

CHAPTER SIX

Interview Results and Analysis

Introduction

In this chapter, attention shifts away from the class as a whole to examine in depth the thinking of six students who each participated in two interviews. The point of the examination is to show how their collective responses inform the aspects of variation in the revised framework outlined in the previous chapter.

There are three females (SP, EM, and DS) as well as three males (GP, JM, and RL) who were selected to be interviewed. They were selected mainly because on the PreSurvey, their collective responses seemed to span all the coding levels. In that sense, they seemed somewhat representative of the class. Moreover, they provided enough written detail to convince me that they would have no problem sharing their thoughts in an interview situation.

Each of these six students, who I will also refer to as my six cases, participated in two videotaped interviews, with each interview lasting about 45 minutes. The PreInterview was given within two weeks of administering the PreSurvey, and was conducted before formal instruction on probability and statistics in Math 212 had begun (recall that the first four weeks of the quarter in Math 212 was spent on geometry). After four weeks of doing lessons and activities on probability and statistics (this time frame corresponds with weeks 5 – 8 in the ten week quarter), the PostInterviews were conducted. The PreInterview and PostInterview instruments are found in Appendix B.

As mentioned previously, the framework that began to emerge at the end of the last chapter provides a way of organizing the discussion in this chapter. The summary of the student thinking unfolds first through both the PreInterview and then the PostInterview, with primary attention being given to how the subjects' responses fit into and are captured by the framework for analysis. Any noteworthy comparisons (from within the PreInterview or PostInterview, or across from the PreSurvey to the PreInterview or PostInterview) will be mentioned, and then in the next chapter the cases will be compared to each other as the elements of the framework are pulled together in summary.

PreInterview

Although the entire PreInterview script can be seen in Appendix B, for ease of reference the questions are annotated in tables within this chapter as needed. Along with the question number, a brief name and description of the question is appended, and the contexts of the questions are also specified. The structure of the PreInterview was similar contextually to the PreSurvey in the sense that the sampling context was placed first, followed by data and graphs, and finally by probability. Table 58 shows the first two questions.

Question 1

The first question on the PreInterview was identical to the Q1 on the PreSurvey. This was asked as a way of easing the subjects back into the context of pulling handfuls of ten from the jar with 60 red and 40 yellow candies.

PreInterview Questions #1&2		
Question Number & Nickname	Brief Description	Context
Q1a "One Trial"	A jar has 60 Red, 40 Yellow. You pull a handful of size 10 (this is one trial). How many reds in your handful? Why?	Sampling
Q1b "Repeated Trials"	You do several trials (as in Q1a). Will you get this many reds every time? Why?	Sampling
Q1c "Six Trials"	Suppose 6 people did trials (as in Q1a). How many reds might they each get? Why?	Sampling
Q2 "Lists"	Five lists are shown, each showing a different "Six Trials" set of choices. Which list best describes what might happen?	Sampling

Table 58

The two main aspects of the framework reflected in the responses on this question were expecting and interpreting variation.

Expecting Variation: In this aspect, both *what* and *why* were informed by students' responses. First of all, concerning *what* was expected, one area of focus was on the expected value (6 red). EM, DS, and JM agreed in the "Six Trials" part of Q1c that results should be close to the expected value:

- EM: Well, my thinking is that you're going to get close to 6
 JM: You're gonna pick out close to that 60% each time
 DS: It's more common to get closer to your percentage

Also, even though five of the six cases did give 6 reds as the expectation for "One Trial", JM was emphatic about results not being the expected value each time. When I asked why not, he said: "Well, because it's... it's impossible." This concern about repeating values goes beyond just the repetition of the expected value, as SP's comment shows:

- SP: I'd be more surprised if the same number kept showing up, as opposed to if it was just completely random

Notice the implications about randomness in SP's response, whereby random behavior seems to preclude having the same number keep appearing.

Another facet of *what* is expected is shown by SP in that she focused on range when thinking about "Repeated Trials," saying "I feel like it'd be somewhere in the same range." She also expected a result to be somewhere in the midrange on "One Trial", and this idea of a middle value is shared by others in later questions. Here is how SP expresses this in response to Q1a:

SP: Well, I guess instinctually I would say that it'd be somewhere like in a median, like uh... 4 , 5 ... just instinctually.

What SP is calling the median, I am classifying as the midrange because it seemed that how she obtained a "median" was to look at the options of 0 red and 10 red and went for the middle.

Second, concerning the *why* behind their expectations, one dimension of reasoning had to do with possibilities and likelihoods, such as the notion that the expected value was the most likely outcome:

DS: Six is the best shot
 RL: Six is the most likely
 EM: It's just I think you're more LIKELY to get six

Also, a common theme echoed throughout the interviews was that extreme values are possible:

JM: You know, of course you CAN pick out ten red, or you can pick out zero red
 SP: And so there is a chance that you could grab all yellow, or all red.
 RL: It is entirely possible that if there were 99 candies and 1 yellow, you could pick that one yellow every single time.

As RL's response shows, for him not only are extremely unlikely outcomes

possible, but those outcomes are “entirely possible.” At the same time, however, RL and others agree that extremes are also unlikely:

- RL: It’s on the far end of a bell curve, it’s extremely unlikely, but it COULD happen.
- DS: And because, it’s going to be rare to get fewer reds, or a lot more reds
- SP: Well, I guess I just never want to say 0, and never want to say 10.

It is worth noting that RL’s response showed some distributional reasoning with his use of the bell curve, which he employed when thinking about the “Six Trials” part of the question. The most prevalent form of explaining *why* was to use proportional reasoning, exhibited by all except SP on this question.

Interpreting Variation: The three categories of *cause*, *effect*, and *influencing expectation and variation* were also illustrated by responses to this question.

Regarding *cause*, for three subjects the nature of the actual mixing was seen as a source of variation, with GP being particularly adamant about how the candies were selected:

- GP: Uhh, random being ... You can grab from so many different places in the jar. So your hand can go this way, <He demonstrates with his hands>, this way, this way, this way, this way... And then, while you reach in, the candies move, all over the place, and so... You just... Your hand creates randomness. <Laughs>

As for an *effect* of variation, I have included the notion of SP (and others) that “It can be anything” in this category because it seems that the

students point-of-view is an indirect result of variation. At three different times in discussing Q1, SP expresses this notion:

- SP It could be ANY amount, and I can't guess, you know.
- SP Logically that's what my brain is telling me, is it can be absolutely anything.
- SP Yeah, because I feel like it really can be anything. And so making a guess is just like.... Just saying anything.

SP is not by any means alone in this view, as will be seen in future responses to other questions. Her responses above are reflective of the Outcome Approach described in Chapter Two, whereby students focus on correctly predicting the next outcome.

Another effect of variation, which seems to be related to the notion that "It can be anything", is a difficulty in deciding or predicting results. SP notes the difficulty above by saying "I can't guess," and EM plainly shows her dilemma as she wrestles with "Six Trials" :

- EM Ohh. I don't...I don't know... <Sighs>. Well, I think, like, I don't know...I don't have any set computation, but I think it's somewhere around 6..."

My thinking is not that students necessarily perceive their own views of "It can be anything" or "I don't know" as an effect of variation, but it is because of the variation inherent in the situation that results in uncertainty, leading in turn to the difficulty students have with making a choice. Variation (and the resultant uncertainty) means you don't know for sure what will results, and if you don't know what it'll be, then it could be anything, and so it's hard to decide. Note too how EM's reference above includes the possibility that a mathematical computation might somehow "answer" the question of "Six Trials" in a

definitive way.

A last effect, mentioned only by RL here in this question but also by two others in future questions, is how reality is somehow different from probability. At first this idea is somewhat vaguely expressed by RL, as he explains why he put all sixes at first for his “Six Trials”, saying that he was “being very strict as in probability-dictated reality, as distinct from described likelihoods.” Further probing showed that he meant he first put all sixes because that is supposedly what probability says to him *should* happen, but then he changed his mind because, he says, “I thought, okay, reality is going to impinge on the strict likelihood by a given thing.” So, an effect of variation is that reality does not always match with what probability says should happen, at least for RL in Q1.

Next, I’ll share some responses to Q1 that I have included as ways students think of *influencing expectation or variation*. On a more basic level, several of the cases spoke of the actual numbers of candies as having an influence on results, and this contrasts with proportional reasoning because the attention is not on the ratio, but the sheer quantity. On a more sophisticated level of reasoning, DS links the number of trials with the resultant variation, as illustrated in this brief exchange in Q1b:

- DC Let’s suppose you do this several times. Are you going to get that many again?
- DS The more I choose candies, the more chance there will be that I’ll get different than six reds. Either fewer or more.

JM is more specific about what he thinks happens with more trials, and he

says that “I think on average, if you did it enough times, you probably average uh... 6 reds.”

Question 2

After the “Six Trials” task in Q1c, this question provided examples of lists of six choices that other people are said to have chosen. The list that four out of six cases liked the most was List (ii) - “6, 7, 5, 8, 5, 4” - which of all the other lists is in fact the most likely to occur. GP preferred not to pick a list as most likely, and SP chose List (v) – “3, 10, 9, 2, 1, 5” – which although wide is at least consistent with SP’s notion of “Anything can happen.” The same aspects of expecting and interpreting variation, as drawn out in Q1, were also illustrated by responses in Q2.

Expecting Variation: Looking first at *what* was expected, all six cases expressed thoughts about how the results should look with respect to the expected value of 6 reds. Five cases mentioned was that results should not be 6 reds every time, as DS make clear in reference to List (iii), “6, 6, 6, 6, 6, 6” :

DC: And what do you think about choice number 3?

DS: Well, choice number 3 , I mean, that’s sticking with your, you know, your 60%,but it’s very rare that it’s going to be exactly the same each time you pick it. So, I don’t think that’s going to happen.

Instead, as suggested by all except GP, results should be on both sides of 6, and three of the six cases even mentioned wanting a range around 6, as exemplified by JM’s reaction to List (i) - “7, 9, 7, 6, 8, 7” :

JM So, I’d say, I don’t like it as much as, something that has a greater range from below and above 6.

In addition to the expected value of 6 reds, SP is unwary about having other values repeat also, and she says about List (i) “If there’s three 7s... I just, I just don’t feel like that would be as likely.” Focusing more attention on range or extremes, EM, SP, and GP talked about not wanting too many high or too many low numbers in their lists of results, but instead wanted to see most numbers in the middle, or midrange area:

- EM You pull like 4...4 reds and maybe 6 yellows instead. So I kind of like that this is middle-of-the-road.
- SP Yeah, like 4s , 5s, and 6, to me, is somewhere like in the middle

In SP’s case, the numbers she gave in the example above were referred to by her as “median numbers.”

Two of the same dimensions of explaining *why* showed up in the responses for Q2 as also did for Q1, and the first of these dimensions concerns possibilities or likelihoods. Specifically, the expected value is again seen as the most likely result, as RL states succinctly that “on any given pull, you’re more likely to pull six reds than any other number.” Also, in claiming that the extremes are unlikely, GP uses some unique language in talking about why he likes List (ii) – “6, 7, 5, 8, 5, 4” - the best:

- DC Any reason why?
- GP Probably less radical numbers <Laughs>
- DC Like, what would be a radical number?
- GP Well, the 1 and the 10... you know, 9 is kind of radical
- DC Radical meaning?
- GP Extreme, less likely, you know

But extremes, though unlikely, are still possible, as RL discusses when

considering List (v) – “3, 10, 9, 2, 1, 5” :

RL Choice five strikes me as the worst of both worlds, in terms of outliers. 10, getting ten reds, seems to me least likely of all the numbers, for ten or zero.... But, well, yeah, I would say, you know what? It's possible. It could happen.

The second dimension of reasoning *why* again has to do with proportional reasoning, but this was only reflected in the comments of RL and JM for this question.

