The Law, such as it is
Season 3, Episode 7

Larry

This is Larry Lessig. This is the seventh episode of season three of the podcast, "The Law Such As It Is." In this season, we are reviewing the decision to revoke the tenure of Francesca Gino at the Harvard Business School.

That decision was based on four claimed acts of academic misconduct, basically data fraud. And so far, we have addressed two of those four charges. In this episode, we're going to address allegation number three, which will then leave just one final allegation to review, allegation number one.

If you've been listening to this podcast, the critique of allegation number three is going to repeat some of the same themes that we've covered in allegation two and allegation number four.

First, as with allegations 2 and 4, the claim of the Hearing Committee with this allegation is that the anomalies identified in the data all tended to support the conclusions of the paper. That claim by the Hearing Committee in this case is once again and overwhelmingly false.

Second, as in the earlier allegations, the Hearing Committee concluded that Francesca must have been responsible for the changes identified. Once again, despite it being uncontested, the data cleaning, coding and processing was work done by research assistants — and in this case, research assistants working with the data for more than 500 days. Never once were those research assistants questioned about their work with the data.

That's not all that's wrong with this allegation, but it is a start. Let's get to the particulars.

Once again, I'm joined in this conversation by an anonymized version of a friend who has studied this allegation carefully, as I described in the last episode. My friend's voice has

been modified by fancy AI, so this is not what my friend sounds like in real life, although I gotta say the AI is not too bad either. So welcome Ava.

Ava

Thank you, Professor.

Larry

Why don't you introduce the allegation?

Ava

Sure. Allegation three involved a paper titled "*Evil Genius: How Dishonesty Can Lead to Greater Creativity*." The paper was published in *Psychological Science* in 2014, and the aim of the paper was to examine whether there was a correlation between acting dishonestly and creativity.

In one study in that paper — study number 4 — participants played a game where they were to guess the outcome of a virtual coin flip. After they guessed the outcome, they flipped the virtual coin, and then they were to report whether they had guessed it correctly or not based on the first flip. They were asked to flip the virtual coin a couple more times to ensure the toss was legitimate. The software knew what the outcome of the virtual flip was, and so the researchers could tell whether the participant reported it correctly or not, meaning the researchers could tell whether the participant lied or not. The participants then completed some creativity tasks. The hypothesis was that those who cheated would be more creative.

Larry

And is that what they found?

Ava

Yes, that's what the study found. Participants who cheated performed higher on measures of creativity.

Larry

So when was the data for this study collected?

Ava

Though the paper was published in 2014, the study in question, study 4, was conducted in late 2012. So we're back to some of the issues you covered in prior episodes when you talked about the reasons behind the six-year statute of limitations on reporting research misconduct: it is really hard to reconstruct exactly what happened almost 14 years after the data was collected. And as we'll see, that point applies to this allegation especially.

Larry

Right, I've said plenty about the limitations point: The University promises that academic charges of academic misconduct six years or older "will not be investigated." But literally millions have been spent investigating these charges, some by Francesca and multiples of that by the University. But we're going to pass over that flaw for now.

Let's start with the context of when this paper was written, because this is a paper where there's particularly strong evidence that there were certain errors made before publication. Is that right?

Ava

Yes. They are what we've been talking about throughout this podcast: the anomalies. It's clear there are anomalies in this paper. The question is what the source of those anomalies is.

Larry

You mean the question is whether Francesca herself introduced those anomalies, or whether they were introduced by the research assistants. Correct?

Ava

Correct.

Larry

And why would you think that the anomalies would have been introduced by research assistants?

Ava

Well, remember that one constant theme in this podcast has been the Hearing Committee's constant mistake in claiming that the anomalies they've identified all tended to strengthen the conclusions of the paper. That's not true. It's not true in the other allegations we've considered. It's not true in this case either.

If you have anomalies that don't strengthen the conclusion of the paper, that would lead you to think that these are not anomalies that Francesca herself would have intentionally introduced into the paper. You typically don't commit academic fraud by intentionally reducing the significance of the results you're discussing in the paper. Instead, anomalies that weaken the conclusions of the paper would more likely have been made by research assistants.

Larry

What did the research assistants say about this allegation?

Ava

Nothing.

Larry

What do you mean, nothing?

Ava

Well, the research assistants said nothing because the research assistants were not interviewed with this allegation.

Larry

The research assistants were not interviewed?

Ava

No. The research for this paper continued over two years, there were at least five separate research assistants who

worked on the research, but Harvard did not interview any of them.

Larry

So let me make sure I understand this correctly. We know that the anomalies could have been introduced either by Francesca intentionally or by research assistants accidentally. We also know that the anomalies in this case are not all supportive of the conclusions of the paper, so some of the anomalies here are most likely the result of the research assistants. To understand whether and how the research assistants could have introduced these anomalies, a proper investigation, a complete investigation, would have had to consider the work of the research assistants, but Harvard did not do that. Instead, Harvard raced to the conclusion of guilt at the business school stage and then tripled down with 100,000 lawyers to affirm that conclusion throughout this multi-year process.

