Notes and Script
Bold Signals 3.01: Science and Technology with Brian Nosek

Cover Art: "Homebrew" by Robert Tinney [Byte Magazine, (1980), 3(5)]

Introduction

Media	"Enterprise 1" by Languis

Welcome to the third season of Bold Signals!

A podcast about the people who produce, consume, and apply science.

My name is John Borghi, I am a cognitive neuroscientist and library post-doc.

In this episode:

I start a new segment "Scenes from the Replication Crisis" which discusses the methods scientists use through the lens of the people who created and critiqued them. First up, an introduction to the infamous p-value and the life and times of Ronald Fisher.

I interview Brian Nosek about his research on implicit bias and his work in making science more open and reproducible.

And finally, in another sorta new segment, I examine how science is presented and communicated outside by looking at some science documentaries. The first documentary I'll be covering in the Bold Signals Documentary Club is Cosmos: A Personal Journey. The seminal work by Carl Sagan.

Scenes from a Replication Crisis

Media	"Trees Don't Sleep" by Zachary Cale, Mighty Moon & Ethan Schmid

Introduction

In the past few seasons I've used this first segment to talk about everything from politics to systemic issues in science to the interaction of those two things. Well, in the time between the second and now third seasons of this podcast I've thought a lot about what it means to talk about science when words like "fake news" and "post truth" are becoming so common.

As a scientist turned librarian, I believe that things like evidence and expertise matter. My entire professional existence is predicated on the importance of data integrity and accessibility. Basically, that good things happen when people are given accurate information.

I'm a scientist who has spent a lot of time talking about openness and reproducibility. This episode features an interview with a scientist whose work is dedicated to those same things. But recently it has occurred to me that it would be valuable to add more context to those conversations. So let's talk about open science and reproducible science and what it means that we're in a replication crisis. But, while we do that, let's also talk about how we got here. Since science is a human enterprise, let's talk about the humans who developed the methods scientists use and the human impulses that may have led to those methods being applied in ways that are not necessarily ideal. Let's talk not about

Let's talk about scenes from the replication crisis.

And let's start, if not at the beginning, then in the first act with Ronald Fisher and one of the most significant phrases in modern science: $p<0.05$

The P-Value and Ronald Fisher

It's 1925

Ronald Fisher is a geneticist and statistician working at Rothamsted Experimental Station, an agricultural research institute located in the English countryside.

Before coming to Rothamsted, Fisher was instrumental in reconciling Charles Darwin's notion of evolution by natural selection with Gregor Mendel's Laws of Genetics. Basically, if you've ever wondered how Darwin's observations of Finches and Mendel's experiments with pea plants led to our modern understanding of evolution, one of the people you have to thank for that is Ronald Fisher.

It's also worth pointing out, given Fisher's influence on genetics, that he was an outspoken eugenicist. After all, this was the early twentieth century and the history of science is not exactly a straight line of people or non-horrific views on society.

Anyway, back to the countryside.

Long-term experiments with wheat, grass, and roots abound at Rothamsted, giving Fisher a bumper crop of data to analyze. However, though the overall quantity of data is high, sample sizes are low. An influential study of the effects of rainfall on wheat incorporates data from just thirteen plots of land.

Concerned with generalizing the results of such experiments, after all, the point of this type of research is to increase crop production, Fisher synthesizes several recent advances in "small sample statistics" into a framework known as significance testing.

He takes a statistical test called the Student's t-test, which was initially developed by statistician to monitor the quality of Guinness of all things, and develops a complementary test known which he calls the Analysis of Variance (ANOVA).

To ensure these innovations are accessible to the research community beyond Rothamsted, Fisher publishes Statistical Methods for Research Workers. Central to the book, and significance testing more generally, is the null hypothesis- the position that there is no significant difference between groups of data. In Fisher's conception, devices like t-tests and ANOVAs are tests of the null hypothesis. The results of such tests indicate the likelihood of observing a result when the null hypothesis is true. In quantitative terms, this likelihood is expressed as a p-value.