Interpreting Variation: Again, case responses fell into the categories of causes, effects, and influences on expectation or variation, and EM's thinking about *causes* of variation sounds very similar to the earlier reasoning shown by GP on Q1, as this exchange shows:

EM <Laughs> Yeah, Ok. Um, well, just, you know, they're random in the jar, and...
 DC What does that mean, by the way?
 EM Well that means you can't tell where they're going to fall in the jar. And so, I think that in some places, you're not gonna always have a red/yellow red/yellow, or, you know. So, in some places, they'll be yellows that have collected together.

The way I began to interpret the theme I picked up on from GP and EM was that specific attention was given to the physical nature of the candy mixing.

The hand doing the grabbing might go to an area where there are more yellow, or maybe the reds have somehow gotten together in a little cluster – It seems that in some students' minds, these physical characteristics are seen as the reason results are different, and thus are a cause of variation.

Effects again include the “It Can be Anything” phenomenon, which on the previous question was expressed by SP. For this question, the idea is

captured by JM's succinct opinion on different lists as he says, "Well, they're all likely." Not only can anything happen, apparently, but all things are likely for JM at this point in the course. GP had difficulty in choosing the list he liked best, and in fact he declined to select a list as the one he thought was most likely:

- GP Mmm-hmm. Uh, kinda hard to do, I mean, they're kinda similar...
I guess, I dunno...
- DC Because...
- GP Because there are just so many different choices, and so many different choices in the other one, that it's kinda hard to decipher for me...

Further probing confirmed that GP found it hard to choose one list over another, and this seemed to be a result of the variation in the choices he was faced with.

Finally, in terms of *influencing expectations and variation*, DS again draws a connection between variation and the number of trials. Also, EM supports thinking of the expected value as more of an average, saying "that would be my average number." I believe that students who talk about getting an "average" equal to expected value see the number of trials as influencing the cumulative average. That is, whatever variation they may see in one, six, or many trials, if more trials are done then the average of those results should be closer and closer to the expected value. The key idea is the link between variation, expectation, and the number of trials, and this idea is expressed in several forms throughout both interviews.

Question 3

I call this question “Fake: 30” because it shows the purported results of 30 trials from the same sampling context as Q1 and Q2, in both tabular and graphical form. Because of the table and graph, this question spans both the sampling as well as the data and graphs context. Question 3 is described below, and since both of Questions 3 and 4 have similar formats, they are paired together in Table 59 below:

PreInterview Questions #3&4		
Question Number & Nickname	Brief Description	Context
Q3 “Fake: 30”	The supposed results of 30 trials are shown. What is most likely – The results are real, made-up, or no one can tell?	Sampling & Graphs
Q4a “Real: 300”	The supposed results of 300 trials are shown. What is most likely – The results are real, made-up, or no one can tell?	Sampling & Graphs
Q4b “Compare 30 & 300”	How does the shape of the graph in Q3 compare to the shape of the graph in Q4a?	Sampling & Graphs

Table 59

Although students do not compare multiple graphs in Q3, I was curious to see what, if any, features of the graphs they attended to in discussing the question: “Which do you think is *most* likely?” The question was followed by options suggesting that the results were actual, made up (fake), or that no one can have much confidence if the results are made up or not. As can be seen from the graph (see Figure 38), the “results” consist only of 5’s, 6’s and 7’s (with a mode at 6). These are in fact fake results, and highly unlikely to actually occur in thirty trials.

Expecting Variation: The only response that dealt directly with *what* was expected came from GP, whose comments suggested that results should be

on both sides of the expected value.

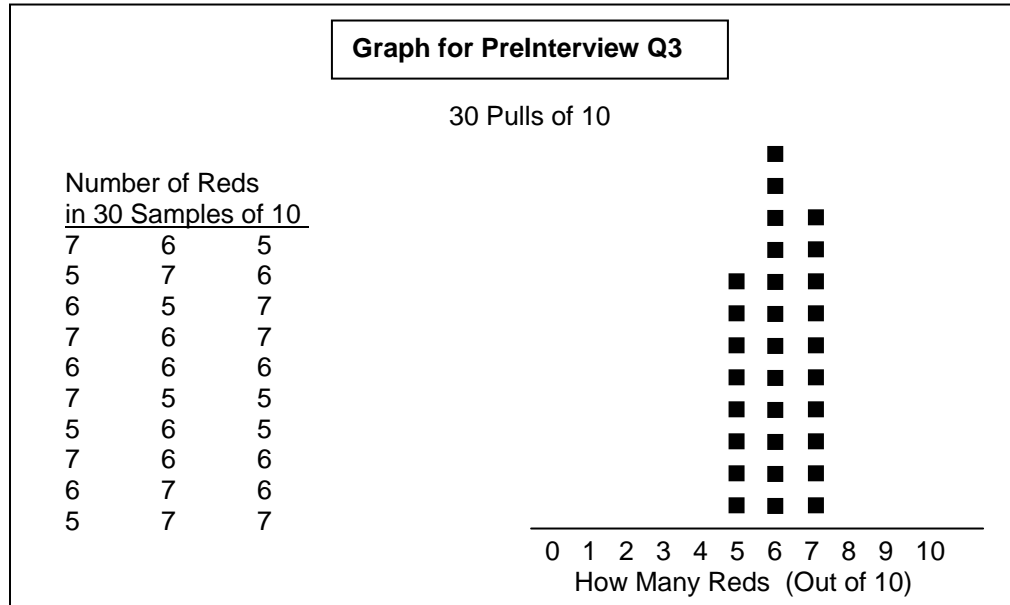


Figure 38

For the *why* component of this aspect, again DS and RL expressed how the expected value was the most likely to occur. A new dimension to *why* was added by both SP and DS, as they shared the opinion that sometimes you just don't see much variation when you do an experiment such as thirty trials of this nature:

- DS Because six is our odds-on favorite, and they just didn't have a lot of variation when they picked out their candies.
- SP Because we REALLY can never know, because it , it COULD happen, there's always the chance that it could happen.

Notice how SP turns the focus of the notion of what is possible away from the extremes and toward the limited range of the results shown in the graph. That is, she is essentially looking at the way that all results are at 5, 6, or 7 (which is reflective of minimal variation) and saying that it "could happen."

Displaying Variation: This is the first time in the PreInterview that this aspect of displaying variation got addressed, since it was the first question to contain a graph. Two dimensions of displaying variation are illustrated by responses to this question. The first dimension involves noticing or *evaluating* graphs, in terms of the features of the graph that students attended to. All six cases gave comments in some form showing their attention to the limited range shown in the graph:

- DS Because here's six, and they're only one away from six, on each side
- EM Well, I see that they're only using three numbers, 5, 6, and 7
- GP Because they only have it in the three groups here
- JM I can't believe that it was only three numbers. I mean, it's possible of course
- RL It's curious, that there were only three possibilities out of ten that were actually seen out of thirty.. That strikes me as odd
- SP Because they're all 5, 6, and 7s... And three numbers in a row and... All clustered

Considerable attention to range was given by four cases, who suggested the graph needed to span a broader range. EM's response was especially articulate:

- EM I think that you would have at least one or two 2s and you'd probably have some 3s... I think you'd have actually a range from at least 2 to 9, some more in there.

There was also elements of distributional reasoning evident in evaluating the graph, shown by four cases. GP, who so far in the interview has had more lower-level responses (attending to the way his hand might grab the candies, for instance), now shows that he has some definite ideas of what a reasonable distribution might look like for this problem:

GP And there should be... you know, more even, like, more of a pyramid or something. I think 6 should be probably the highest, usually the highest, and then kind of spread on down [Using both hands to show tapering from the peak]

It was in my interactions with GP that I first started noticing the reliance on gesture in describing variation, and although I found two articles in the *Journal of Educational Psychology* (Flevaras & Perry, 2001; Goldin-Meadow, Kim, & Singer, 1999) discussing gesture and mathematics, I know of no research that focuses on gesture in the context of learning or teaching probability and statistics. But gestures have shown in both interviews to constitute, in a sense, a language all its own. And in this question, GP visually shows via gesture the kind of distribution he pictures in his mind. JM actually draws on his paper, as this exchange illustrates:

JM Well, I'm thinking more of a, you know, more of like a classical, uh, what do you call it, not a bell curve, but a ...<He draws an inverted V over the graph>

DC That's more what it ought to look like to you?

JM Yeah, just a little bit more broad distribution <Shows with his hands a bigger range>

JM uses his distributional reasoning to go on and illustrate the second dimension of displaying variation, which is making conclusions about graphs.

JM Well, if I believe, as I've stated, that there should be a more broad distribution of the numbers, <Points to his inverted V> then they must be made up

However, later JM ends up not having confidence in this conclusion, whereas three other cases (EM, GP, and SP) were firm in their conviction that the graph showed made-up results. DS and RL felt the results were real.

Interpreting Variation: Causes and effects were both brought out in this question, and only by JM and SP , but they are worth mentioning because as themes they persist throughout both interviews, for JM and SP and other cases as well. The physical arrangement of the candies, which had been shared by EM and GP so far, now gets a nod from JM.:

JM Uhhh, <Long pause>. You know, I guess I'd have to see how they fit into your hand. <Chuckles> , Maybe that has a bearing on it possibly, right? And when you reach into a container, and pull them out, and if they're completely mixed, whereas one red is lying against, or there'd be, what is it, 60%? So you'd have almost ... 2 reds around 1 yellow. Maybe? Something like that?

The point is that the physical nature of the problem seems to be a real issue for some of these preservice teachers, and it certainly is for children. Also, whereas JM first had concluded that the graph in Q3 showed made up data, he had a marked change of opinion after he compared Q3's graph to the one in Q4. In fact, he also argues that Q3 showed actual data, and eventually found it safest to say that no one could have much confidence in saying one thing or the other. I have categorized this indecision, or rather a decision to not make a conclusion, as an effect of variation because it was very clear that it was the variation (or lack thereof) which accounted for his waffling. Too little variation allowed SP to insist, as shown earlier, that the results could have happened, and thus she stated her attraction to the option of having no confidence. I see a lack of confidence in making a conclusion on this question as an effect of variation, or at least an effect of not understanding variation.

Question 4

This question extends Q3 by showing (as a graph without a table) the purported results for 300 trials, and I call the first part of the question, Q4a, “Real: 300” because it shows the actual results as generated by the Fathom software (see Figure 39).

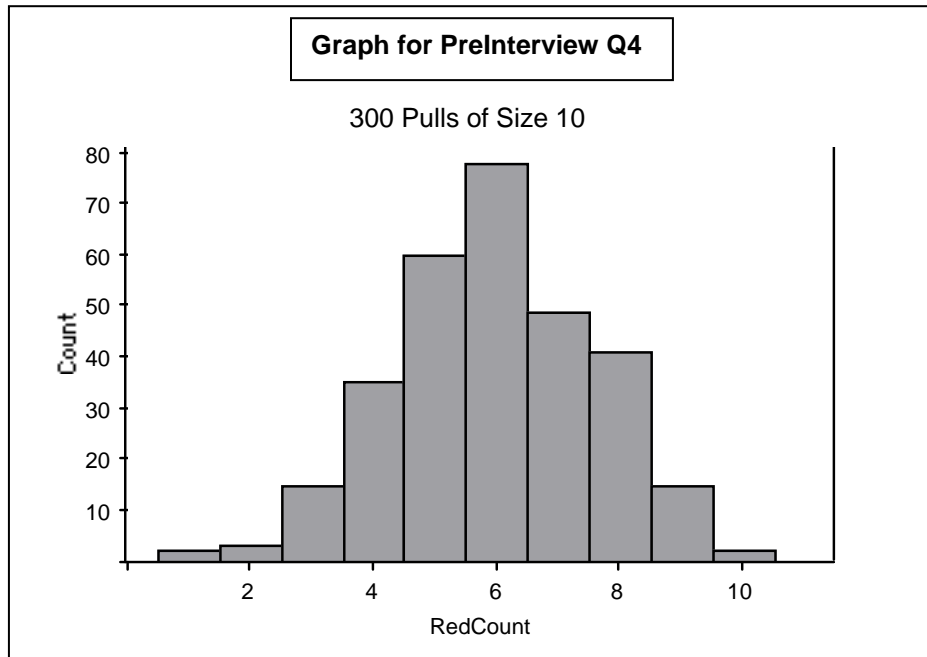


Figure 39

The second part of the question, Q4b, is called “Compare 30 & 300” because it asks how the shape of the graph in Q3 compares to the shape of the graph in Q4a. As in Q3, all three aspects were brought out by this question.