Ava

100,000 lawyers sounds like a lot, but yes, it's a good summary and an important summary, because the point should be obvious: no good prosecutor's office would think it had completed its investigation when it hadn't even spoken to people who could have easily created the problem being investigated.

Larry

I don't think anybody has ever accused Harvard of being a good prosecutor's office.

Ava

Fair enough. But I want to make an even stronger point: where the evidence exists in this allegation, we can say with certainty that it was the research assistant who introduced the anomaly.

Larry

And how can you say that?

Ava

Let's divide the research into two parts — an objective part and a subjective part.

The objective part of this study, study 4, related to a virtual coin flip. Participants were asked to predict the results of the coin flip. They indicated their prediction. The coin was flipped. They were asked whether they had guessed it correctly or not. Most people told the truth. Some people lied — some said they had guessed it correctly when they hadn't. The aim of the study was to see whether those who lied were more creative than those who didn't.

The subjective part that's relevant here can itself be further divided into two parts, each involving a creativity task participants completed and that research assistants then needed to code.

The first task involved a remote association test with a tight time constraint of five minutes. Participants were given three words and asked to provide a word that was suggested by the three words.

Larry

So how would that work?

Ava

Well, for example, the participants would be given "blank-white-lines" and then asked to enter a word that connected all three — in this case, "paper" is a logical choice.

Larry

And so how many questions like that were there?

Ava

There were 17.

Larry

So, a participant would give a word association for 17 triads of words.

Ava

That's right.

Larry

And then how would they be evaluated?

Ava

Well, first, the RA would have to clean up the data. Sometimes the participant would misspell the word p - a - p - a - r rather than the right spelling of p - a - p - e - r, for example. Or sometimes they'd put a question mark at the end of an answer they were unsure of. Regardless, the RA would go through the Excel sheet and clean up the data to make it consistent and meaningful.

Larry

And when was this cleaning done?

Ava

We don't have the original Qualtrics data, but we do have a file that was created about 11 hours after the study was completed. That file was saved on November 18, 2012; the file used for analysis before the paper was submitted for the first time is from July 24, 2013 — almost 250 days later.

Larry

Okay, but you said this was the first bit of the subjective part of the research. Was there a second bit, or a second part to this subjective part of the research?

Ava

There was. The first subjective part we'll call the RAT task, for "remote association test." The second, we can call it a "Usage Task."

In addition to being asked to make a word association with three given words, the participants were also asked to name as many uses of a newspaper as they could in one minute. Like the RAT, this is a common creativity task, often used in the psychological literature.

Larry

The uses of a newspaper?

Ava

Yeah. Like, what could you do with a newspaper? I mean, the obvious use would be to read it, but you could make a paper airplane with it, or use it to start a fire, or frame it, or wrap a fish in it. I guess, lots of things you could do with a newspaper.

Larry

And so why were the participants asked to supply possible uses of a newspaper?

Ava

The goal was then to grade the creativity of their replies. Not only the number of distinct uses people came up with, but also how different they were from one another, and how original they were. If you said you would read the newspaper, that's a solid, boring, uncreative use of the paper. If you said you could use it to train puppies or make a paper hat, those would be very creative uses of the paper. The objective was to distinguish between people who could identify many creative uses of a newspaper and those who could just identify a few.

Larry

Right, because the question of the study was whether liars were more creative.

Ava

That's right. I mean, we know they're creative with the truth. The question was whether that creativity carried over beyond the truth.

Larry

Okay, but this grading of these answers feels a little bit subjective, doesn't it?

Ava

It's certainly subjective, and that created an issue in the preparation of the paper. Initially, when the authors submitted the paper to the journal that eventually published it, they received what's called a "revise and resubmit" response. That means the journal liked it, but they wanted it revised, and the journal described what that revision should be.

As it related to these creativity measures, the journal suggested that they evaluate the creativity along two additional dimensions of creativity. So rather than just one measure of creativity, there would be three. That's where the subjectivity issue arose: If the responses for the first dimension were evaluated by one person, but then the responses the next two dimensions were evaluated by a different person, the difference between those evaluations could produce bias in the results. To avoid that bias, after this revise and resubmit request, the authors asked for one RA to evaluate all three dimensions. That RA thus produced a new evaluation along all three dimensions.

Larry

Okay, but then what was the committee's allegation of fraud here?

Ava

Here's where the difference is between these two text-based responses — the RAT task and the Usage task — are so illustrative. Because we actually have the original coding data for the Usage Task. And so we can actually see where the error for this task arose.

If we compare how the research assistants coded the data — after the revise and resubmit, the three dimensions of data — to the final data set that was submitted, we can see a number of places where the final data set was produced by copying the wrong part of the original data set.

Larry

What do you mean by that?

Ava

I mean that the person who imported or copied the data into the final dataset — and there were two, because this part of the data was imported at two different points of time, one initially, and one after the revise and resubmit request — that person copied data from the wrong part of the source dataset into the final data file. With the first error — what the hearing committee identifies as 7 values that had been changed — we can show the copying is from the right tab of the data file, but the wrong rows. With the second — what the hearing committee identifies as 18 other errors — we can show that the error was in the copying of the data from the wrong tab of the source data sheet.