Fitting it's origins in applied research, the utility of Fisher's framework is best demonstrated with a practical example. Suppose Fisher and his colleagues want to study the effect of a particular method of fertilization on the growth of grass. To do this, they obtain yield measurements from ten plots that use the method and ten that do not. These numbers are small, but reflective of the time and effort that goes into harvesting good data. Before examining the two groups of data, Fisher reminds his colleagues that the null hypothesis stipulates that there is no difference between the fertilized and unfertilized plots. This is a really abstract way of talking about something as exciting as watching grass grow, so he reiterates that the null hypothesis is essentially that the fertilization method has no effect. Then, he runs a t-test.

A resulting p-value of 0.50 indicates that, assuming the fertilization method has no effect, the probability of Fisher and his colleagues obtaining their yield measurements is fifty percent. A resulting p-value of 0.10 indicates that the probability is ten percent. In Statistical Methods for Research Workers, Fisher introduces an informal criterion for rejecting the null hypothesis: $p < 0.05$.

"The value for which $p = 0.05$, or 1 in 20, is 1.96 or nearly 2 ; it is convenient to take this point as a limit in judging whether a deviation is to be considered significant or not."

We've arrived finally at $p<0.05$

Almost a decade after the publication of Statistical Methods for Research Workers, Jerzy Neyman and Egon Pearson address what they view as a fundamental asymmetry in Fisher's framework. Namely, though it's intended to help researchers evaluate the results of experiments, the focus on null hypotheses doesn't really give researchers any way to evaluate experimental hypotheses. Basically, the argue with increasing volume, you can use Fisher's methods to evaluate if there's a difference between two groups- but you can't used it to make a statement about what's causing it.

Though their "hypothesis testing" framework draws heavily from Fisher's, Neyman and Pearson's has a fundamentally different goal. Rather than giving researchers tools to evaluate the results of agricultural experiments, their goal is determining the most optimal test for deciding between competing hypotheses. These hypotheses include Fisher's null hypothesis, but also a

variety of "alternative" or experimental hypotheses. To this end, they introduce three important concepts to the burgeoning field of research-oriented statistics: Type I Error- The probability of incorrectly rejecting the null hypothesis, Type II Error- the probability of incorrectly accepting the null hypothesis, and Power- the probability of correctly rejecting the null hypothesis correctly.

Disagreements between Fisher and Neyman and Pearson soon escalates into open antagonism. No seriously, reading accounts of these debates you get the sense that Fisher's true talent wasn't in biology or statistics, but in expressing his ego mostly through yelling.

However, despite the controversy, the two frameworks are soon combined and presented as one in research methods textbooks. What emerges is an enormously and immediately influential model of statistical testing that incorporates Pearson's null hypothesis, Neyman and Pearson's alternative hypotheses, and a focus on observing p-values less than 0.05.

So when we talk about p-values and things like p-hacking, we're talking about a method for evaluating the difference between groups of data that was designed by an evolutionary biologist and eugenicist for use in agriculture. We're also talking about a debate about how what this number means and how to use it that has been ongoing for almost 80 years.

Next time, we'll talk about the effect of all this on the research literature.

Additional Reading

Box, J. F. (1987). Guinness, Gosset, Fisher, and small samples. *Statistical Science, 2*(1), 45-52.

Halpin, P. F., & Stam, H. J. (2006). Inductive inference or inductive behavior: Fisher and Neyman-Pearson approaches to statistical testing in psychological research (1940-1960). *The American Journal of Psychology, 119*(4), 625-653.

Fisher, R. A. (1918). The correlation between relatives on the supposition of Mendelian inheritance. *Transactions of the Royal Society of Edinburgh, 52*, 399-433.

Fisher, R. A. (1922). On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 222*, 309-368.

Fisher, R. A. (1925). *Statistical methods for research workers*. Oliver and Boyd.

Neyman, J., & Pearson, E. S. (1933). On the problem of the most efficient tests of statistical hypotheses. *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character, 231*, 289-337. doi: 10.1098/rsta.1933.0009

Lenhard, J. (2006). Models and statistical inference: The controversy between Fisher and Neyman–Pearson. *The British Journal for the Philosophy of Science, 57*(1), 69-91.

