LOIS LANE, SUPERMAN, AND IRON MAN: HOW PERSPECTIVES OF STATISTICS RELATE TO STUDENTS' IDENTITIES AND CAREER PURSUITS

Kelly Findley[1], Nicola Justice[2] & Florian Berens[3]
[1]The University of Illinois at Urbana-Champaign, United States
[2]Pacific Lutheran University, United States
[3]University of Goettingen, Germany
kfindley@illinois.edu

Incoming university students' unique experiences with and perspectives of statistics may shape their learning trajectories, recruitment, and retention in the discipline. We completed qualitative case studies with three first-year statistics majors. Each student shared their views of "who" statistics is to them, how they imagined a career in statistics, and recalled their prior experiences and motivations for studying statistics. We contrast the students and remark on how their varying experiences relate to perspectives on their chosen field. Results varied in the perceived nature and content of statistics, while experiences engaging with rich data sets were associated with more substantive perspectives of the discipline. Future research will use follow-up interviews to examine the trajectories of students' changing perspectives, participation, disciplinary identities, and interests.

The American Statistical Association's (ASA) Guidelines for Undergraduate Statistics Programs highlights several important practices for the next generation of statistics students, including: increased experience in data science and computing, working with complex data, using diverse models and analytical approaches, and communicating insights and methods to diverse audiences (ASA, 2014). As students enter undergraduate university programs with varied experiences and perspectives on the nature of statistical work (e.g., Gordon, 2004; Justice et al. 2020), we see two reasons why it is important to understand student perspectives about the discipline. First, research conducted from constructivist theories of learning suggests that students learn best when they connect new content with their own experiences and conceptual frameworks (e.g. Bransford et al., 2004). Therefore, it is important for teachers of statistics to understand what experiences and perspectives students bring into their courses so they can build on them. Second, early classroom experiences that help shape students' perspectives of the utility and creative nature of statistics may incline more students to explore careers in the discipline (National Science Foundation, 2021).

This study explores the statistical perspectives of three newly-enrolled university students intending to major in statistics. We asked these novice statistics majors to describe their perspectives of what statistics looks like and to reflect upon their prior experiences with and initial interest in statistics. This is an initial report based on the first interviews of a longitudinal study aiming to understand the first-year experiences of statistics students and how the content of their first-year courses interacts with their perspectives.

BACKGROUND

We use the term *perspective* to capture a broad swath of constructs, including conceptions a student may have constructed about a topic, as well as the student's perceptions, impressions, and loosely held associations with it. Perspectives involve what the student perceives in relation to who and where they are, and has nuances of being less-developed and more amenable to change than fully-formed conceptions. It is important to note that students' perspectives about statistics are descriptive (e.g., statistics involves calculations), and therefore not the same as the students' attitudes towards statistics, which are by nature evaluative (e.g., statistics is difficult; stats is fun).

Research on Experts' and Students' Perspectives of Statistics

Most statisticians agree that the discipline of statistics involves using data to grapple with uncertainties involved with real-world problems. Davidian and Louis (2012) offer a definition of statistics used by the ASA: "statistics is the science of learning from data, and of measuring, controlling for, and communicating uncertainty" (p. 1). As Wild and Pfannkuch (1999) document, the process of doing statistics is much more complex than knowledge of methods or the ability to follow a step-wise analytical process. It is also the process of developing productive dispositions (e.g., curiosity,

skepticism, perseverance) and demonstrating a deeper awareness of how to connect statistical principles to real data (e.g., transnumeration, acknowledging variation, reasoning with models). Statistics is dynamic, data-oriented, and innately interested in drawing meaning (De Veaux & Velleman, 2008).

Several studies have examined how students conceptualize statistics, with two common example conceptions being *statistics as a series of tests-and-procedures* and *statistics as a way of making meaning in the world* (e.g., Bond et al., 2012; Gordon, 2004; Reid & Petocz, 2002; Rolka & Bulmer, 2005). Students' conceptions of statistics are often viewed hierarchically, with expert conceptions of statistics being more dynamic and focused on meaning.

Rather than a strict hierarchical structuring, Justice et al. (2020) and Findley and Berens (2020) examine students' conceptions in the form of two spectrums. In both frameworks, the authors offer one spectrum representing objective, static views of data against views that acknowledge variability and uncertainty. A second spectrum contrasts theory-centered, procedure-driven views against real-world-centered, data-driven views.