Expecting Variation: Again showing that *what* she expects is numbers in the midrange, SP liked the graph in Q4a because “it’s spread out, there’s a little bit in everything, and then, that most of them are somewhere in the middle.” The dimension of *why* was included by how the expected value is most likely to occur, and that extreme values are unlikely:

- RL Where, yeah, you see the most likely thing the most times.
 JM And this one I think, goes out to the extremes <Pointing to 1 and 10> where, I don't think it's possible to get a handful of either all red or one red.
 DS They're VERY unlikely, and so, there are hardly any, you only have like one person chose zero, and one person chose ten here, maybe two..."

So although for JM the extremes are not possible at this point in the interview, and for DS they are unlikely, RL said "I like the outliers. I guess they're not really outliers...The extremes." The main idea is that, in looking at conceptions of variation, there is this contratenion between the viewpoint that says extremes are unlikely and the viewpoint that says extremes are possible (Torok & Watson, 2000).

Displaying Variation: In comparing the two graphs, several students commented on how both modes were at 6 reds. Attention was also paid to the range, or spread of the graphs, as shown in this exchange where DS is comparing the shapes of the two graphs:

- DC Any other differences or similarities of shape?
 DS Well, this {Q4a} has a broader range of picks, of number picks. {Her finger traces the range along the horizontal axis}
 DC Why do you say that?
 DS Because it goes all the way from – I'm assuming this is zero – to ten, {She's referring to Q4a}, where here {Q3} they only chose between 5 and 7 reds

Moving beyond just attention to the range, some responses also showed what I considered to be more of a distributional form of reasoning. This reasoning could include mentioning how the data was distributed, or the shape of the distribution, or using multiple characteristics of the distribution. For instance,

GP attends to both center and spread in considering Q4a, and SP comments on the shapes of the graphs in both Q3 and Q4a:

- GP Well, it seems like it {Q4a} spreads out, and is, you know... Um, you have a few extremes out here {Pointing to them}. And then it kinda goes up, where it's more likely in the middle here {Points to the mode of 6}
- SP This {Q3} is just a more extreme curve {Points to the mode in Q3} and this is a more gentle, gradual curve {Traces out graph in Q4}

Regarding the more bell-like shape of the graph in Q4a, DS found this characteristic to be appropriate, while SP found it questionable:

- DS 'Cause I think when you're picking random things out of a given sample set, that you're going to end up with a bell curve <Laughs>
- SP Uhhh, the only... well. The only reservations is how perfect it {Q4a} is.
- DC And by perfect, you are meaning ?
- SP As in, it, you know, it's like this perfect curve. Whereas I don't know if something's randomly being chosen, that you can get this perfect curve. <Traces the curve with her finger>

I think the contrast between these two cases is quite stark: They both use the unsolicited language of random sampling, they both refer to the shape, but they draw different conclusions about the veracity of the shape given the nature of the situation. These responses fit into the aspect of evaluating and comparing graphs, with an emphasis on distributional reasoning, precisely because of the attention to the shape of the distribution.

Finally, in the aspect of displaying variation there is also the dimension of making conclusions about graphs, and all six cases at some point felt that Q4a displayed real results. JM did have some equivocation in Q4a, as he did

earlier with regard to Q3,. In the end he felt that the far extremes shown in Q4a were just not possible. Meanwhile, EM and GP show how they are comfortable with the distribution in deciding that the graph shows real results:

- EM Well, uh, I would say that these look more...actual, like they could be the actual results of the class. For the reason...that there is more variety in the numbers {She is tracing out the graph, on either side of the mode} . So there are some 3s reds... 3 reds were pulled sometimes and... So I like that
- GP I think this is more like the pyramid, what I would see. Um, this looks more legit to me {Shows an inverted V with hands}

Interpreting Variation: For causes of variation, the nature of the candy mixing gets a reference by RL, who now joins EM, GP, and JM as subscribers to this facet of the situation:

- RL You don't expect to see somebody to pull 10 candies, or zero candies of any particular color, because they have to move their hand around a lot of other candies.

RL's language helps show again why this is categorized as a cause of variation. It may be true that extremes are not expected, but why? Because the hand that does the grabbing is moving to "other candies" – This is seen as the reason for why a person would get different outcomes. For effects of variation, I have again listed JM vote of no confidence in this category, since it seemed that a result of variation was that he could not be certain what was real and what was made-up.

The dimension of influencing expectation and variation again is brought out in this question, by RL, who suggests that the number of trials has something to do both with the amount of variation and the shape of the graph:

- RL You can pull out the {one} yellow candy, it's possible. It's not going to happen that often, but if you do it enough times, sooner or later, it's bound to happen.
- RL If you did this ten thousand {times}, you'd probably have a really nice bell curve. So, I attribute the more-bell curve-looking design to the number of trials.

Question 5

Question 5 (“Car Brakes”) was the first of four questions set in the context of Data and Graphs. All four questions are listed below in Table 60:

PreInterview Questions #5,6,7,8		
Question Number & Nickname	Brief Description	Context
Q5 “Car Brakes”	Three graphs show different ways of showing data. How do the graphs differ? Which graph shows more consistency?	Data & Graphs
Q6 “Reasons: MAX”	21 students tried to measure the duration of a MAX train ride. What are some reasons they did not all get the same result?	Data & Graphs
Q7 “Rounding: MAX”	The 21 students then tried graphing their data in two different ways. How do the graphs differ in the stories they tell about the data?	Data & Graphs
Q8 “MAX Wait-Time”	Two different graphs are shown: Wait-times for the Eastbound and the Westbound trains. What can you conclude?	Data & Graphs

Table 60

In “Car Brakes”, a table shows twelve data points that represent repeated measurements for the stopping distance in inches when testing a particular car's brakes. Three graphs are then shown, all of which present the data in a different way (see Figure 40). The axes are scaled differently on two different dot plots, and the third graph is a sort of histogram. All of the responses to this question fell in the aspect of displaying variation.

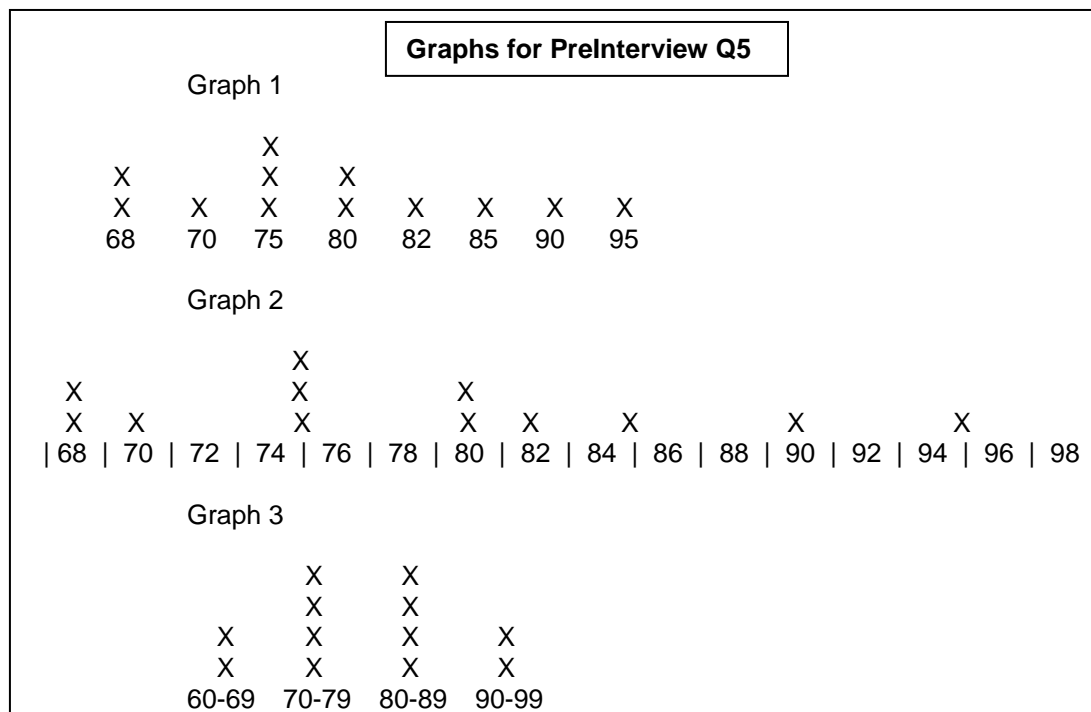


Figure 40

Displaying Variation: Within this aspect, most of the responses had to do with comparing the three graphs. Particularly for this question, the issues of level of detail shown or the usefulness of the graph were emphasized at some point by all the students. On a more fundamental level of graph sense (reading graphs), several cases took note of scaling issues, such as how the scale on Graph 2 was evenly spaced, but in Graph 1 the spacing was uneven:

- RL I like {Graph} 2 because the increments between each measurement that they are counting for are equally spaced
- JM Well, it {Graph 1} goes from 68 to 70, then 70 to 75, and 75 to 80, and then 80 to 82... so...No, I don't like that one, that's a little confusing for me
- SP Well, the first graph, which doesn't a ton of sense to me, but, she just wrote just the distances that she got. She didn't write anything in-between, and so you're just getting like, 68 jumping to 70, to 75, so here we have a distance from like 2 to 5, 5 again, 2 now... And so it's just sort of, doesn't represent what would be in-between.

Whereas JM and SP mention not liking Graph 1, DS says that Graph 1 (which has on the axis only the actual data points obtained from the twelve measurements) is “more specific” and also more concise:

DS At a glance, it's easier to see , if something's presented in a concise, efficient manner, you can look at it and go, okay, most of the times, it broke, you know, 68 to 75, but it did have these trials that were higher. And it's, you know, there aren't a lot of extra numbers in there, which is good, you just have the numbers that it broke. That the brakes worked.

Thus, for DS it is actually a kind of benefit that the axis on Graph 1 omits numbers which aren't part of the data set and effectively distorts the variation. EM said of Graph 1 that for her, “the impact is not as great.” Four of the cases commented on how Graph 2 gave more information, and for RL, Graph 2 was seen as more useful: “And so, it's {Graph 2} a more visual, it's more intuitive visually, more useful visually, graph.” Concerning Graph 3, although DS felt that graph did not include enough detail, she and EM agreed that it showed the general idea of how the brakes performed. Graph 3 was also seen by EM as easier to use for comparing the data, and several of the students had comments that suggested Graph 3 was somehow more concise.

EM Ok. Well, when I can see where the distances fell, {She is pointing to Graph 3} and if they're closer together, then it's easier for me to see how they compare, I guess you would say
GP Graph 3 really seems compact.

In comparing the graphs, the issue of the consistency of the brakes' performance came up, with Graph 3 showing the brakes to more consistent (especially in contrast to Graph 2):

- EM I think that Graph 3 actually tells me ... I get a better sense of where that car is generally braking, or where it's consistently braking.
- GP But, you know, if she showed them Graph 2, it would look like the car was really not being very consistent in its braking.
- JM Whereas we look at something like this {Graph 2}, it looks much more inconsistent.