Larry

Right, I read the that the committee said the data was copied from a different study into this study, and that they used that to evidence the fraud.

Ava

With respect to the 18 participants, yes, it was copied from the wrong tab or wrong rows.

Larry

And so you're saying that was error. And again, how do you know that it's just error?

Ava

Because the consequence of that copying was to weaken the study, not to strengthen it.

Larry

It weakened it. The anomaly weakened the study, right?

Ava

It weakened it. So the idea that this copying was made as an act of intentional falsification of fraud is just wildly implausible. There was no motive to falsify data to weaken findings.

Larry

Okay, so that's with respect to the 18 participants in the usage test. What about the other seven here?

Ava

The difference did make the result slightly stronger. But even here, if you reverse the mistake– if you put the original text back with the correct participants — the result is still highly significant. The change thus did strengthen the results, but they strengthened an already significant result. Which means there could have been motive, but not much: there was no need to strengthen the paper at that point, and in that way, because the paper was already strong enough.

Larry

Okay, so with the usage tasks, your claim is that you can point to exactly where in the source file the data was copied from, and because of that, we can make an inference about whether the copying was fraud or not?

Ava

Exactly. With respect to the largest error, the one affecting 18 participants, the copying looks like an error. We can find the place where the data was copied from. The copying weakens the strength of the creativity measure and the overall strength of the study. Why would you falsify data to weaken the conclusion of a paper? You wouldn't: which is why I'm saying this is likely an error.

With respect to the seven, we can see the values were taken from the wrong row of the correct study. Here, there's an increase in the statistical significance of the study, p equals .03, down to p less than .01. So that suggests a possible motive.

But arguing against that possibility is a simple fact: the finding gets would be strong either way. There is a difference, but not a difference that either reviewers or readers would ever worry about or even think about.

Thus, for the first change, there was no motive for the change. And for the second, there was, at most, an awfully weak motive. And from my perspective, the presence of the first undermines any negative inference about the second.

Larry

What do you mean by that?

Ava

Well, again, if the anomaly weakens the finding, that suggests errors. Not all errors are going to be in one direction, so it's not surprising that the second error was in a different direction, which means the first error weakens the suggestion of fraud in the second.

Larry

With respect to this set of anomalies.

Ava

Right. We're just talking about one of the two sets of subjective evaluations. One of those two is most likely error, not fraud.

Larry

And what about the other subjective evaluation?

Ava

Well, here, unfortunately, we don't have the same source data.

Larry

What do you mean?

Ava

I mean, with the other subjective task, we had the coding data from the RAs who coded it. Here, we don't. So here, we can't be as sure about what exactly explains this error.

Larry

Okay, well, so first describe this error.

Ava

In this part of the charge, there are four participants out of the 178 whose scores basically don't add up.

Larry

What do you mean by the scores don't add up?

Ava

Remember, this is the RAT task, the task where participants were given three words and were to answer what those three words were associated with. The example I gave before was "blank-white-lines"; and for those three words, the clearest answer was "paper." For these four participants— out of 178 — if you look at the individual coding from Excel for each of the 17 answers to see whether the participants got the right answer, they don't add up to the aggregate or the total represented for all the answers together.

Larry

Why would that be?

Ava

We can't figure this part out. Part of the problem was caused by a mistake in the coding of the original file. Remember the original file — the file produced 11 hours after the study was over — is the closest we have to the raw data. But that file already has coding in it.

Larry

It already has coding in it?

Ava

Yes, it isn't just the raw data from Qualtrics. It included coding — meaning formulas designed to convert the answers into usable data. That means that file had already been worked on by RAs. And that means, as with allegation 4, we don't actually have the original data, but are making inferences from a file that had already been worked on.

Larry

Okay, so what were the mistakes in that coding?

Ava

Well, we can say that the formula that tested the response was wrong in at least two cases. For example, one answer was supposed to be "thief," but the coding formula spelled thief incorrectly. E, I, rather than I, E. Another answer was supposed to be "beer," but the coding formula spelled "beer" as "bear," B, E, A, R. But even after you correct these formulas, you don't fix the problem. With respect to these four participants, the totals for these four still don't add up, and because we don't have the RAs' coding sheets, we can't say why these four don't add up correctly.

Larry

Did the changes strengthen the paper?

Ava

Yes, they did. These changes did marginally strengthen the paper. P was .012, versus p equals .04. But here again, if you took these four participants out of the study, both with these subjective questions as well as the uses tasks, it wouldn't have changed the finding of the paper. The paper was statistically strong without them, and both could legitimately be reported as p less than .05, so again, the referee and the readers would not know. This again raises the question: why commit fraud when it doesn't change the publishability of the ultimate paper?

