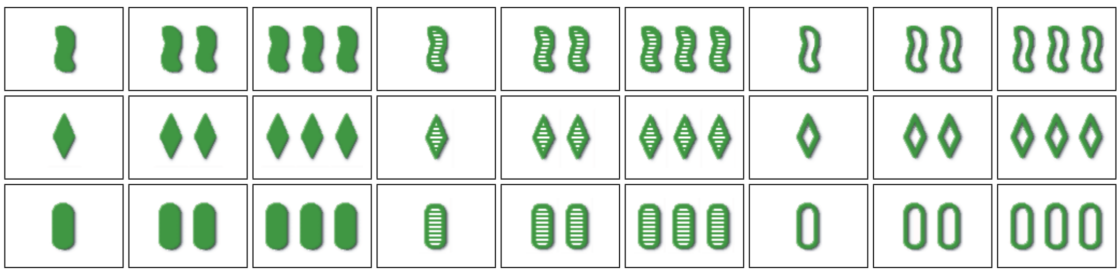
	X	
X		X
	X	

,

X	X	
X	X	

, etc.

What if we considered one more characteristic? Instead of fixing both the color and the shading, let's just fix the color. So we'll look at all the green cards. (As in the previous case, I assure you that if this document were printed in color, the cards below would most definitely be green. The letters to denote color are again omitted for the reason cited previously.)



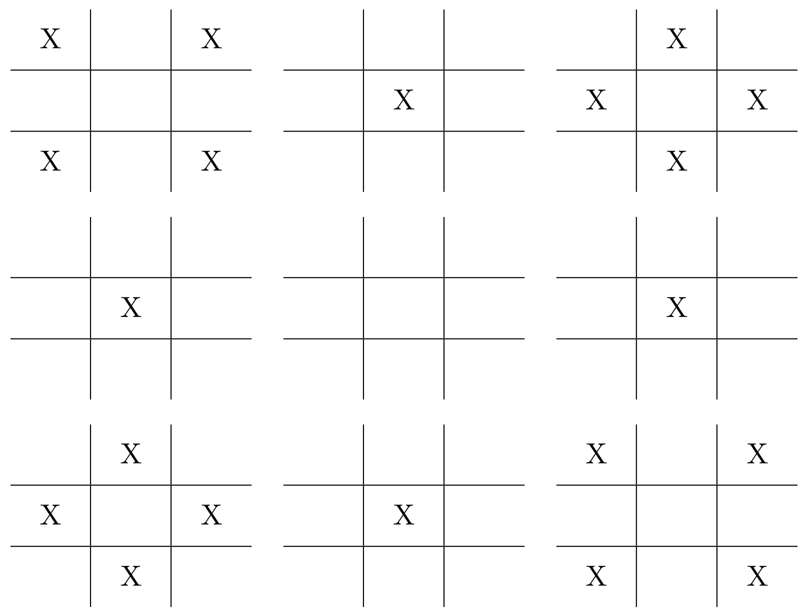
We can pick the cards as follows and get an anti-SET of size 9. (Note that this is not the only way to pick an anti-SET of size 9.) We again use tic-tac-toe boards as a convenient and simple illustration of our choice of cards.

X		X
X		X

	X	

	X	
X		X
	X	

We observe that there is some sort of correlation between the case of fixing 2 characteristics (i.e. 2 free characteristics) and the case of fixing 1 characteristic (i.e. 3 free characteristics). We can try to extend this pattern to work for the case when we let all four characteristics run free, that is, fix none. This would give us the entire SET deck. We structure the array of SET cards so that we have three clumped rows of grouped colors. That is, we write the previous formation of 27 cards down three times, with a different color each time. (Notice that this is analogous to the way we transitioned from writing down all the striped green cards to writing down all the green cards.) We then match the following array of tic-tac-toe boards to the array of SET cards appropriately, and observe that we can choose cards in the following way to get an anti-SET of size 20.



It is crucial to keep in mind that we haven't said anything about what the *maximum* number of cards we can collect without taking a set—all we've shown is that we can pick *at least* 4 cards if we only consider two characteristics, at least 9 cards if we only consider three characteristics, and at least 20 cards if we consider all four characteristics. In the language of anti-`SET`s, we know that we can certainly find anti-`SET`s of the above sizes, but we don't know whether it'd be possible to find a larger one in any of these situations. So how do we bridge this gap and show that the anti-`SET`s we found above are actually the largest possible? I told you earlier that I was going to talk about the math behind SET, so let's do exactly that.