In this paper, we look more broadly at students' perspectives of statistics. We examine how three prospective statistics majors have found their way to the discipline and identify the broad disciplinary perspectives that have guided their motivation to pursue careers in statistics. We make connections between their statistical perspectives, their participation in the discipline thus far, and what these connections reveal about their developing disciplinary identities.

METHODOLOGY

In this exploratory research study, we conducted qualitative case studies (Creswell & Poth, 2016) to understand and relate the perspectives of incoming freshman majoring in statistics. Three students responded to our solicitation to interview, which was sent to all first-year intended statistics majors at a large Midwestern U.S. university. We make no attempt to generalize these results to a larger population; we see these three students as what Creswell and Poth (2016) describe as instrumental cases, designed to help understand the phenomenon of students entering and remaining in statistics programs.

Two of the authors led semi-structured interviews together for each student. Modeling loosely after the *Draw a Scientist Test* (Chambers, 1983), we asked, "Who is statistics?" and invited students to draw a picture that personifies statistics and explain their drawings. We also asked students to identify skills or character traits a successful statistician might need, their prior experiences with coursework or projects, and their motivations for studying statistics. The semi-structured format allowed for a conversational style where we could follow up on important ideas and encourage students to elaborate. Interview transcripts and students' drawings formed the primary data sources for the study. Secondary data were the researchers' memos recorded throughout data collection and analysis.

Our analysis was shaped by the theoretical lens of *disciplinary appropriation* (Levrini et al., 2015) in the context of *legitimate peripheral participation* (Lave & Wenger, 1991). That is, we examined to what extent students see links between statistical work and their developing identities, with an eye toward how experiences may involve legitimate peripheral participation in statistics. With this perspective, the extent to which the discipline is perceived to be responsive to new ideas and allow for creativity will relate to the extent to which the students participate and believe themselves to belong as legitimate members or contributors.

The first round of analysis involved three statistics education researchers completing independent In Vivo coding. Next, the researchers met to discuss their codes—initially line-by-line and eventually with a more holistic approach, comparing themes and challenging each others' interpretations. To continue to ground results in the data, direct quotes were re-examined to determine whether interpretations and claims were supported by data.

RESULTS

All three participants depicted statisticians as different types of heroes, and we will use three to illustrate our results. (1) Lois Lane, the heroine in the superman saga, is an investigative reporter with a commitment to finding truth and sharing that truth with the world. She has no supernatural powers, but she is nonetheless a hero as she uncovers important stories. (2) Clark Kent has two contrasting identities: a friendly, neighborly business employee who aspires to a simple life; and Superman with out-of-this-world strength and knowledge. Whether he likes it or not, the world depends on Clark to use his powers. (3) Tony Stark is a brilliant, confident engineer-hero who uses his knowledge of science

and technology to create powerful tools—most notably the Iron Man suit. Tony finds satisfaction in his innovative creations and finding new shiny things to play with.

To be clear: none of the students claimed to be superheroes themselves. These three fictional personalities emphasize certain powers and missional orientations that were key to understanding each student's perspective. Lois (named after Lois Lane) depicted a statistician who served as a data journalist—uncovering truth and making it known to her audience. Clara (named after Clark Kent) depicted a statistician with a double life: he aspires to simple, quiet home life often intruded upon by work demands when others depend on his valuable skillset. Tony (named after Tony Stark) painted a picture of a modern, "sexy," versatile and popular picture of data science. His statistician is a data engineer who uses machine learning and cutting-edge algorithms to do incredible things.

Lois Lane: Statistician as a Data Journalist

Lois is drawn to the discipline of statistics through opportunities she anticipates in the job market as well as her own personal success in her high school AP statistics course. She likes that statistics is balanced between technical skills and soft skills; balanced between using math, but in a very applied and useful context; and balanced between working alone at a computer, and working with people. She sees statistics as modern and relevant. In particular, she thinks that statistics is useful for helping businesses and other fields make effective decisions.

Lois personifies statistics as carrying a phone because she is "always keeping track of stuff" and communicating. In her perspective, there is a single underlying Truth that can be found in the data, and it is the job of the analyst to find that truth and communicate it to clients. She says of data, "I think it's like a *sure fire* way to help a business grow because it's like it's *proof*. [emphases added]" When recounting her experiences in her AP class, she seems to identify a threshold for evidence in determining Truth and value: "it was cool seeing statistical significance… how much evidence there needs to be for something to actually be relevant." She sees the statistician as a data journalist committed to sorting through the irrelevant information in order to find underlying Truth in the data generating process.