Okay, but here again, they didn't talk to the research assistants. Conceivably, the research assistants would have had the coding sheets that would have made it possible to see exactly how the error was made, if indeed it was an error. If they talked to the research assistants, there was a chance those research assistants would still have had the coding data if they had the coding data. Then, as with the first subjective scoring, you could have figured out why. In this context, there was also an error. Harvard didn't interview the research assistants and didn't think it was important to have a complete investigation. So we don't have that information here.

Larry

Okay, so summing up the story about these subjective tests: With respect to the first, because we have the coding data, we can be pretty confident that this was just an error by the research assistants. With respect to the second, we don't have the coding data, so we can't absolutely negate the idea that this was intentionally, fraudulently tweaked. But even here, if you remove the observations for the four scores that don't add up, the paper would have been statistically strong even without them. So there's a motive, but not a clear motive.

Ava

That's right.

Larry

Okay, so let's turn to the objective test. Explain again the issue around this virtual coin flip.

Ava

In this part of the study, participants were asked to guess what the results of a coin flip would be. They then flipped the coin and could see the result. They were asked next whether they had predicted correctly how the coin flip would come out. The majority of them told the truth, and about 24% of them lied.

Larry

And I take it the study knows this because the program recorded what they had guessed, what the result of the coin flip was, and whether they said they had guessed correctly.

Ava

Maybe. You see, we just don't know exactly what the program did, because we don't have the program, and Harvard never asked the programmer who wrote it to provide the program.

But the more I've studied this allegation, the more I've realized that both sides just misunderstood what was happening here. Both sides were really talking past each other, and we can simplify the story pretty dramatically if we assume that what happened is what Harvard said happened. Again, we don't know that's what happened. We don't have the data because we don't have the program. But the point is, even if we accept Harvard's description of what happened, we can show this is not fraud.

Larry

Okay, so what did Harvard say happened here?

Ava

Harvard said the coin flip program was rigged.

Larry

How could it have been rigged? What does that mean?

Ava

It'd be rigged to basically give the opposite of whatever you guessed. If you guessed heads, the program would simulate a flip and show you tails. You guessed tails, the program would simulate a flip and show you heads. That way everybody in the survey thought they had gotten it wrong. The question then was whether they would tell the truth about whether they got it wrong and how their truth telling would affect their creativity.

Larry

Okay, so then what was the problem here?

Ava

Well, there are 178 participants in this data set. On Harvard's assumption about how the study was conducted, all of them were told they got the coin flip wrong. When they were asked whether they had gotten it right or wrong, the majority were honest and reported that they hadn't guessed it correctly. But about 24% lied. About 24% said they had guessed it correctly, when in fact they hadn't, because, again, on Harvard's assumption, the coin flip was rigged, so everyone was told they had gotten it wrong.

Larry

Okay, and so what's the problem with that? Where was the source of the alleged manipulation of data here?

Ava

Well, there were 12 participants who, like the majority, had said they had not said that they had guessed it correctly. They had been honest about the fact that they hadn't guessed it correctly. Nonetheless, these 12 were marked as cheaters. Or more specifically, between the first file that we have available and the final file we have available. These 12 went from being marked as non-cheaters to being marked as cheaters. Harvard's claim is that the change was itself cheating, allegedly by Francesca.

Larry

What do you mean marked as cheaters?

Ava

Well, remember the file we have here is a file that was produced about 11 hours after the study was completed. And as I said, it wasn't the raw data from the Qualtrics survey, because the file already has coding in it.

Larry

Remind me again, what do you mean by coding in it?

Ava

Well, already there are formulas that the RA had put into the file to begin to process the data, to make it usable to conduct the study. Formulas that would translate the raw responses the participant had given into usable data that could be analyzed.

Larry

Okay, so what data was there related to cheating?

Ava

In this file, there was one variable or column in the data set called "reported guessed correctly" that reported whether the participants said they had guessed correctly or not. If they didn't say they had guessed correctly, then the value of that variable was zero. If they said they had guessed correctly, the value of that variable was one. The people who said they guessed correctly were lying. So those people, at least we know, were cheaters. These 12 participants, like the majority, had honestly reported that they had not guessed correctly. So these 12 had a zero in their "reported guessed correctly" column in the first version of the file.

Larry

So again, no one had guessed correctly, because on Harvard's assumption about how this data was created, the survey was rigged. If you had a zero in the "reported guess correctly" column, you're telling the truth, and if you had a one, you're telling a lie. And presumably, the study treated those who had a one in the "reported guess correctly" column as liars or cheaters. Is that right?

Ava

Almost. It turns out, right next to the "reported guessed correctly" column, there was a second column. That column was marked "cheated." In the first version of the file

produced 11 hours after the study was complete, the values in that column were the product of a simple formula.

Larry

A formula?

Ava

Yes, a formula. A formula to calculate whether that person would be marked as cheating.

Larry

And so what was that formula?

Ava

The formula looked at the value in the previous column, the "reported guessed correctly" column, and, if the value was zero, meaning they didn't say they had guessed correctly, then the value in the cheating column was zero, meaning they hadn't cheated. If they said they had guessed correctly, then the value in the cheating column was one.

Larry

So you're saying the values in the "cheated" column were exactly as the same as the values in the "reported guessed correctly" column?