The consistency of the brakes, and the conciseness of Graph 3 in contrast to Graph 2, largely fit with how the focus was more specifically on the larger apparent range or spread of the data when talking about Graph 2. All the cases except for RL made specific comments suggesting that they were attending to the larger visual spread of the data in Graph 2:

- DS Well, graph 2 you can see that also, except that it's kind of spread out, and it's not as, at a glance it's not as, like, you kind of your eyes go "wooaahh" {She dramatically waves hand from one side of table to the other}
- EM Graph 2 to me is too spread out, so I'm... I like seeing the Xs next to each other, so I can compare them easier, whereas Graph 2 is kind of spread out and I can't really read... There's more left up to interpretation.
- SP This graph 2 shows it more spread out, and it'd be like "Wow! 68 all the way over here, to ... 95" {Her hands are moving across the range}

Note how similar especially DS and SP's reactions to Graph 2 were, and how the bigger apparent range made such an impact on them. The final issue in comparing graphs was one of overall distribution, by which I mean the bell-shape exhibited by Graph 3, which was remarked upon by DS, JM, and RL, the latter of whom stated "The graph 3 looks like a pretty good bell curve."

Aside from comparing graphs, the other dimension of the aspect of displaying variation which came through in this question was that of making

conclusions about graphs. In this situation, when asked which graph an engineer should use “to suggest that the car was fairly consistent in its braking power,” all six cases agreed that the engineer should use Graph 3. Here are some sample responses given to explain why:

- DS Because it has a bell curve. And so, mostly, it does consistently brake between 70 and 89 inches.
- RL Oh, I’d go with {Graph} 3, because the {Graph} 2, is too, it’s spaced out, it’s kind of like you dropped an egg on the floor, it just oozes all over the place. It’s hard to pull it together, it’s hard to say something about it, and people generally make graphs so that they can justify what they have to say.

For RL, notice that he mentions a reason for even having graphs, which is as a justification, and this brings up the matter of whether or not a justification is really warranted. SP and GP go a bit further in terms of suggesting a misleading use of Graph 3 by the engineer:

- SP I guess this one {Graph 3}, because it seems less spread out. The graph 3. And so , it’d be like “Oh, look how close the graph is” {Laughs}
- GP Probably Graph 3, to show that it’s more consistent...Kinda fool them {Laughs}

So now, there is not only the issue of graphs to justify, but also to mislead. I probed RL about whether or not he thought Graph 3 was misleading:

- DC Does this misrepresent the data?
- RL It doesn’t misrepresent the data, but it does suggest more flexibility in interpretation, I guess.

RL’s comments are apropos of a true statistician or politician, and certainly worthy of Huff: Statistics don’t lie, they just offer “flexibility in interpretation.” It

is the variation in the data set, and the subsequent ways of representing the data – alternatively maximizing or minimizing the apparent variation – which lets this task be a good way to see what subjects speak about what they notice and what it means to them. Finally, in making conclusions, I was curious to see if anyone would mention the context of the task, which was the braking distance for a car. Both DS and SP had comments that reflected their attention to the importance of the context:

- DS Well, I think I would say, I would go with graph 1, because it's a little more specific on the inches, which could be a life-saving difference. And that having your, quite a few tests come up 82 or above could mean that they'd want to go re-adjust brakes.
- SP I dunno, with the ranges, it just seems like the range is pretty large {in Graph 3} And so... What is it? 70 to 79, especially when you're talking about braking distance. It seems important that you know more individually as opposed to clumping them together in 9 inches

Braking distance is important to these two subjects, and although the engineer can use Graph 3 to suggest the brakes are more consistent in performance, DS and SP want more specific information for decision-making purposes.

Question 6

The scenario painted in “Reasons : MAX” was that a class of twenty-one 6th- grade students wanted to find out some information about MAX train rides. Their first goal was to find out the duration of a ride from Washington Park to Gresham. They all got on the same train, but they sat separately and kept track of the time on their own. Later in class, they were surprised to find that they did not have the same results. After reading through this scenario

with each of my six cases, I asked for some possible reasons for why the class did not all get the same results. The aspect into which all of the responses fell was that of interpreting variation.

Interpreting Variation: Only one dimension of this aspect was brought out, as was the intention of the question, and this dimension concerns causes of variation. Plausible causes were suggested by each of the cases, and these causes all seemed to be one of four different types. The description of these types, as well as who among the cases mentioned this as a cause, is listed below in Table 61:

Summary of Causes in Q6						
Description of Cause	Subjects Mentioning the Cause					
[1] Where or When to Start or Stop Timing	DS	EM		JM	RL	SP
[2] Watch Functionality or Type	DS	EM	GP	JM	RL	
[3] Student Error			GP		RL	SP
[4] Where the Students are Sitting	DS		GP	JM	RL	

Table 61

An example of what each case suggested is provided next, and cumulatively these examples help illustrate each of the causes:

- DS Well, they could've started their watches at different times, like, within a second or two, and they could've stopped them at different times.
- RL Three reasons I can think of, are they didn't start timing at the same time. They didn't stop timing at the time, and the devices they used to time didn't run at the same speed.
- SP Probably how they were defining when it started. Like maybe some started right when they got on and sat down, maybe some started right when the train started moving... And when they stopped it, same thing.

- EM I don't know, they could also be all using different types of timing devices... Somebody could be using a stopwatch, somebody could be using a watch, electric, you know, or electric watch
- GP Maybe... You know, they don't have very precise watches, maybe the watch, something's wrong with the watches, you know...Need a little more grease in the watches. Some might've forgot to stop their watch and then : "Oh, no!"
- JM They're all sitting in the same car? They're not. Ok, so when are they supposed to start? When the train starts moving?

Each of the cases had no trouble thinking of multiple causes for variation for this situation, but the age of the students in the task and the context (the twenty-one 6th graders on a field trip, a scenario which experience suggest can be rife with chaos) seemed only to get the most peripheral of emphasis. As shown, some of the responses mentioned the need to synchronize the starting and stopping procedures for timing the duration of the train ride, but very little was said to explicitly suggest that the middle schoolers simply were unfocused on the task, which I would probably have cited as a fundamental cause of variation. GP's response brought this cause out a little, as did some allusions by RL and SP.

Question 7

This question flows directly from Q6, keeping the same scenario and the same set of 21 data points (the times collected by the 6th graders).

This time, the data has been graphed by the class as a stacked dot plot in Graph 1, with each data point rounded to the nearest 15 seconds. Also, the class graphed the data as a stacked dot plot in Graph 2, rounded to the nearest 5 seconds. The two graphs for Question 7 are shown in Figure 41 on the next page.

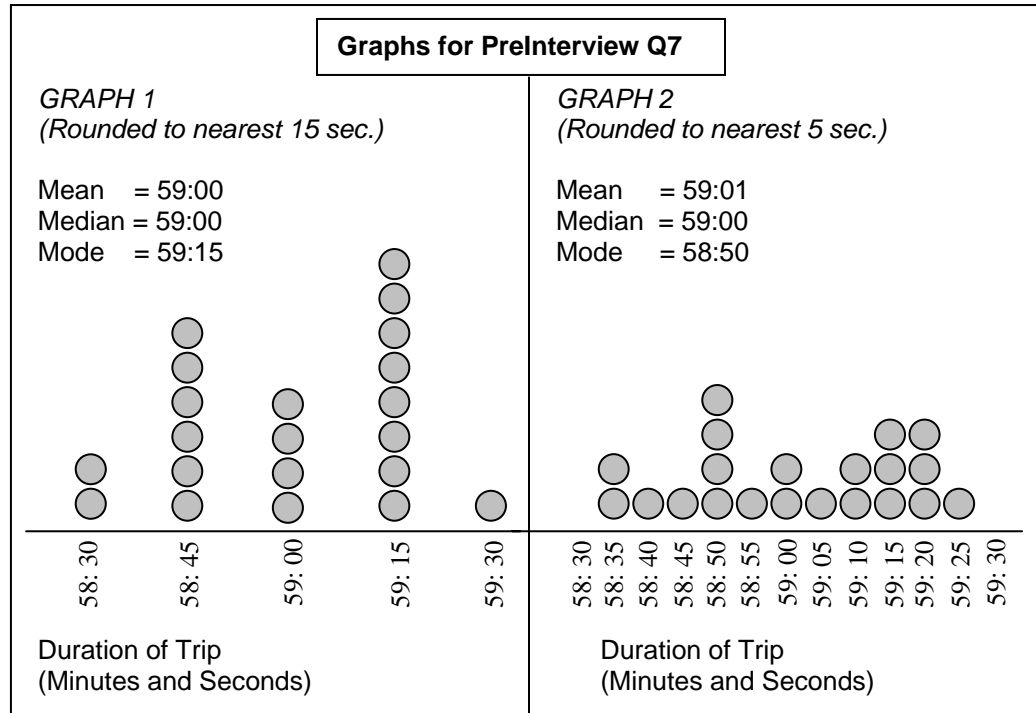


Figure 41

I told the cases that the 6th graders wanted to compare these two graphing styles and see which one they liked better. The specific questions I used either as direct prompts or as probes were:

- Q7a: How do these graphs differ in the stories they tell about the duration of the trip?
- Q7b: Some members of the class argue that the trip was really under 59 minutes, while some argue that it was over 59 minutes. Others claim it was exactly 59 minutes. What do you think about the true duration of the trip, and why?
- Q7c: Does one graph help you more than the other in making your conclusion?

Sometimes, after only reading Q7a, the responses elicited enough information that I felt Q7b and Q7c had already been addressed. Collectively, the responses fell into the aspects of displaying and interpreting variation, but primarily into the former.

Displaying Variation: The dimension of *evaluating and comparing graphs* received the most attention for this question. All six cases attended to measures of average. In addition to having the table of raw data on the same page as both graphs, there were summary boxes within each graph showing the mean, median, and mode for each set of rounded data (see Figure 40). In a stacked dot plot, the mode tends to be a visual attractor, and so it was not surprising that all the cases commented on the modes. GP makes this point:

- DC You can draw on this too, if you like.... What about Graph 1?
 GP Graph 1 just looks like, um, the 59:15 is really tall.
 DC Ok, and that might make you think...?
 GP That most of the kids, that was the time. But then you go on the other one, and it's at 58:50 is the highest one. So, um...So it kinda shows something differently. I mean, your eye usually goes to the tallest one, there, and you see... And so, they come off kinda {like} different results

Another feature of the graphs that drew some attention was the spread of the data, with the range as a particular focus for EM and DS:

- EM The Graph number 1 tells me that there's a whole minute difference between the time, you know, of all the times they took. And then this other one {Graph 2} tells me it was, like, 55 seconds, 50 seconds difference.
 DS Because it took somewhere between 58:30 and 59:30 {She traces out the range of Graph 1 with her fingers}

However, DS then considered what I'll call a sub-range, meaning she discarded the two data categories on the edges of Graph 1 and looked at the data within the trimmed range:

- DS Well, because, more people experienced that time frame {She points to the three high bars in the middle of graph 1} than...the 59:30 or 58:30. So there's a higher chance that those people on

the outside were incorrect {She points to the extremes of Graph 1}

The response of DS showed a shift from the just the range to a consideration of how the data was distributed, and four additional cases showed how they were either looking at different intervals besides the entire range, or how they were actually counting data points to see how many were in different areas of the graph:

EM Well, I can see that there's more people clocked 59 seconds or above, just by looking at it {Graph 1}. Now, if I look at graph number 2, I would say it's about even, without counting the little dots, or whatever, that actually, more or about the same clocked under 59, so ... 59 or below. So, almost, graph 2 almost tells me something different. But, I'd have to count it to find out.

One of the more interesting examples of attention to distribution came from GP, who was the only case to actually try re-distributing the data to see how it balanced out:

DC Well, what are you thinking of? Why don't you explain it to me.
 GP Well, I'm seeing here {Graph 1} there's more 59:15's and there's less 58:45 here {Graph 1}, and uhm... So, you kind of have to, so if you're going to go one here, one here, one here, one here...{He re-distributes, from the columns adjacent to 59:00, onto the 59:00 column} ... Um, you know, you'd have more left over here... Two more left over here {In column just to the right of 59:00} So, it would be more, it would be a little bit more than 59. I dunno {Laughs}

GP continued this re-distributing of values in Graph 2. His thinking was possibly influenced by class experience, since a visual model of redistributing values as a way of understanding the mean had recently been featured in the Math 212 classroom. Students had built towers of tiles of varying heights and

re-distributed to find the average. In this situation, GP moved data around in an effort to compare the two graphs, and I found his thinking in this task to be reflective of a high level attention to distribution and to variation. Explicitly referring to the relative amounts of variation, SP says about Graph 2 that “it shows that there WAS a greater amount of variation.”