In her perspective, creativity comes with the challenge of communicating results to an audience. In that sense, the data journalist is a helper with two purposes: first, to serve as a consultant for which the input is the data and the output is the answers that reveal the Truth underlying the data. Secondly, to communicate those results in a way that the client can understand, make sense of, and implement in advantageous ways (improving business practices, gaining more sales, furthering science, etc.). In both purposes, the data journalist serves as an intermediary helper or bridge between the client and Truth that will help move the business or research forward.

With regard to analytic methods, Lois envisions there being a correct analytical approach to take, and that she has much to learn. These perspectives largely stemmed from an internship experience in which she recorded data to a spreadsheet and gained awareness of the larger process of statistics. She feels she did not know enough of the "data stuff" to explore results and play with the data. She frequently positioned herself as a novice with quotes like, "I'm pretty much a beginner…" and "I didn't get to work with [the data] at all because they knew I hadn't had much experience…" The primary skills that Lois sees herself as lacking are computing skills. She also views coding skills as a primary source of appeal on the job market, and the skill that distinguishes adept statisticians from novice learners.

Overall, Lois' perspectives on statistics relate to her view that there is an objective process to doing statistics. As a future statistician, it is her job to carry out these methods with an open mind and get that information out to others. Lois also views herself as a novice who is on the periphery—or perhaps even outside—of statistical practice. Contributing to these feelings are her self-expressed lack of higher level content knowledge and coding skills. She is eager to learn more statistical methods and coding and to gain more experience with real data.

Clark (Clara) Kent: Statistician as a Data Superman

Clara recognizes that data analysts play a critical role in the business world. Much like Lois, Clara also found success in AP statistics, and chose statistics in part because it mixes different skill sets. She thinks of statisticians and data analysts as being good at math, but also getting to be creative and interactive. Clara's personification of statistics was a young, nerdy guy living in a big city who she characterized as a "workaholic." At the same time, he is trying to balance a fast-paced environment at work with the desire for a simple, leisurely lifestyle. Clara describes statistics as someone who has a

dog “against his better judgment,” and as someone who likes to stay at home on the weekend. Rather than having the most modern technology, Clara describes him as having an antique analog watch and carrying a briefcase with his hard copies of spreadsheets.

Clara thinks there may be an objective Truth underlying the claims that statisticians strive toward, but she emphasizes that it is hard to identify that Truth in statistics. Claims involving data are not only restricted by levels of confidence, but also clouded by biases and bad assumptions. She explains, “I was really shocked at how much of a gray area there is…” and, “you can pretty much manipulate data into however you want it to be…it definitely made me more skeptical of things that I see. They'll be like, oh, 99% and I'll be like, oh really?”

While Lois had some experience collecting real data, Clara largely drew on her perceptions from AP statistics, as well as impressions she gathered from a relative who works with data analysts. Clara had few specific data experiences to ruminate on her interview. This difference in previous experience with data seemed to explain a nuanced distinction in Lois’ and Clara’s perspectives: Lois views the role of a statistician as filtering out noise to identify patterns and report the facts (a transparent reporter). In contrast, Clara views data analysts as finding creative ways to say something relevant from data (a knowledgeable sage).

Clara recognizes a data analyst as using some type of skillset to do their work, but this skillset is still largely a black box for her. She did not bring up coding in her responses (except to say that her first-semester data science class involved coding). She did say that data analysts were good at math, and she frequently brought up communication. She emphasized that whatever it is they did, it was extremely valuable to companies and in high demand.

We see Clara’s views of the modern data analyst much like a dual-personality superhero. This person offers incredibly important skills at work, but whose exact work is difficult to understand. Additionally, Clara offers a picture of someone living two disconnected lives, saving the company during the week, and recharging as an ordinary person on the weekend. Clara’s disconnect between these two personalities might stem from her lack of authentic data experiences. Without more detail to contextualize statistical activities, she was not sure how to characterize the work of statistics. We see Clara still largely as an outsider who is peeking into the world of statistics based on some positive class experiences and anecdotal perspectives.