3 A Mathematical Perspective

We can think of a set card as a point in the 4-dimensional vector space over a field of size 3, that is, a point in $\mathbb{F}_3^4$. (When you see $\mathbb{F}_3$, you should think of $\mathbb{Z}/3\mathbb{Z}$, or, in the standard PROMYS notation, $\mathbb{Z}_3$.) So we assign every card an ordered 4-tuple (x, y, z, w) with each coordinate entry x, y, z, w being 0, 1, or 2 (i.e. any element of $\mathbb{F}_3$ (think $\mathbb{Z}_3$)). We do this in the following way. Recall that every card is uniquely determined by four characteristics. We assign each characteristic a representative coordinate (which makes sense since we have 4 characteristics) and each variation within each characteristic an element of $\mathbb{F}_3$ (which makes sense since we have 3 variations within each characteristic). The natural question to ask, then, is in this language, what is a `SET`?

Recall that a `SET` is a collection of three cards with the property that, within each category, either all characteristics are exhibited or exactly one characteristic is exhibited.

Now assume that you have 3 points in $\mathbb{F}_3^4$ that correspond to three cards in a `SET`. Call these points A, B, and C. Let's look at an arbitrary coordinate of A. Call it x. For convenience of language, we'll say that x is the first coordinate (we can apply the same logic to the remaining coordinates). Since A, B, and C form a set, then we know that either both B and C also have x as their first coordinate, or that one has entry $x+1$ and the other $x+2$. This means that the the sum of the first-coordinate entries in this set is either $3x$ or $x+(x+1)+(x+2) = 3x+3$, which are both 0 in $\mathbb{F}_3$. What we have, then, is that $A+B+C = 0$ in $\mathbb{F}_3^4$.

We make another observation. Since we are working in $\mathbb{F}_3^4$, we have that $B = -2B$, so by substitution, our equation becomes $A - 2B + C = 0$. From this, we get $A - B = B - C$, which gives us directly that A, B, and C are collinear. It is also easy to check that if A, B, and C are collinear, then the cards corresponding to A, B, and C form a `SET`. Hence we have shown that A, B, and C form a line[2] in $\mathbb{F}_3^4$ if and only if the cards corresponding to A, B, and C form a `SET`. In fact, if we had a game of `SET` with 5, 6, or any number of categories, we could make the same claim, replacing $\mathbb{F}_3^4$ with, respectively, $\mathbb{F}_3^5$, $\mathbb{F}_3^6$, or $\mathbb{F}_3^d$ where d is any number of categories.

Now let's introduce a definition.

Definition 2. A d-cap is a subset of $\mathbb{F}_3^d$ that does not contain any lines.

So remember our main goal? We wanted to know what the maximum number of cards we could have without including a `SET`, or in a different phrasing, we wanted to know the maximum size of an anti-`SET`. Recall that the first thing we did was restrict our observations to only 2 categories. Then we broadened our search to 3 categories, and then finally considered 4 categories. In the "mathier" language that we have just introduced, our examination of the cards in these three cases translates to an examination of a 2-cap, a 3-cap, and a 4-cap, respectively. So instead of asking about the maximum size of an anti-`SET` considering d categories, using our new vocabulary, we can ask the following: What is the maximum size of a d-cap in $\mathbb{F}_3^d$? Notice that with the introduction of some new words with precise definitions, we've transformed a sort of vague, everyday question—What is the minimum number of cards we need to guaranteed the presence of a `SET`?—into one that we can really feel and taste and work with in a mathematically precise way.