Ava

Yes, there was a formula that just carried over the values from one column to the other.

Larry

Okay, so that's weird. Why would there be two columns reporting essentially the same data?

Ava

Well, this is the point that no one seemed to ever try to work out. Why *indeed* would there be two columns reporting the very same data? But as I've studied this more, it seems pretty clear why there were two columns. That's because

whether someone was a cheater did not depend solely on whether they said they had guessed correctly.

Larry

Well, what else could it have depended on?

Ava

Well, certainly it depended primarily on whether they had lied on having guessed the coin flip correctly. But the fact that there was a second column shows pretty clearly that there could be some other reasons why someone could be marked as a cheater.

For example, the survey itself had rules. If you could determine that the participant was violating those rules while completing the survey, then yes, that participant was a cheater, not necessarily because they had lied about the coin flip, maybe they told the truth there, but because they had violated the rules of the survey.

And the point is that if you wanted a way to mark more people as cheaters than simply those who reported that they had guessed correctly when they hadn't guessed correctly, you would need two columns to do that. The first reporting the lying about the coin flip, but the second, giving the RA freedom to mark more than the coin flip liars as cheaters.

Larry

Okay, so is there any indication that the 12 who had been marked as cheaters were cheaters for some reason other than that they lied about the coin flip?

Ava

There sure is. There are indications in the data to suggest that these 12 might have been cheaters, even though they reported truthfully that they did not guess correctly.

Larry

How else could they have been properly marked as a cheater, if, in fact, they reported honestly that they did not guess the coin flipped correctly?

Ava

Remember the people answering these questions are being recruited from something Amazon runs called Mechanical Turk.

Larry

What's Mechanical Turk?

Ava

Mechanical Turk is a site, really a complicated online application where you can earn money by participating in studies that people are running, psychology researchers, among many others. You sign up, you get tasks to do. You do these tasks, and you get paid based on the number of tasks you complete. Some people do it for a hobby. Some people do it for extra money. I imagine some people in some countries are living off the money.

Larry

Okay, but what does the fact that people are coming from Mechanical Turk have to do with someone's incentives to cheat or not?

Ava

If you're a Mechanical Turk worker, you want to be hired. You're not going to be hired if you're seen as a flake. In theory, you could go through a task and give the same answer to every question, yes, yes, yes, yes, but the people evaluating your answers will mark you as not a serious worker, that will make it harder for you to get hired again. Specifically for questions where there is an objective answer, you need to get that answer right, or at least close to right, otherwise, your reputation as a reliable MTurk worker would go down.

Larry

So for questions with an objective answer, how do you make sure to get the answer right?

Ava

Ask any teenager doing their homework on a computer connected to the internet, you go to Google, you search for the answer.

Larry

How are you going to search up the answer to a survey being run by a business school professor about word associations?

Ava

Well, it turns out that this particular word association task, what we've been calling the RAT task, is common in psychology. So there are lots of examples of it on the web. A Mechanical Turker eager to get the right answers could just search these answers and copy the answers from the web into the submission.

Larry

Okay, but that would suggest that these 12 did really well with respect to those objective tests. Did they do well?

Ava

Yes, they did. Their overall mean performance score (the RAT_perf variable in the data) was 11.2 compared to 7.7 for the rest of the participants. Running a statistical test, a t-test, confirms this difference is highly significant, P equals .02. And 100% of these weird 12 successfully solved specific difficult puzzles like RAT2 and RAT3, that stumped over 30% of the broader group. That's telling, and it suggests they might have used external help, like a web search. That meant they were doing something against what they were told for the RAT task, where the instruction said, and this is a quote, "Please do not use any help other than your own knowledge."

Larry

So that is saying that these 12 were particularly good with respect to these RAT tasks. But why does that show that they were cheating?

Ava

Well, they weren't particularly good with the other creative task, the one that's harder to look up on the web. On the usage tasks, these 12 were just normal. The 12 generated an average of 6.83 alternative uses, while the rest of the group generated an average of 6.61 alternative uses. A statistical t-test shows this difference is insignificant, P of 0.77.

Larry

So you're saying that an RA could notice this difference and reason that these 12 were cheaters because they were not following the instructions they were given about how they should answer the questions. Namely, not looking the answers up, because they were obviously looking the answers up when they were told not to. It was reasonable to call them cheaters.

Ava

Exactly. If the cheat variable measured more than just whether they lied about the coin flip, then an RA could look at this and believe that it identified participants who were cheating.

Larry

But wouldn't it take more time to go to the web and get answers than just to give a first guess of whatever the answer should be?

Ava

Yes, it would.

Larry

And did these workers actually take longer to complete their survey than the others did?

Ava

They sure did. These 12 took longer, on average, to complete the survey (about 21 minutes) compared to the rest of the group (about 15 minutes). If they were simply click farming bots, they'd be faster. A few minutes ago, I posited maybe they were inattentive, clicking Yes, yes, yes, yes, yes, almost randomly, in a hurry to get the work done. But if inattentive users would maybe do the work in a hurry. Well, these 12 actually took longer. And if these 12 were humans, reading a puzzle, opening a new browser tab, typing the words into a search engine, reading the results and typing it back into the survey, their completion time would naturally increase.