In fact, amount of variation shown and the rounding procedures used did come to bear on which graphs the cases found more useful or accurate. Except for DS and GP, the other four cases commented on the effect that rounding had on the resulting graphs, and SP cites the amount of variation of Graph 2 as reason for why she likes that graph:

SP Just because it shows that there is greater variation in the times that they were {getting}. It shows, it’s closer in reality to what they did get.

EM agrees that “Graph 2 is definitely more helpful because it’s closer to the accurate time, and although GP first had the thought that Graph 2 was more helpful, he later changed his mind (the numbered lines of the transcript are provided in square brackets to show how close in time these following two comments are):

GP [428] I think graph 2 is more helpful.
 GP [473] I think this one {Graph 1} is easier... This one {Graph 2} gets a little confusing, you know. This one {Graph 1} is, you know, if you were going to talk about it, it’d be easier to do this one

It seems that the fewer categories on Graph 1 compared to Graph 2 changed GP’s mind, and I think this also fits with his earlier attention to the mode, or

high bars, which have more impact in Graph 1. Some sample responses that explicitly tied the rounding procedures to the variation shown or to the usefulness of the graph include:

- SP Well, because these {Graph 2} are rounding to 5, then you're getting greater variation in the answer
- EM Yeah. I like the idea that they just rounded to the nearest 5 seconds, because if you're trying to track the time, you'd want to get closer to that.
- JM You get a more precise time with the second one, with graph 2. With the numbers at that particular time, since it's in 5-second increments.
- RL If it comes down to "I need accurate information," I want smaller increments.

Even though SP's response does seem to link the rounding to a cause of variation, I have categorized responses about rounding here with the others about the level of detail and usefulness of graphs. I made this decision because the main emphasis of the question was on how different claims could be made about the duration of the trip, and also because her response had more to do with variation in the way the display of data looked rather than what caused variation in the data that was explored earlier in Q6.

The last dimension receiving attention from responses to Q7 is *making conclusions about graphs*. EM, GP, and JM both readily volunteered that Graph 1 would be influential if 6th graders wanted to argue that the true duration of the trip was over 59 minutes, and also that Graph 2 would be useful in arguing that the trip was under 59 minutes. These opinions also correlate with where the mode lies on each of these graphs. When asked what they themselves felt about the true duration of the trip, all six cases

expressed comfort with an average answer, meaning that they gravitated towards one of the summary statistics (mean, median, and mode) presented within each graph:

- DS It look, uh, about 59 minutes. {Laughs}
 EM I'd actually... It looks to me like 59, I'd pick that time. That it TOOK 59 minutes
 GP Well, why don't we just say 59, right in the middle?" {Laughs} You know?
 JM Well, I would conclude that it was right at 59, right?{Laughs} I'm just reading the... I'm reading the average {Laughs more}

It should be pointed out that 59 was the median for both graphs, and that was the only summary statistic to be identical for both. DS and EM also had mentioned comfort with a range answer, with EM adding that "looking at graph number 1, I would say, on average it took over 59 minutes."

Interpreting Variation: Among responses top Q7, there were only two brief references to this aspect: One was RL's reference to *causes* of variation. This is where he chose to make explicit the idea of student error as a cause, saying "they, you know, had ate too much candy, they weren't thinking about it, whatever, they were busy thinking about something else." That is, he supplements his earlier Q6 response even as he was thinking about Q7. For *effects* of variation, I have included SP's feeling that you could use either graph to make arguments either way about the duration of the trip. I have coded this feeling of SP's as an effect of variation because it seemed to stem from the variation in the data itself. In other words, it is because of the data's

Displaying Variation: In *evaluating and comparing graphs*, all six cases focused on the averages, noticing that the mean and median were the same, For JM this was his immediate observation:

- JM Well, looking at the graphs, they're both the same, right? {He points to the boxes of summary statistics}
- DC Looking at the graph, they're both the same? Because your finger's pointing up here...{to the summary statistics}
- JM Well, not looking at the graph, but I'm looking at the BOX, right. I'm just reading the mean and the median. And, uhhh...

Conversations with the cases then tended to proceed to other things they noticed, with all six cases having some comments comparing the ranges, such as how the eastbound trains had less spread or how the westbound trains had more spread:

- RL Well, it looks like the {Eastbound} train, the range of wait-times is pretty limited
- JM On the Eastbound train it goes from 8 and a half minutes to 11 and a half
- GP Well, they're closer, they're all closer together here {Eastbound}. This is more spread out {Westbound} So... {He shows with his hands what he means}
- EM But you also have two ends of the large spectrum on the Westbound train, and a shorter spectrum on the Eastbound train.
- SP It {Westbound} goes from 7 all the way to 14, and so that's... seven minutes.
- DS Because it {Eastbound} varies between 8-and-a-half and 11-and-a-half minutes, so it's a shorter time variance.

The differences even in the language they use to attend to range is noteworthy, as the subjects also use terms like spread, spectrum, and variance, in addition to using gesture. Some responses that included more elements of distributional thinking were those that had language suggestive of

relative tightness, clustering, or comparing portions of the distribution as JM does in this example:

JM Even though we have 2 thirteens and a fourteen or a fifteen minute on the Westbound train, that's just 3 trains out of , what? Out of ten? Ten trains? So 30% of the trains take longer than 12 minutes on the Westbound train, and then , um, you know, it looks like 100% of the trains are under 12 minutes on the Eastbound train.

Similar to the "Car Brakes" (Q5), in *making conclusions about graph* Question 8 elicited comments about the consistency or reliability of the different train lines. Whereas in Q5 I specifically introduced the word "consistent" as something the engineer was trying to promote, in Q7 I just asked "What can you conclude?" DS, EM, and GP specifically noted that the westbound trains were less reliable, and all six cases pointed out that the eastbound trains were more reliable or consistent.

SP Well, the Eastbound train is more consistent
 GP Well, it's just the Westbound train is less consistent, you know. But the Eastbound is more consistent. It'd be more, you know... Closer together.
 JM Well, the Eastbound's are much more reliable.

Moreover, as DS did earlier, so too do EM and RL explicitly refer to variance or variation:

EM Um, from looking at this, I would say that the Westbound trains are less consistent in their wait-times. There's more variance. So, you can be waiting 7 minutes, and you can be waiting 14 minutes. And then, the Eastbound trains are pretty consistent
 RL Looks like the Eastbound is more reliable. I would say that the Eastbound is more predictable, and less variation.

RL's comments fit well with the earlier notions of variation that were explored in the PreSurvey. That is, variation implies for many a lack of predictability – More variation, more uncertainty, and less variation, more predictability. Lastly, in making conclusion, it is interesting the way in which both EM and DS make the context a bit more personal, picturing themselves in the situation of waiting for the train:

- EM Well, I would rather be waiting for the Eastbound train, because I know “Hey, I’m either on that train somewhere between 8-and-a-half and 11-and-a-half minutes”, and I like that consistency.
- DS Well, I would say that if I was waiting Eastbound, I would be more, umm, sure that my train is going to arrive, and I’m going to get where I’m going on time. Westbound could just show up whenever it wants to!

I found it interesting how the language of consistency and reliability came unsolicited from these subjects. I wonder whether this choice of words had more to do with earlier questions and discussions we had in the interview or in class, or more to do with the way that all the cases were familiar with riding the bus or the train.

Interpreting Variation: *Causes* for variation in this question were brought up only by DS, who had a keen observation about why the wait-times were not all uniform. She said, “And as the day progresses, the trains, you know, stuff happens, people step out in front of the train, you have more red lights, and so, then the schedule is not going to be as precise.” As for *effects*, again the personalized notion of how long one is willing to wait for a train came

to bear, as well as the likelihood of having to wait a certain amount of time.

Note in the following examples how GP and DS use probability language:

- SP You have a greater assurance that you're NOT going to wait more than 11 and a half {Eastbound} . Where HERE {Westbound} it could be 14 minutes. Even if SOMETIMES it might be 7.
- DS But I might NOT get really lucky, and then I'd be 14 minutes waiting. So I'd rather have ... know how much time I'm going to wait, than {to} risk. Especially if I have a meeting...
- GP Westbound train seems more of a , uhh, more of a gamble, you know.

Of course, the ultimate effect of the variation in the wait-times, is that you can't be certain. Both RL and SP had this view:

- RL Right, well, it's not even so much the waiting, it's the not knowing that the waiting's going to end.
- SP <Laughs> Well, it just seems really hard to make any conclusion, because there is such variation...

I was pleased with how the subjects all seemed to connect to this task, in the sense that very little probing was needed to get them to talk about what conclusions they drew, or what the graphs meant to them. It may have helped that the culture of Portland State University, being an urban campus, is very much tied to issues of transportation, and public transportation in particular. In any case, it underscores for me the importance of having tasks and contexts that draw in the subject and naturally gets them talking about and reacting to what they see and understand.

Question 9

The context of the interview now shifts away from Data and Graphs and into Probability. The first three questions (Q9, Q10, & Q11) in the Probability

context all depended on tossing a die, and thus are listed together in Table 62:

PreInterview Questions #9, 10, 11		
Question Number & Nickname	Brief Description	Context
Q9 "Sixty Tosses"	A regular, fair die is tossed sixty times. Fill in the table to show how many times you think each number might come up.	Probability
Q10 "Who Cheated?"	A teacher asked 4 students to do the activity in Q9. Only one of the students really did it & the rest cheated. Who really did it?	Probability
Q11a "Repeat Trial"	Consider how many "5s" you put in Q9. If you rolled another sixty times, do you think you'd get that many 5s again?	Probability
Q11b "Six Trials"	Six friends each rolled the die 60 times. Put down how many 5s you think each friend might get in 60 tosses.	Probability

Table 62

For the first question in this context, Q9 ("Sixty Tosses"), students are asked to consider the results of throwing or tossing a regular die sixty times. A table is shown with the six different possible outcomes, and a blank space by each outcome where the student can predict the frequency of each outcome. The wording of this task is critically important, because if students interpret the question as asking if they know that each outcome has a one in six chance of occurring, then they are apt to write all tens for the frequencies. My wording of the ask was to "Fill in the table below to show how many times you think each number might come up," and either the subjects read this for themselves, or else I read it to them with an emphasis that this was asking them for what might happen if they really did the experiment. It usually became apparent that the emphasis was on what they think *really* would occur. Then, if they thought that all tens would actually happen, they could defend that choice. The framework aspects drawn out by this question were *expecting* and *interpreting* variation.

Expecting Variation: For *what* was expected, four of the six cases immediately wrote tens all the way down the table. EM wrote “8,9,10,11,12,10” and GP put “7, 13, 8, 12, 9, and 11”. While any select list may have a low likelihood of occurring, the task was not about a right or wrong choice but about what their selections said about variation. From a probabilistic point of view, the safest thing to say about the actual experiment is that it is very likely that the numbers will not be same, and very unlikely that all tens will result. Yet from the outset, DS, JM, RL, and SP seemed to expect zero variation. RL, in particular, was unshakable even as I made sure that was what he thought would really happen:

- DC What do you think might happen?
 RL I think that's {10, 10, 10, 10, 10, 10} going to happen.
 DC Ok. You're very confident. You think that's going to happen. I feel like getting the die right now. Let's throw it 60 times, Rob
 RL Alright. I will describe it a little differently. The first time I do it might be an outlier. For all I know, I might roll 6 sixty times {He writes 0, 0, 0, 0, 0, 60 next to his original 10, 10, 10, 10, 10, 10}. Can it happen? Sure it can happen.
 DC So you're telling me you think it's 0,0,...60?
 RL I think that's {10, 10, 10, 10, 10, 10} going to happen.