Tony Stark (Iron Man): Statistician as a Data Engineer

Like Lois and Clara, Tony also participated in AP Statistics in high school. For Tony, however, the AP Statistics experience was only the initiation point for his journey with statistics. His teacher recommended the website Kaggle as a way to continue learning. Tony took up this impulse with great interest and started various data projects. Tony frequently experienced roadblocks with algorithms he had not learned, but he found help and inspiration via internet searches and YouTube videos. Even though Tony describes himself as a beginner, he reports that he has been able to achieve high accuracy scores with his models. “That was just something that was very cool,” he comments. The topic relevancy of his projects, as well as the challenges he overcomes, motivate his pursuits.

Kaggle as the main encounter space with statistics has also shaped Tony's view of statistics, which he also referred to as “data science” seemingly interchangeably in the interview. He sees statistics as an attempt to model data in a way that allows for the most accurate predictions possible. Statistical expertise here is knowing many methods, algorithms, and codes that can be used to model the data. Doing data science mainly involves trying out the different possibilities and combining them in such a way that an optimal prediction is achieved.

In contrast to Lois and Clara, the answer to a statistical problem is not represented by a tangible truth, but rather as an optimal solution. The optimization criterion is the accuracy of the prediction of the model. The value of statistics comes more from the fact that predictions can be useful. This perspective seems a significant shift from the others by avoiding an epistemological consideration in the statistician’s role. To Tony, statistics (or perhaps more specifically, data science) is pragmatic.

He personified statistics as "the new girl on the block." She parties on weekends, wears cool T-shirts, communicates clearly, and stays active. At the same time, she is also intelligent and able to explain complicated issues in an understandable way. It is here that we see balance once again in the personification—statistics is cool and popular, but also smart and nerdy. This balance is important: to be truly useful to a business with statistics, the statistician must be able to present the meaning of her

results to other stakeholders. “communication is probably like number 1… like story telling. If you're a good storyteller, kind of the ability to tell a story, umm, using data, like why things occurred in the past. What might that mean about the future.”

Overall, we see Tony as a data engineer. The task of the data scientist is then to design a “machine” that can do incredible tasks efficiently and find an optimized solution. In his descriptions, Tony reminds us of Marvel's Tony Stark (i.e. Iron Man). He sees statistics as agile, powerful and modern. At the same time, both are engineers who build something from known basic materials that is more powerful than solutions before and capable of improving the world. With his experience working with real data and self-learning methods, Tony appears already to be on an inbound trajectory with respect to the discipline.

CONCLUSION

In one way or another, all three students depicted statistics personas with unique, relevant, and useful powers. Whether by uncovering the truth that few others could find (Lois), using mysteriously acquired knowledge to offer sage insights when the company was counting on him (Clara), or designing tools for prediction that few others have mastered (Tony), all three saw the work of a statistician as meaningful for making positive change.

All three liked that statistics uses math and saw themselves as being good at math, yet perceived statistics to be more exciting, interesting, or useful than pure, theoretical math. When asked about the content of their AP courses, all mentioned some form of statistical inference, with little emphasis on large projects with messy datasets. AP Statistics then was a critical entry point to the discipline, but with little or no experiences working with more authentic data-based questions.

One major difference between our three cases was the extent to which they have engaged with open-ended projects or grappled hands-on with messy, or difficult data sets. Tony has spent hours tinkering with difficult models and trying out different approaches to improving their performance. Lois’s internship experience collecting data to answer a business question gave her some insights into the investigative cycle outside of classic textbook problems. Clara, on the other hand, shared that she had no experience working with larger data sets on more open-ended problems (in her course, students could either choose a project, or to take the AP exam). We hypothesize that this gap in her experience may explain her under-developed perspective of what a statistician actually does in their daily work, which was evident by her focusing more attention on her statistics person’s lifestyle.

In addition, there was a striking difference in the value and role of computer coding. While Lois and Tony recognized a need for coding skills, Clara’s picture of statistics was stuck in a world without digital devices. Her statistician was somewhat proud of his ability to use his analog watch (which he thought was somewhat of a dying art), and he carried paper spreadsheets of data in his briefcase. In this sense, his powers were a sort of ancient wisdom that he had the key to uncover.