Larry

Wait, wasn't there a five minute time limit for the RAT? And if there were a time limit, and most subjects used all the time for this tricky task, then how could these 12 take longer to complete the task?

Ava

It turns out Francesca's so-called time limit was on the honor system. That created an additional opportunity for participants to cheat – take more than the budgeted time on the survey page. With the RAT test, Qualtrics automatically recorded the timing of the very last click on that page before the final submission and the timing of the very first click on the page. Looking at the difference between the two times, you see that the 12 participants took 436 seconds on average, that's a bit over 7 minutes, while the rest took 340 seconds on average, well under 6 minutes. A t-test confirms the difference is statistically significant, P equals .07, though only marginally different.

Larry

So the theory is that these research assistants noticed that the higher RAT scores, including the perfect scores on two of the hardest RAT questions, plus the fact that it took these 12 subjects longer, on average, to get the answers and the fact

that they were not great on the other usage tasks. Together, these indicated that they were possibly or likely cheaters. Are there other indicators that these are the sorts of people who are likely to be cheaters?

Ava

Yes, at least one. These 12 scored low on a test designed to measure caring about the rules. Questions like, "if a supervisor or person with authority gave me some direction, right now, I would follow them." So that would be consistent with them being found to be cheaters, given the other evidence that they had looked up the answers for at least some of the questions.

Larry

Okay, so you're saying adding all these things together, an RA could have looked at these different signals from the results these participants gave, and determined that these participants were cheaters, even though they did not say that they had gotten the coin flip right.

Ava

That's right. And this way of understanding it makes sense of the second column. When you ask, "Why is there a second column marked 'cheated'?," the answer must be that "cheated" was not simply a function of whether they said they flipped the coin correctly or not. Instead, that "cheated" column would have been determined based on a number of factors, maybe first by whether they lied about getting the right coin guess on the coin flip, but not just that. The fact there is a second column means the study required the RA to make a broader judgment.

Larry

Okay, but are there other counter examples? Are there other participants who fit this particular pattern, who maybe were not marked as cheaters?

Ava

Yes, I don't want to overstate the claim. I'm not claiming that this is a clear rule that the RA applied 100% consistently. All I'm saying is this: *if* the RA had been asked to make a judgment, these 12 would have been appropriate for the RA to mark as cheaters. Again, we don't know whether the RA was asked to make a judgment, because, again, Harvard did not interview any RA for the study. The critical exculpatory evidence was just never collected.

Larry

So this is a great example of why even a very simple interview would have gone a long way in resolving what actually happened here. While an RA is not going to remember every detail of working through a dataset, they might well have remembered whether they had been asked to make a judgment about whether a participant had cheated, beyond simply recording whether they had reported correctly that they guessed the coin flipped correctly. They might have remembered concluding that someone had used Google to figure out the RAT task. This is an obvious example about how failing to do the most basic investigation wildly undercut the fact finder's ability to know what exactly happened.

Ava

Right, if they had interviewed the research assistant and said to the research assistant, "What led you to mark someone as a cheater when they *hadn't* falsely claimed to get the coin flip toss right?", that research assistant could either have said, "I didn't make any changes in the cheating field column," which would tend to incriminate Francesca. But if the research assistant instead said something like "I was applying a more multi-dimensional analysis, I was looking at this other stuff too," well, that would tend to exonerate Francesca. Imagine if the research assistant gave remarks like what we described above, how RAs all know that RATs can be gamed, so if you see scores too good, plus a long time, you can infer cheating. But once again, a basic part of the investigation was

just not conducted because Harvard assumed that she was guilty. They weren't investigating whether she was guilty. So yes, they didn't even interview the RAs.

Larry

Wow. Okay, so then what about the third part of the charge made by the Hearing Committee?

Ava

Here, the committee found that Francesca had fiddled with the data to hide her earlier fraud, or what they said was her earlier fraud when one of the researchers from Data Colada asked her to provide the underlying data for this study.

Larry

And did she alter the data?

Ava

Here again, as with the rest of the charge, it's impossible to know with confidence without talking to the research assistants involved. Because Francesca is clear about one important fact: she said she did not create the data file that was shared with Data Colada.

The question for the research assistant would then have been quite simple, "did you create the data file or not?" If the research assistant said, "Yes, I worked on the data file," then it would be important to understand what the research assistant had done and why. But if the research assistant had said, "No, I didn't work on the data file, I remember Francesca said she would take care of it," that, again, would have been completely damning for Francesca. The failure to do this most basic step of investigation means we don't have an answer to this fundamental question.

Larry

Okay, but why did the Hearing Committee conclude that her manipulating this data was fraud?

Ava

The Hearing Committee flagged two changes that happened in the data. There were others. For example, column headings were clarified to make it easier to understand the data. That is the sort of work that research assistants would do when preparing to post data publicly or share it. But beyond that innocent change, there were two other changes that the committee didn't find to be innocent.