DS was the first of these four students, initially comfortable with all tens, to recant. This occurred while she reflected back to a similar context involving sampling:

- DS But then, going back to that question about picking the colored candies and the 6,6,6,6... That's kind of ... 10,10,10,10 is kind of like 6,6,6,6... so it probably will vary somewhere.

Meanwhile, EM and GP vocalize their feelings that the same results would not always occur:

- GP Well, I knew it was going to be random, and so I first looked at the 60, and I said, well, these all have to add up to 60. So I divided by 6, and I said 10 each. And then I said, well, it's not going to be happening, 10 for each one
- EM Ok, out of 60 times you roll ten 1s, ten 2s, ten 3s, but I just think that won't happen. I don't have a calculation, but in my head, if I threw it, I think I'd see the same number at least once every seven times.

For EM, this was among several instances that she referred to some supposed mathematical tool that might help her definitively address questions of expectation, perhaps yielding the "right" answer. She was even more explicit at another point in discussing this question, saying "I'm sure there's a formula for this. But, I don't know." For the situation of Q9, however, there is some sort of timetable in her mind, a threshold of throws after which she does expect to see the same result.

As far as the reasons *why* behind the expectations, all six cases offered up reasons based on what I am categorizing as probability arguments, and I link these more with arguments about possibilities and likelihoods than with strictly proportional reasoning:

- GP Well, each, each side has the same chance of getting a number, so...
- SP Because I'm just giving them each an even chance. {She sweeps her hand} 'Cause each time you roll the die, it could be any of the numbers. There's not like, a greater chance of getting one than the other.
- RL Because, rolling any given number is no more likely than rolling any other number.

Although I believe that the one-in-six notion is fully behind all these kinds of comments, the main thread that links these comments together is that they all

deal with chance or likelihoods – The probability of one thing happening versus the other. The above responses do not emphasize ratio, but do suggest what probably would or would not happen. EM includes personal experience in her reasons for variation, explaining that she relied on

“Instinct, thinking about when I’ve played games, and how often sixes came up, and how often fours...”

Interpreting Variation: The only dimension to come out of this question was that of the *effects* of variation, which in this case were that anything could happen, and therefore one can never know for sure.

JM I mean, I don’t know what I’m going to get. I might shoot , I mean, it’s like the game of craps, right?

SP I just... I dunno, I just... I can’t guess. I have trouble making guesses because...Because you KNOW when you play that it could be... You COULD possibly get all 2s, the entire time. I mean, it’s most likely not going to happen. Again, I’d be really surprised. But, there’s a possibility. I guess, you know, I can never know

More information on how these cases think about the dice situation of Q9 develops over the next couple of questions, which also involve trials of sixty tosses.

Question 10

Whereas Q9 asks subjects to imagine what would happen if the die were to be tossed sixty times, Q10 (“Who Cheated?”) supposes a classroom where the teacher told his students to go home, do sixty tosses, and record the results. The purported results from four different students are then shown, but the narrative of the task indicates that only of the students really did the

experiment, and the other three just made up the results. The question asks what is most likely – That Riki (one of the students) really rolled it, or Lynn, or Lee, or Pat. These supposed results are listed in Table 63. Another option for what is most likely is that “No one can say. Any of the 4 students is equally likely to have rolled it.”

Supposed Results of Four Students (PreInterview Q10)				
Side That Faces Up	Riki	Lynn	Lee	Pat
One “1”	7	10	10	2
Two “2”	12	11	10	15
Three “3”	6	10	10	10
Four “4”	9	10	10	28
Five “5”	14	9	10	1
Six “6”	12	10	10	4

Table 63

In a practical sense, Lynn exhibits too little variation, Pat exhibits too much, and Lee exhibits none. That leaves Riki as having the most plausible list of results. Responses to this question helped show more about the aspects of expecting and interpreting variation.

Expecting Variation: In terms of *what* is expected, all six of the cases thought Riki was most likely to have really rolled it, and all except SP were outspoken about how Lee’s list was unbelievable. Especially for DS, JM, and RL, who put all tens (just like Lee) as their first thought for Q9, a fuller portrayal of the conversations is warranted, because it exposes a sharp dichotomy in thinking.

- DC Explain your reasoning, please.
 DS Well, because really, the 10,10,10, would be so unusual that it would come out that way.
 DC Ok, it would be so unusual, and yet that's exactly what you said you thought might happen {Turns back to Q9}
 DS Well, I don't really think that it's going to happen.

With DS, she sees her own choice of all tens as unusual and something that it not going to happen. For JM, all tens is possible, and likely, but still not going to happen:

- JM But I don't think, even if I rolled them 60 times, I would not get 10 numbers each.
 DC Now, hold on, hold on. You just told me... {Turns back to Q9}
 JM {Laughs} I think it's possible, likely, because it's one out of six times, but I don't think I could roll that and that'd actually happen. It's not going to happen.

And for RL, he takes care to show how all tens is a reasonable choice, but it's clear that he doesn't believe anyone is going to get that result in an actual experiment.

- DC Now, what are you saying about Lynn and Lee? You think they really did it?
 RL No, I don't. I don't believe that they actually got that. But, I think it is entirely reasonable that they would say beforehand, like I just did {on Q9}, that doing that is more likely...
 DC And yet on the previous page, you said that that's what you think would really happen? {Turns back to Q9}
 RL We're living in the real world, this is not going to be 10,10,10...We're going to be, here's a 12, but here's an 8... {On Q9, he's changing his choice to 12, 15, 9, 11, 8, 5}

RL offers additional reasons for this change of mind, and these reasons will be looked at shortly. For now, it is interesting to wonder what it is about the structuring of Q9 and Q10 that brings out the following contrast: In an a priori situation, where they are asked what might happen (and these expectations

are listed before the experiment would be done), all tens seemed reasonable to some of the students. But in a posteriori situation, where the experiment is said to have already been done and we are looking at results that purportedly came out of the experiment, all tens became unreasonable. SP was less emphatic, although she did soften a bit on expecting all tens:

- DC Let's talk. What do you notice about Lee, for example?
 SP She's all... Lee's just like my guess, and I...{Flips back to Q9}.
 This is not exactly what I would really expect to see.

GP is more direct, saying: "I look at it and go: Come on, Lee! There's no way that this happened!" Some sense of what the subjects *would* expect includes the idea that outcomes (such as ten for each face of the die) won't repeat, and this focus on repeating or non-repeating numbers also came up in the sampling context. Some cases also mentioned expecting a range or spread of results.

- SP I like more spread as opposed to more of like one thing
 EM I think that just sometimes, you know, that you'll get more sixes than you'll get fours, I just think that will happen, I don't really have any scientific evidence for that. Or mathematical evidence.
 JM I don't know what I would expect to get, but I don't... I think it would differ. I don't think I would get 10 each time
 RL The reason why I would not go for this {10,10,10...} after all is because you're going to see a range of results.

The response of RL clearly shows how a thought can be coded as both showing *what* is expected and *why*. He expects to see a range of results, and this is offered as a reason for avoiding Lee's list. So, in theory you could have a subject say something like "I expect outcomes to vary," and this seems to speak to *what* that person expects. But that same subject, in talking about

why he or she has that expectation, might say “because there’s going to be variation anytime you do an experiment like this.” The point is that I have tried to make a distinction of emphasis, so that when subjects are suggesting ideas as reasons for their thinking, I coded these ideas in the *why* dimension of expecting variation. Often, a response from a subject could show both *what* and *why*.

The idea of randomness as a reason for expectation was given by DS, EM, GP, and SP. As in DS said: “Because it’s... a random thing.” And by this she meant the nature of the experiment – the rolling of the die – was random. For SP, Riki’s results were seen as more likely “just because it seems more random.” The language of probability came through clearly in many explanations, as cases commented on the chance of certain outcomes, or relative likelihoods, and RL reiterates the following idea regarding Lynn and Lee’s list: “Because, strictly speaking, those are most likely. To roll around 10.” However, there was agreement from all six cases that Lee and Lynn’s results were possible but unlikely. This consensus, held by all the six cases, was interesting because it showed a softening on extreme results:

- GP Yeah, it {Lee’s} could happen, but uh...Well, I mean, {it’s} very very unlikely
- JM It {Lynn’s} could happen, I just think it’s ... it’s more unlikely for that to happen.
- SP It {Lee’s} could happen. But it most likely wouldn’t, for some reason?
- DS It {Lee’s} could but it doesn’t.

While Lynn and Lee’s lists of results in Table 63 were characterized by a lack of variation, Pat’s had too much variation, and in particular the extreme

outcome of getting only one “5” and getting twenty-eight “4”s was deemed unlikely (but possible):

- EM I mean, it {Pat’s} could happen, but, that seems a little more far-fetched. Yeah, the 28 makes me nervous. That seems like too many 4s.
- RL Pat has got the 28 here. Which is SO unlikely, that either, okay, it’s just one of those unlikely things, or it was, she’s trying to appear like unlikely.
- DS Well, Pat’s is a little TOO broad. For only 60 rolls. So it’d be very rare that you’d get {a “4”} 28 times
- SP It just seems awfully high to get 4 twenty-eight times.

The last theme for reasons *why* had to do with personal experience, and was illustrated by EM, GP, and JM. They each explicitly drew from their own memories of playing games:

- GP I’ve played board games, and I’d roll dice, and you get 5 more than that {Pat}, you know
- JM I’ve played Yahtzee enough that you just can’t ...the chances of it happening all the time like that {Lee’s} you know... have the distribution just flat like that, {Shows with his hands a uniform distribution} each one gets ten times, in 60 times, um, it’s just not as likely. I don’t what the odds are.
- EM Instinct, thinking about when I’ve played games

Note that when they made their own lists on the previous question (Q9), GP and EM did not put all tens but JM did, although “it’s just not as likely” in his mind.

Clearly these two questions, Q9 and Q10, generated considerable commentary from the subjects on a wide range of both *what* is expected and *why*. For the four people who wrote all tens on Q9, they went back and changed their lists after denouncing Lee’s list. The six cases, taken together, discussed expecting to see a range of results, and how they wouldn’t expect to

see Lynn's or Lee's or Pat's results, but that Riki's were reasonable. They offered many reasons, such as how results should look random, or how extremes (both too much and too little variation) might be possible but unlikely, and what some lessons were from their own experience. RL was alone in offering a reason for expectation that was based on a synthesis of both average and spread, and he used this reason first as a basis for changing his own list from Q9.

RL I'm changing my mind {on Q9} because I'm making the same mistake that I was accusing some kids earlier of, and that was I was considering average but not considering variation. And you need to consider variation to get the full picture. So, yeah, if you're going to see a range, the average of that range will be 10, but not every response will be 10.

It is important to note that, although RL has moved on from an "all tens" position, he still has the sense that "the average...will be 10." He also uses average and spread in his evaluation of Riki's list:

RL I like two things about Riki: One, there is a range of results, from 6 to 14. But, the general average is roughly 10. So, it strikes me as more realistic.

Interpreting Variation: Concerning *effects* of variation, the idea that reality is different from probability came out earlier in the sampling context. Here the same idea shows up again, as in RL's earlier comment that "We're living in the real world, this is not going to be 10, 10, 10...". For example, GP describes Lee's results as "too perfect", and JM refers to his uniform distribution earlier as "perfect." When talking probabilities, there is a common convention of talking about the expected value as the perfect result, as if that

is what might happen – results might turn out perfectly – if it weren't for variation. Thus, I categorized this kind of thinking as an *effect* of variation. DS was the most vocal spokesperson for thinking about results as “perfect” versus what really happens:

DS Because it {Lee's} is TOO perfect. Life doesn't happen that way. There's chance involved, so, whenever there's chance, then things won't necessarily turn out perfectly. Like in a perfect world situation, where the dice was loaded.