Proposition 1. *The maximum size of a* 2*-cap is* 4*.*

Proof. We showed by example that it is possible to find a 2-cap of size 4. So what remains to be shown is that it is, in fact, the maximum size of a 2-cap. Suppose that we can find a 2-cap of size 5. Let's call these points x_1, x_2, x_3, x_4, x_5. Now, $\mathbb{F}_3^2$ can be decomposed into the union of 3 parallel lines, each containing at most 2 points in the 2-cap (since three collinear points in $\mathbb{F}_3^2$ comprise a line and a 2-cap contains no lines). So one line

[2] In this paper, I will use the word "line," even though what I actually mean is a one-dimensional affine subspace of $\mathbb{F}_3^4$.

contains exactly one point. Call this line L^*. Without loss of generality, let us assume that L^* contains x_5 (we can always rename things). In $\mathbb{F}_3^2$ there are exactly four lines that contain x_5. We know one of them—L^*—and let's call the other three L_1, L_2, L_3. Note that these four lines cover all the elements in $\mathbb{F}_3^2$, i.e. $\mathbb{F}_3^2 = L^* \cup (\cup_{i=1}^3 L_i)$. In particular, this means that $x_1, \ldots, x_4$ must each be in at least one of these lines. Since L^* contains none of these points, then they must each be in at least one of L_1, L_2, L_3. By the pigeonhole principle, one of these lines must contain two of $x_1, \ldots, x_4$, and since this line also contains x_3, then this means that we have found a line in our supposed 2-cap of size 5. This is a contradiction, and hence we conclude that the maximum size of a 2-cap is 4. □

For the case $d = 3$, we make a similar argument. Recall that we showed (by example) that it is possible to find a 3-cap of size 9. Now we proceed to argue the following proposition by contradiction.

Proposition 2. *The maximum size of a* 3*-cap is* 9*.*

Proof. Suppose that there exists a 3-cap of size 10. Note that $\mathbb{F}_3^3$ is the union of three parallel planes, just as $\mathbb{F}_3^2$ is the union of three parallel lines. Note also that the intersection of a 3-cap with one of these planes is just a 2-cap, and by Proposition 1, we have that a 2-cap can have at most 4 points. Let P^* be the plane with the fewest points of the 3-cap (keep in mind that P^* is a 2-cap). Then P^* must have 2 or 3 points. This is easy to check by seeing that it can't be otherwise. If P^* had 0 or 1 point, then our 3-cap would only contain at most 9 points (since no plane can contain more than 4 points). If P^* has 4 points, then our 3-cap would contain 12 points, but we want 10 points, so that doesn't work either.

Now let $x_1, \ldots, x_{10}$ be the ten points of our 3-cap and without loss of generality let x_9, x_{10} be the points contained in P. There are three other planes that go through x_9 and x_{10}. Call them P_1, P_2, P_3. Notice that these four planes cover $\mathbb{F}^3$, i.e. $\mathbb{F}_3^3 = P^* \cup (\cup_{i=1}^3 P_i)$, which means that $x_1, \ldots, x_7$ must be contained in $\cup_{i=1}^3 P_i$. (Note: We exclude x_8 since we only know that P^* contains 2 or 3 points.) By the pigeonhole principle, we get that one of these planes (say it's P_1) must contain 3 points of our 3-cap. But P_1 also contains x_9, x_{10}, which means it contains 5 points of $\mathbb{F}_3^3$. But this means that P_1 is a 2-cap of size 5, which contradicts Proposition 1. Therefore we have that the maximum size of a 3-cap is 9. □

Because of the similarities between the proofs of Proposition 1 and Proposition 2, it would seem reasonable to expect that the proof to verify that the maximum size of a 4-cap is 20 would follow in a similar vein of thought. But in fact it doesn't. The proof is significantly more complicated and I will not present it here. If you are interested, please see the reference at the end of these notes.

It turns out that this problem of finding the maximum sizes of d-caps is actually a very hard problem that gets complicated very quickly. For the case $d = 1$, it is trivial, for the case $d = 2$ and $d = 3$, we saw that it was not so difficult, but for $d = 4$, things get harder, and for $d = 5$, progress has only been made in recent years. This problem is so hard, in fact, that for $d = 6$ and larger values of d, very little, if anything, is known about the size of the maximal d-cap in $\mathbb{F}_3^d$.

If we think back to where we began this discussion, it's quite miraculous how far we have come. We took the game `SET`—something that seemed concrete, tangible, and seemingly innocent—and kept asking questions and rephrasing our questions until we came out with something that we could generalize easily to more characteristics, more dimensions. And then suddenly we were working with something significantly harder than what we began with. It just goes to show the importance of asking the right questions, and if we do fall upon the right question, we may end up sitting on top of some very hard, unsolved problems.

References

[1] Davis, Benjamin Lent and Maclagan, Diane. "The Card Game Set." The Mathematical Intelligencer, 25, (3):33–40, 2003.