These differences may explain the varying connectedness of the identities of the statisticians presented with the interviewees themselves. Lois presented a modern, young, business-like statistician who was female (like herself). Lois seemed to identify with the statistician’s organization, consultant role, and strong quantitative skills. Tony presented his image of statistics as a clever, sexy new-girl-on-the-block with brand name clothing. She was versatile, in hot demand, and full of new ideas. Clara’s statistician was more dichotomous and disconnected: sage hero that everyone depends on by day; low-key, friendly homebody in the evenings. It was difficult to find ways that she might identify with the daytime superhero, and she seemed to identify only with his “simple life” outside of work watching movies and cooking food with neighbors. We asked Clara: *what genre of movies does he watch?* Her answer? Superhero movies!

FUTURE RESEARCH

We hope to follow up with these three students again. Second-round interviews would create an opportunity to check in on their first-year experiences and explore how their evolving disciplinary identities might relate to their ongoing engagement in legitimate participation in the discipline. We hypothesize that the identities and trajectories observed in our interviews with regard to experiences working with real data or acknowledging the role of computing will be more pronounced as students gain more experience in their major courses. Additionally, a larger-scale study involving more students could detect trends in what early experiences in statistics are most critical to helping students find their

identity in the discipline and how to attract and retain a more diverse group of students who will enter the story of becoming their own kinds of statistical heroes.

ACKNOWLEDGMENTS

We would like to thank the students for their willingness to talk about statistics with us! And we would also like to thank Liz Fry for her important contributions early in this project.

REFERENCES

American Statistical Association (2014). *Curriculum guidelines for undergraduate programs in statistical science*. https://www.amstat.org/asa/education/Curriculum-Guidelines-for-Undergraduate-Programs-in-Statistical-Science.aspx

Bond, M. E., Perkins, S. N., & Ramirez, C. (2012). Students' Perceptions of Statistics: An Exploration of Attitudes, Conceptualizations, and Content Knowledge of Statistics. *Statistics Education Research Journal*, *11*(2), 6–25. https://doi.org/10.52041/serj.v11i2.325

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn* (Vol. 11). Washington, DC: National academy press.

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Thousand Oaks, CA: SAGE.

Chambers, D. W. (1983). Stereotypic Images of the Scientist: The Draw-a-Scientist Test. *Science Education*, 67, 255–265. https://doi.org/10.1002/sce.3730670213.

Davidian, M. & Louis, T. A. (2012). Why statistics? *Science*, 336(6077). https://doi.org/10.1126/science.1218685

De Veaux, R. D., & Velleman, P. F. (2008). Math is music; statistics is literature (or, why are there no six-year-old novelists?). *Amstat News*, *375*, 54–58.

Findley, K**.** & Berens, F. (2020). Assessing the disciplinary perspectives of introductory statistics students. *Proceedings of the 23rd Annual Conference on Research in Undergraduate Mathematics Education* (pp. 1090–1095). Boston, MA. http://sigmaa.maa.org/rume/RUME23.pdf

Gordon, S. (2004). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal*, *3*(1), 40–59. https://iase-web.org/documents/SERJ/SERJ3(1)_gordon.pdf

Justice, N., Morris, S., Henry, V., & Fry, E. B. (2020). Paint-by-number or Picasso? A grounded theory phenomenographical study of students' conceptions of statistics. *Statistics Education Research Journal*, *19*(2). 76–102. https://doi.org/10.52041/serj.v19i2.111

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation.* Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Levrini, O., Fantini, P., Tasquier, G., Pecori, B., & Levin, M. (2015). Defining and Operationalizing Appropriation for Science Learning. *Journal of the Learning Sciences*, *24*(1), 93–136. https://doi.org/10.1080/10508406.2014.928215

National Science Foundation (2021), "Women, Minorities, and Persons with Disabilities." National Center for Science and Engineering Statistics Directorate for Social, Behavioral and Economic Sciences Report. https://ncses.nsf.gov/pubs/nsf21321/report

Reid, A., & Petocz, P. (2002). Students' conceptions of statistics: A phenomenographic study. *Journal of Statistics Education*, *10*(2). https://doi.org/10.1080/10691898.2002.11910662

Rolka, K., & Bulmer, M. (2005). Picturing student beliefs in statistics. *ZDM*, *37*(5), 412–417. https://doi.org/10.1007/s11858-005-0030-4

Wild, C., and Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review*, *67*(3), 223–265. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x