If a loaded die connotes to DS a “perfect world situation”, then it is probably because there is a higher degree of predictability in a loaded die, as opposed to the unpredictable nature inherent in a regular die. She goes on to compare chance to the perfect world:

DS Well, 10 would be like the perfect world, and he {Lynn} only went, over one, and under one. Where really, chance could probably have a broader range ...

Thus, an effect of variation is that outcomes won't be perfect every time in the real world. Finally, even though results won't be perfect every time, according to JM and SP the results can be anything; and if results can be anything, then you can never know for sure what those results might be. In commenting more on Lee's list, SP shared the following:

SP See, I would look at that and be like, “Oh, he was just forced to guess” and he didn't really do it. Just like I was. ‘Cause I don't think I would really expect it to look like that, but... When you're making a guess, I just do it that way, because you can never really guess. Because there's always a chance that any of those numbers could be anything.

This comment is interesting for three reasons. One, it shows how SP has a strategy for when she has to make a guess, which she used in generating her list of all tens in Q9. Two, it shows what guessing really means to her. It means to guess and then be correct. So when she says “you can never really guess”, she means that you can’t guess and then be assured that your guess will turn out to match the results. Why not? Because of variation, and how “those numbers could be anything.” Three, her comment says the following about the effects of variation: It is because of variation that, for SP, results can be anything and thus, she can never know for sure.

Question 11

The last of the PreInterview questions on die tossing, Q11 re-framed the trial so that out of 60 tosses, only the number of “5”s facing up was recorded. That is, instead of focusing on the frequencies for each face of the die, only the “5”s were of interest, and the choice of “5” was arbitrary on my part. The question starts by referring subjects back to their list (or revised list) of Q9 and making note of how many “5”s they had predicted. The reason I focused on just one face of the die was because I wanted to simplify things. This task asks about sets of repeated trials and I didn’t want confusion to result from keeping track of how many “1”s or “2”s there were. Also, I felt this made the situation more analogous to the sampling context earlier. In other words, if Q9 was similar to Q1a (“One Trial”), then Q11a and Q11b were similar to Q1a and Q1b. The first part of Q11 (Q11a – “Repeat Trial”) asked if subjects thought they would get the same number of “5”s as in Q9 if they did

another set of sixty tosses. The second part of Q11 (Q11b – “Six Trials”) asked for how many “5”s they thought six friends might get in each of their sixty tosses. Aspects of expecting and interpreting variation were again both illuminated by the responses on Q11.

Expecting Variation: Looking at *what* was expected, there were comments focusing on the expected value (10 “5”s in this situation). For example, EM’s and RL’s comments suggested that results should be close to that value of 10:

EM I like the idea of being kind of close to 10.
 RL I’m expecting to see roughly 10.

Another clear sense of expectation came through regarding repeated values. For Q11a (“Repeat Trial”), all except for JM were thinking more in terms of not getting the same amount of “5”s. When asked if they thought they’d get that many “5”s again, JM said “maybe,” but the others were more strongly against that possibility:

EM Ah, yeah, I’d be surprised if it came to be the exact same number as the day before. I mean, yeah, it could happen, but I’d be surprised.
 GP Probably not. I mean, it’s possible, but...
 SP No. I think that it’d be more random. It could be, I think... I feel it would change every time. But, I mean, of course it could be the same, as well. I can’t say in all certainty that it’s going to be different, but I feel strongly that it would probably be different. But there’s always the chance...

The sense from most of the cases is that the value from the first trial probably wouldn’t repeat on the second trial – they wouldn’t expect it to – but it might happen anyway. Another way of seeing this is in the way EM, GP, and SP

talked more in terms of being surprised if they saw the same result on the second trial, whereas JM said “No, I would not BE surprised.” However, when it came time for Q11b (“Six Trials”), none of the cases put all tens, a clear indication that they did not expect the same results on the six trials.

In describing *why* they held their expectations, the only real commonality was that all their comments had to do with possibilities or likelihoods. All except GP offered some reasons that involved the chances or likeliness of different outcomes. For instance, SP spoke in the last response given earlier of how “it’d be random,” and how “there’s always the chance” of something unexpected happening. DS also used language of chance, saying “there’s more chances that it’ll come out different than 5,” and EM referred to the “luck of the dice.” The idea for extreme values is that such an outcome is “highly unlikely, but it could happen” (DS). SP had a moment of reflection about her own choosing procedures, noting that “But for some reason I don’t choose the really low, the 0, or the 60.”

Interpreting Variation: *Causes* of variation were brought up by only one student, who suggested a physical reason behind the way the die was tossed:

EM I just don’t think you’re likely to get the same answer every time. There’s no way you could do that unless you know how to drop the dice, or something.

The next dimension illustrated was the *effects* of variation. The primary point made by the cases was once again that results can be anything, and therefore you can never know for sure what will happen. It is interesting to wonder just

how strongly SP holds the notion that “I know that it could be anything” since, as she reflected above, she also knows that she avoids putting the maximum and minimum possible result. For this question, EM and JM join SP in expressing the notion that you cannot know for sure:

- EM You just never know what you’re going to get.
 JM It’s impossible to know. Because you can’t predict the future. I mean, you can predict probabilities, the odds of it possibly coming out, but the chances of it happening, just like your prediction can be pretty accurate, depending on how many times you do it, but you can’t predict it.

Notice how JM’s response about the impossibility of prediction matches with SP’s earlier reluctance to guess, and the idea that one can “never guess.” The meaning becomes clearer that these two mean you cannot predict with 100% accuracy all the time in situations involving variation. To “predict the future” in JM’s mind seems to equate with “to make a statement about the future which then turns out to be true.” Another effect, as expressed by DS in the last question and then brought up again here in this question, is that variation means things won’t be “perfect.” I took this opportunity to explore with DS a bit more about something she had said several times earlier in the interview:

- DC I would like to push you on that a little bit more: “The more chances you do something, the more chances that it’ll be different from...” {I am quoting her previous remarks}
 DS From perfect, the expected. It is an idea that I hold. So, ‘cause I think that, um, that there IS the chance that it’ll come up perfect, but there’s ... “perfect” is one {Holds up hand to signify one number} , and there’re more things that are imperfect, like, not perfect. {Moves hands to show distribution of other numbers on either side of the perfect number}. So, there’s a lot more options for the imperfect.

This exchange helped clarify for me how she was interpreting variation. The effect of variation for DS is that outcomes will be different, and if there is only one perfect outcome, then all the other outcomes are less than perfect. So variation is responsible for the imperfect results that occur in the real world.

Influencing expectation and variation occurred in responses by DS and RL, and in one comment by JM that was shared earlier. JM had implied that the accuracy of your prediction could depend on “how many times you do it,” and this idea gets at the fundamental connection between the number of trials and the amount of variation. RL and DS expand on this connection as follows:

- DS Right. So we if had more sets of 60, then I would make my numbers go lower {Showing with hands a greater range}
- RL Well, I actually represented here a pretty limited range, just between 8 and 12. But we’re only talking about 6 people throwing, and when you’ve got 6, it’s a pretty small sample size. So, chances are you’re not going to see anything too goofy. You get a hundred people doing this, you’re definitely going to see the extremes pop up more often.

DS and RL saw an expanding range resulting from increased trials. Put another way, DS and RL see increased trials as influencing the results.

Question 12

Still in the context of probability, for the last two questions on the PreInterview I did change the situation away from dice and onto spinners, since the subjects had at least been introduced to questions involving spinners in the PreSurvey. The spinner questions (Q12 and Q13) are listed together in Table 64 on the next page.

PreInterview Questions #12 & 13		
Question Number & Nickname	Brief Description	Context
Q12 "Surprise Spinner"	A spinner is broken into thirds: Two white and one black region. Would you be surprised if you got more black than white in X spins?	Probability
Q13a "Compare Groups"	For the spinner in Q12, Group A was told to do 20 sets of 60 spins, as was Group B. Compare the two groups' supposed results.	Probability & Graphs
Q13b "Expected Graph"	For the two graphs from Group A and B in Q13a, does one graph or the other look more like what you would have expected?	Probability & Graphs

Table 64

Questions 12 and 13 involved only one spinner, fairly divided into three regions of which two were white and one was black. In Q12a they were asked if they'd be surprised if in 3 spins there was more black than white, and in Q12b they were asked the same question for 12 spins. Then in Q12c, a similar question asked if in 60 spins they would be surprised if they got white 30 times. I had wanted to end up with 60 spins because that set of sixty constituted one trial for the next and final question on the PreInterview. The reason I went with the language of surprise is because that is what Truran (1994) had used in his research with schoolchildren, and I thought it might give an opportunity to elicit some different kinds of commentary on the aspects of expecting and interpreting variation.

Expecting Variation: As far as *what* was expected, the nature of this question led the subjects to express expectations largely in terms of what was surprising or not surprising to them, and from this perspective one can interpret what is therefore expected or not expected. All six cases mentioned they would be surprised if, on 3 spins, there were more black than white. This finding alone surprised me, because if the subjects are interpreting the

question correctly, that means that even an outcome of two blacks and a white is unexpected or surprising to them, as this comment shows:

EM So, if it ended up landing all 3 times, or a majority of the times you spin on black, that would be unusual for me. I mean... I would be surprised.

Moving from 3 to 12 spins, all except SP said they would be even more surprised if they saw more black than white, and SP was equally surprised. As mentioned earlier, I was aiming for 60 spins in Q12c because of the relevance of that set of spins for the next question (Q13), but I also found that I could get a sense for what the student would expect (or consider surprising) in 60 spins. The ratio from the spinner – 2 White : 1 Black – suggests that a result of 40 white and 20 black would be reasonable. I started Q12c by asking if 30 white would be surprising and then probing further to see where the subjects' threshold of surprise lay. For DS, EM, and RL, the threshold of surprise was at 30 white or below. By this I mean that they weren't too surprised if 30 white resulted, and in fact DS mention "I would expect, like 25 or 30 to be white", which led me to wonder if she was just seeing the 2-color spinner as a 50:50 situation or indeed as a 2:1 situation. She later said that for 25 or 30 white she'd "be a little surprised" and that 20 white is where she would in fact "be surprised." EM's comments helped me think in terms of a threshold, as she said:

EM I think 30 is probably my threshold for anything...Right, and then if it landed on white less than 30 times, then I'd be very surprised. Like, it can happen like half the time, ok, but um, two-thirds of the time, it's not okay for me.

On the other hand, the threshold of surprise for GP, JM, and SP was always above 30 white, meaning for example that even 35 white was fairly surprising for those three cases. SP in particular had rather high expectations for the number of whites in 60 spins:

- DC Oh yeah. What would you expect, perhaps? Out of 60? How many whites would you expect?
 SP Hmm. Well yeah, I like... Yeah, between 40 and 50 I guess, whites.

Turning now to reasons *why*, a different focus emerged from this question in that several students mentioned the respective white and black areas on the spinner and their thinking was not explicitly linked to proportional reasoning. In other words, just as there were references in the sampling context to “more reds” in the jar in contrast to the ratio of 60 reds to 40 yellows, so too were there comments about how much white or black was showing:

- EM Because, there's less area that's covered that's black, so, I would think that you're more likely to get white
 GP Um, 'cause you're more likely to go in the white, because there's more space.
 RL White is more represented in the range of possibilities, in the same way that the red candies are more represented. And so, you expect to see what is more represented seen in results more often. White's covering a lot of real estate.

In the comments above, there are also themes of possibilities and likelihoods, as have been used before in the PreInterview and also in the remarks that follow:

- GP It's less likely to get on the black area. Very unlikely that it would be even. Um, so it'd be more likely to have more, uh, higher for the white, but... It's definitely possible.
- SP Because... Each time you spin it, you have a greater chance getting white.

There were also comments showing the explicit usage of some form of proportional reasoning made by DS, GP, JM, and SP:

- DS Because, 1 out of 3 would be black, and 2 out of 3 would be white.
- GP Well if you divide it by 3, get 20... And uh, take one of the tens out of the black, 30... Uh, wait. Let me see here... So, so 20, 20, 20 {He is writing on the paper} So 40 is more likely for the white, 20 for the black...
- JM Since two-thirds of it is white
- SP Two out of three chances of getting white.