Larry

And what were the non-innocent changes?

Ava

The basic claims that the data provided to Data Colada had been modified in ways that clumsily covered up earlier modifications. Either covered up or corrected, depending on the scenario.

Larry

And did Francesca acknowledge that these changes had been made?

Ava

Again, what Francesca insisted upon was that she was not in the business of preparing data to be shared with anyone. When a request for data was made, she would have passed it to an RA, and the RA would have prepared the response, which she would then forward along. But no one denies that these anomalies exist. The only question is, who is responsible for the anomalies and why?

Larry

Okay, that could have been her ordinary practice. But in this case, if we assume that the earlier modifications were examples of fraud, we can see why she would have had an incentive to tweak these data to make sure that the fraud was not revealed. Isn't that right?

Ava

That's right. If you don't believe Francesca, then you believe that what happened is she got this request for the data, panicked, realized that she needed to cover up the earlier alleged fraud, and then tweak the data to cover up that earlier misconduct. But then you must also believe that her cover up was clumsy because she was caught because of the inconsistencies that the cover up revealed.

Larry

Okay, so then help us understand what exactly were the changes that either she or the research assistant had made.

Ava

There were two things that were changed at this stage. The first related to the 12 participants who had before been marked as cheaters, even though they had not asserted that they had guessed correctly. For these 12 participants, the data given to Data Colada modified the "reported guessed correctly" variable to set it to a number that was the same as the cheating variable.

Larry

Okay, just to be clear here, to remember what we were talking about: the "reported guessed correctly" variable is zero if they didn't say that they had guessed it correctly, and one if they said that they had guessed it correctly. That data came from Qualtrics right, that came directly from the survey. The "cheated" variable was the derived variable. Derived in part, we suggested, from the "reported guessed correctly" variable (if you reported guessed correctly, then you are marked as a cheater), and in part, as we've concluded, from other factors.

Ava

That's right.

Larry

And what you're saying is that the modification that was made to the data before it was given to Data Colada was to change the Qualtrics data to conform to the result in the "cheated" variable column. Is that right?

Ava

Right. The changes to a column that comes originally from Qualtrics, someone changed that column, the values in those rows in that column, to make those entries consistent with the "cheated" variable. If that someone was Francesca, it would look like an effort to cover up earlier data manipulation. If it was an RA, it could simply have been an RA who didn't understand what the "cheated" variable was, who was trying to understand why the field seemed to be inconsistent. That RA definitely should have flagged the problem to Francesca, but didn't. Instead, the RA just tried to make the data consistent before sharing it with Data Colada.

Larry

Is there a reason in this strategy to believe that it is or isn't a strategy that a fraudster would have deployed?

Ava

Well, if this was a cover up, it would have been a cover up by basically the dumbest means possible. If you were responding to a request from a fraud investigator in a way designed intended to hide your fraud, you could simply have deleted the "reported guess correctly" column, since it wasn't used in the analysis.

Or you could have changed the "cheated" variable back to match the reported guess correctly column, making it look like the reclassification never happened.

Those two strategies would have eliminated the discrepancy, but preserved the original data, which of course, the fraud investigator would have had access to.

Instead, what the modifier did was the opposite: they left the "cheated" column alone, the "cheated" column that had been changed, at least on this theory, and change the raw self-report data, the part that comes from Qualtrics. This just preserves the fabrication and destroys the original data, giving the data fraud investigator the evidence they need to claim something fishy is going on.

So yeah, that sounds pretty stupid. And if you conclude this was indeed fraud by Francesca, it shows Francesca is not only fraudulent but stupid in her fraud.

Larry

What about the changes related to the four creativity scores?

Ava

Remember here, the charge was that the numbers didn't add up for these four participants. The total RAT score is not the sum of each RAT1 through RAT17. Here too, the charge is that someone changed the individual numbers so that they did add up.

Larry

And if that someone was an RA, why would an RA have done that?

Ava

If it was an RA who had been asked to pull together these data, and if he or she saw these inconsistencies, then they plainly should not have done what they did. They should have flagged the inconsistency and asked Francesca about it. But if they didn't, then these changes were clearly aiming to make the data consistent. If the RA didn't know much about what I explained regarding the inconsistencies, then they were just cleaning up the data to make it consistent. It would have been an error, a pretty serious error, but it would not have been fraud by Francesca.

But if, on the other hand, it had been Francesca, then the changes were being made to hide the suggested fraud on the four aggregate scores that had been raised.

And you know, for both sets of changes, there is this overarching background question of what it means for data to be internally inconsistent, and specifically what an RA should do when data is internally inconsistent. If one part of the data says one thing and another part says something else, it's easy for me to imagine an above average RA who thinks of him or herself as diligent, thinks of him herself as not needing to take big-shot professor's time with minor data glitches, who certainly doesn't want to criticize whatever prior RA did the work and created this apparent inconsistency. This above average RA would want to find a way to make the problem go away without bugging Francesca. With data inconsistency, the obvious temptation is, yes, to make the data consistent. Is it wrong? Absolutely. Is it research misconduct? It looks a lot more to me, like general sloppiness, and it looks more like what a 20-something RA might do versus what a professor is likely to do.