The notion of extreme value being unlikely but possible also came through in this question.

Interpreting Variation: The only person who illustrated the dimension of causes of variation was JM, who had a rather pointed focus on the physical setup of the spinner:

- JM Um, well... I want to look at the engineering of the spinner, where do you start the spin, I mean.... Do you start it in white, you know, the velocity, or the force...
- JM {A few sentences later} I mean, it CAN matter of course, yeah. Well, of course, it WOULD matter, I mean, you play like a game that has a spinner, and, if you're a kid, you know if you hit it just the right way, and you start it at just the right the spot, you could... there's a chance of it being in one spot are greater than in another spot.
- DC So this is very well-oiled spinner...Very, very fair spinner
- JM Ok, so this is a GOOD spinner. Yeah. Ok. A fair spinner. Um, yeah. And the spinner is, is flat? A flat plane? It's a fairly spun game?

I should mention that the tone of the above exchange was by no means facetious. JM had, I think, a good sense for what kids would pick up on, which is that spinners aren't seen as random generators in the minds of many.

The responses that fit the dimension of *influencing expectations and variation* had to do with the number of spins performed, and how that affected the results. For SP, DS, JM, and EM, one form of the language used was in terms of the chance for white, with the idea that if more spins were done, then there would be more chances to land on white:

- DS Because there's a larger chance that I'm going to hit the white.
- EM Yeah. I would think it would be even MORE likely to land on white, 12 times spinning, than 3 times spinning, so...
- JM The more you spin it, I would think the chances of it being more white than black are greater. I would think it be more times in the white in than in the black, at 60 times spun.
- SP So, no matter how many times you spin it, there's always a slightly greater chance of getting white

A more sophisticated version of this idea is that as more spins are done, the closer the experimental results should be to a theoretical prediction:

- JM It seems that , over the long run, two- thirds are going to come out white, over black. Now, you might not do it in 3 spins, you might not even do it in 12 spins, but if you do it in 10,000 spins, I think you're going to be closer to 66, 660 whites, you know, or whatever...
- RL Well, as the sample size gets bigger, you expect to see your results map to your sample, uh...

Note that while JM's comments connote the Law of Large Numbers, RL's comments bring to mind the tendency to superimpose onto probabilistic situation the notion of a sample space where none exists in the traditional

sense (Shaughnessy, 1997; Shaughnessy et. al., 1999). Still, the main idea that comes through in both JM and RL’s thinking is that number of trials has an influence on expectation and variation.

Question 13

The last question on the PreInterview involved a scenario of a trial being sixty spins of the (2 White : 1 Black) spinner, and then noting how many of the sixty spins resulted in black. Two separate groups of students in a class were said to have done 20 sets of sixty spins (that is, 20 trials), and the graphs of the supposed results from Group A and Group B are shown (see Figure 42).

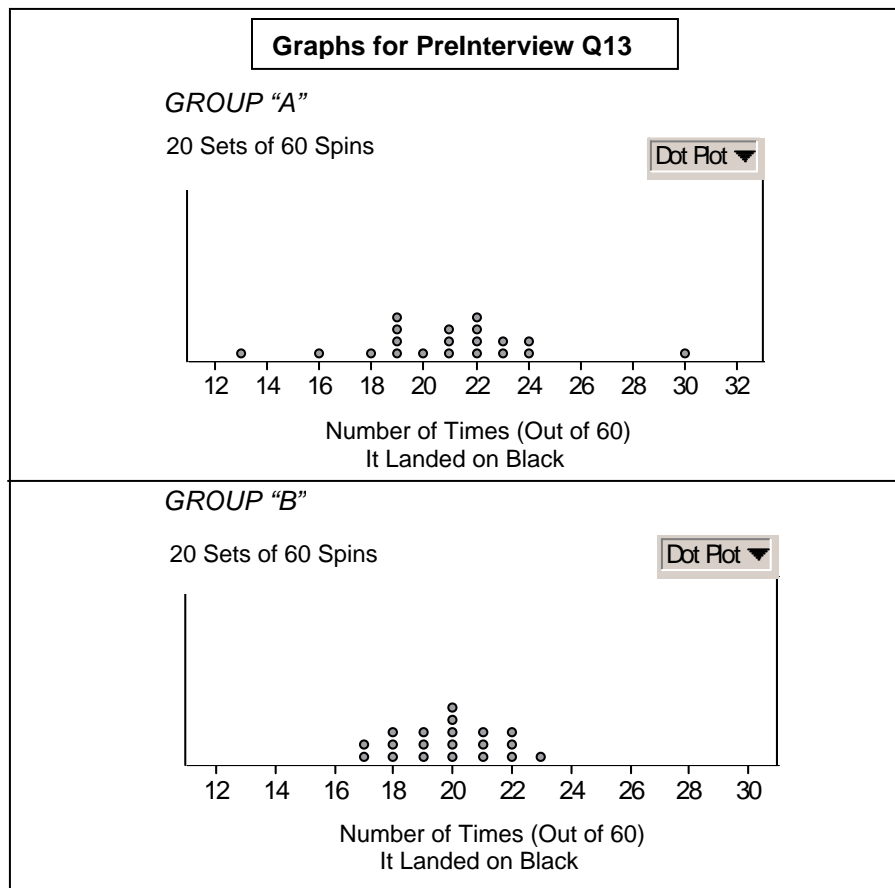


Figure 42

The question asks for similarities and differences that the subjects notice in the two graphs, and also if one graph or the other looks more like what they (my six cases) would have expected. Although I did not explicitly include the language of real versus made-up data, my sense is that this possibility was already in the minds of the six cases. I should point out that, in fact, I did generate the results for Group A by running a simulation using the Fathom software, and the graph looked appropriately varied, bimodal at 19 and 22 black, with the twenty trials ranging from 13 to 30 black (out of 60 spins for each trial). Meanwhile, I did make up the data for Group B, and the graph was symmetric around the mean (and mode, as well as the expected value) of 20 black, with the twenty trials ranging from 17 to 23 black (out of 60 spins for each trial).

Expecting Variation: For *what* the subjects expected, again there was the idea that results should be close to, or on both sides of the expected value, as illustrated by JM and EM:

JM Yeah, I would expect ... about 20
EM 20 times you'd land on black, or less, or a little more.

More information was gleaned as the subjects compared graphs in determining whether one graph or the other looked more like what they would have expected. For example, the low frequency for 20 blacks on Graph A was unexpected for JM, and having most of the data be between 18 and 24 (inclusive) for both graphs was seen as a reasonable expectation by all the

six cases. Also, having some extreme values was seen as expected by all except GP and RL.

Regarding *why* they held their expectations, the language of possibilities and likelihoods was used to describe what could happen:

- SC And so, it would be pretty rare to get only ... to get 30 blacks, in one set. But it COULD happen. But it'd be rare.
- JM Well, I like that, you know, the possibility of it being black 50% of the time is certainly there...you know, <Extends hand out> the chances of it becoming less black and more black are there, in 60 spins.
- RL Not only did someone get the very unlikely result of 13, somebody else got the very unlikely result of 30.

Notice the language of “rare” and “chances” and “unlikely” in discussion possible results. Other explanations relied on proportional reasoning:

- DS Because, your odds are, that 1 out of 3 is going to be black, in a perfect world. And so you're going to have chances that, you're going to have more than the 1-out-of-3 black, and much fewer than the 1-out-of-3 black
- EM Yeah, that would be exactly, really close to one-third of the time they landed on black. Which is what I would guess would normally happen.
- SP Because there's that...That 2-to-3 ratio thing, where there's two whites, and one black, out of these three possibilities

EM not only uses proportional reasoning, but affirms what she said earlier about expecting results that are “usually” or “close to” the expected value of 20 blacks in 60 spins.

Displaying Variation: *Evaluating and comparing graphs* was the main dimension that this question explored. There were many themes of comparing graphs that came out of the responses. One theme was the focus on the

average of each graph. JM and DS were the most explicit about what they were noticing with regard to centers:

- JM I would like to see more 20s in A,
 DS And so, most of those, in Group B, fell in that one-out-of-three {She points to the mode of Group B, which is 20 black}

Another theme was the focus on the spread of data or variation within the graphs, and this included attention to the range and extreme values. For example, several of the cases commented on how Group A had more spread or range:

- JM Um, well, Group A certainly has wider variables {Holds hand far apart} It's gone between 13 and 30.
 RL there's a wider range, more outliers in Group A.
 SP Well, this one has greater variation {Group A}...It's more spread out. It goes from the, lowest is 13, and it goes up to 30.

GP explicitly linked the greater spread to an increase in "randomness", as this exchange shows:

- GP Uhhh, Group A is... is, uh... is more spread out.
 DC Ok. What shows you that?
 GP Um, you have some out here, like 13, 30 {Shows by pointing}. More randomness to it, I guess. The middle is more random too {Group A}

It seemed likely to me at this point that GP was attending to both the range, (since he pointed out the extreme values), and also to the shape (since he commented on the "middle" of Group A). Since Group B was symmetric but Group A was not, it may be that the lack of symmetry was influential in GP classification of the middle as "random." Also, of the four cases just quoted above (JM, RL, SP, and GP) regarding Group A's wider range, all of them

compared that range to Group B with specific comments about how Group was narrower.

By adding more language that focused on the distribution of the data, some responses showed attention to the shape, or where the data was more concentrated:

- DS {Group B} 'Cause it's that famous curve. {Laughs} The majority the spins fell between 18 and 22...
- GP This {Group B} is more bunched up
- RL Group B is tighter, and definitely holds to a center more
- SP This one {Group B} is clustered within 17 and 23
- JM Group B of course, is a much tighter {Brings hands close together} distribution of black.

Some of the language, as in GP's "bunched", SP's "clustered", or JM and RL's "tighter", are all more suggestive of relative distribution of the data than mere attention to the range. Also, the nature of distributional reasoning comes through particularly well in RL's comment, which involves both elements of center and spread.

Making conclusions about graphs was the other dimension addressed by responses to Q13. RL was confident that Group B was what he'd expect, but JM was less certain. In his initial appraisal of Group B, JM had said "I would expect Group B to be, closer to...{Hands in an inverted V-shape}." Later, after comparing the two graphs, JM made his comments (as shared above) about Group B having, in fact, "a much tighter distribution." In the end, JM's conclusion was that "I would like to meld both of the graphs." DS went through three distinct phases along her path to a conclusion:

- DC Does one graph or the other look more like what you would have

- expected ?
- DS So. Well, I think Group B looks more like what I would expect.
- DS {Later} Okay, you know... No, I think either graph would be {okay}
- DS {Later} Well, now I'm thinking that, you know, that it's not going to always end up in this perfect graph picture. So this {Pointing to Group B} would be, if you were going to fake a graph? This would be a fake graph. I would think, if somebody made it up, it would probably be "B". Back up. I think Group A looks more real.

At first, Group B seemed attractive to DS, I think largely because of the relatively large frequency at 20 black compared with Group A, since DS had earlier said: "In 20 sets, your number of blacks stays pretty close to your {20 Black and 20 White} ." The language of "perfect" leads me to think that it was the symmetry in particular that led DS away from Group B, although she was not explicit as to her reasons on this change of mind. GP also had a comment about Group B looking too "artificial," and he – along with DS, EM, and SP – concluded that Group A was more in line with expectations.

Interpreting Variation: The only dimension of this aspect to receive any responses was that of the *effects* of variation. SP again commented that real life and probability do not always coincide, as she reflected on the reasons for her preference for Group A:

- SP Because there's greater variation {Laughs}. And that's just... I feel that's life. That's just how things work out when they're randomly done.

SP's remarks indicated, I thought, a good appreciation for variation. Of course, the extension of this thought is once again that results from a single trial "could be any number," as shared by EM, GP and SP, and thus you can never know or control the results.