Larry

Okay. So for this part, the Hearing Committee quoted Francesca as saying, "If you want to make sure that everything is accurate in the data set and the sums are not there, you might change it so that it's consistent." The Hearing Committee pointed to that statement and then said, "the actions in this instance were antithetical to science. Instead of fixing the sums to reflect the actual data and contacting the journal to correct the scientific record, Professor Gino manufactured the data to support the sums." That does sound pretty bad.

Ava

Yeah, except the committee completely mischaracterized what Francesca said. The committee assumes that Francesca is describing *herself*, because it assumes that Francesca was the one who would put this data together to share it with Data Colada. But she was not describing herself. She was

describing how someone cleaning files would have worked with the files. She had already testified that she didn't clean or prepare data files; she certainly didn't clean or prepare these data files. Thus, when she said "you" in that statement, she wasn't talking about herself. She was talking about the RAs who would have been preparing the data.

Larry

Okay, so we have two radically different theories here.

1. One theory is that a research assistant is asked to clean up the data to send it off to this data fraud investigator. They do some cleanups that are perfectly reasonable, like clarifying the titles of a column, and they do other cleanups that are not reasonable. One seems pretty innocent: they didn't understand why there's an inconsistency between the "reported guessed correctly" column and the "cheated" column. So they just correct the "reported guessed correctly" column, stupidly, if you're trying to hide fraud. But that's what they would have done. More troublingly, they just changed the numbers related to the elements that added up to some aggregate score for these four variables, wrong, but wrong done by a research assistant, not by Francesca.

2. The alternative theory is that Francesca made these changes. The changes to the 12 participants would have been intended to hide earlier changes to 12 participants. But again, that change would have been done in the dumbest way you could have imagined if you were trying to hide the fraud. With respect to the four creativity scores, this would have been fudging to hide the fact that the creativity scores had been amplified.

So two radically different assumptions about what went on here. What would you need to know to be able to tell which of these two assumptions actually reflected reality?

Ava

What you really need is an interview. Francesca's records indicate that there are at least five RAs who could have been involved in the process of preparing the data to be shared with the Data Colada researcher. Harvard knew the names of these five RAs. A conversation with them about preparing this data could have determined whether they had in fact prepared the data, and in preparing the data, whether they had in fact noticed the inconsistencies that the innocent account suggests they changed. That question could have at least been asked. Said, Hey, maybe one of them would have emails about this, emails that Francesca didn't keep. If it had been asked, and the research assistant said, "I don't remember," we're in no better position than we are now. But if it had been asked, and the research assistant said, "Yes, I did that, and I probably shouldn't have," then we've established Francesca's innocence. If the RA said, "Yeah, I cleaned the data up. I didn't touch anything about that. I just changed the column headings," that would again tend to support the hypothesis that Francesca was guilty.

Larry

So once again, a basic investigation here could have provided the basic evidence to disambiguate between these two radically different hypotheses about what actually happened that the changes were the byproduct of a messy research assistant, or that the changes were intentional fraud on Francesca's part. Once again, Harvard failed to take the essential investigatory steps to allow that ambiguity to be resolved. Instead, they simply assumed that she must have made these changes, rather than an RA making these changes, and that assumption convinced them or allowed them to conclude she was guilty.

Okay, so let's bring all of this together. This allegation number three: this is a hard allegation to understand. Multiple people were involved over more than a year, key files are unavailable, and at least some of the anomalies that were

identified were indisputably created by research assistants. We don't have the data to see whether all of the anomalies were produced by research assistants. And most of all, we don't have any investigation by Harvard that looked at what the RAs said or did to evaluate whether these anomalies were Francesca's doing.

Some of these questions may be answered in the subsequent investigation that's ongoing in the ongoing litigation, but whether they are answered there or not is irrelevant to the fundamental point that I want to make here.

That point is this: the big claim Harvard makes is that it concluded Francesca was guilty after investigating her guilt. But to reach that conclusion, they did not do a fair or complete investigation. They went through the motions with a presumption in their head, the presumption that she was guilty, they failed in the most fundamental way to adduce the evidence that could distinguish between fraud and a mistake.

Okay, I'm grateful once again to my friend appearing through the voice of Ava.

Stay tuned for the analysis of the final allegation, number one, which I will publish two days after I publish this episode. I'm again sorry for the delay it's taken to complete this allegation. I personally had a book deadline that I had to clear before I could give it the attention it required.

I am really eager to get this one out and the next one out and then one final episode that will draw together what all these suggest and where I think Harvard should go from here.

Thanks again for listening. Stay tuned for the next episode. These episodes are produced by Josh Elstro of Elstro Productions. I'm grateful to him. They're not funded by Equal Citizens.Us. They are funded by me.

You can find this podcast where you find podcasts. Again, thanks for listening. Stay tuned for the next episode in 48 hours.

This is Larry Lessig.