Student. (1908). The probable error of a mean. *Biometrika, 6*(1), 1-25.

Interview with Brian Nosek

Media	“Shoegaze” by Jahzzar

Interview Recorded: November 9th, 2016 (8:00 PST)
Recorded Remotely via Skype

Introduction

This week on Bold Signals I talk to Brian Nosek!

In the first half of our interview we talk about how he came to do the work he does now and what it means to study implicit biases.

In the second half of the interview We talk about the reproducibility project psychology, the CoS, and the Open Science Framework (OSF).

These things are easy conflate, so it's worth pointing out the the reproducibility project was an effort to reproduce 100 psychology studies. The OSF is an open source software platform that was used by researchers in the reproducibility project to plan and share data and things like that. It’s free to use and now includes tools for addressing the entire research lifecycle. The COS is a non-profit that grew around the OSF. Its stated mission is to "increase the openness, integrity, and reproducibility of scientific research.

For more information on anything discussed in this interview, check out Bold Signals.com

Links

Brian Nosek on Twitter: https://twitter.com/BrianNosek
Brian’s Website: http://projectimplicit.net/nosek/
Project Implicit: https://implicit.harvard.edu/implicit/
The Center for Open Science: https://cos.io/
The Open Science Framework: https://osf.io/

Recommended Readings

Greenwald, A. G., Banaji, M. R., Rudman, L. A., Farnham, S. D., Nosek, B. A., & Mellott, D. S. (2002). A unified theory of implicit attitudes, stereotypes, self-esteem, and self-concept. *Psychological Review, 109*(1), 3-25.

Nosek, B. A., Banaji, M., & Greenwald, A. G. (2002). Harvesting implicit group attitudes and beliefs from a demonstration web site. *Group Dynamics: Theory, Research, and Practice, 6*(1), 101-115.

Nosek, B. A., & Bar-Anan, Y. (2012). Scientific utopia: I. Opening scientific communication. *Psychological Inquiry, 23*(3), 217-243.
Nosek, B. A., Spies, J. R., & Motyl, M. (2012). Scientific utopia II. Restructuring incentives and practices to promote truth over publishability. *Perspectives on Psychological Science, 7*(6), 615-631.

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science, 349*(6251), aac4716.

Bold Signals Documentary Club

Media	"Lights of Tomorrow" by Starover Blue "Not a Song" by Scrapple "The Cosmos is…" monologue by Carl Sagan

Cosmos: A Personal Voyage
Episode 1: The Shores of the Cosmic Ocean
Original Airdate: September 28, 1980

We're introduced immediately to the conceit of the series, an exploration of the Cosmos - defined extremely broadly as all that is, all that was, and all that ever will be. At least in this episode, that exploration will be done on a ship of the imagination. A device that I first thought was meant to be allegorical, until Sagan starts talking from its control room.

What follows is part astronomy tour, part philosophy, all of which hedges very closely to science fiction. Its tone is way closer to Fantasia than a science documentary. We even visit an alien world. There are ample references to "majestic", "wonders". Lots of Carl Sagan staring out the window of his imaginary starship in Mars. Lots of analogy, lots of certainty.

Halfway through the episode, we finally get to earth and the story of Eratosthenes calculating the circumference of the earth.

Then we wander around the Library of Alexandria for a bit and make some assertions that are really speculations:

> "There may have been a portrait of Alexander."
> "First research institute in the world."
> "Genius flourished here - Euclid, Archimedes, Ptolemy."

We end with a conversation about the Cosmic Calendar, a visualization introduced by Sagan in popularized by Carl Sagan in his book *The Dragons of Eden*. Essentially, it's a teaching tool that condenses the lifespan of the universe into a single year to show the recency of human development.

> January 1st: Big Bang
> December 31 (10:30 pm): First humans
> December 31 (11:59.20 pm): Domestication of plants and animals

So, I have thoughts on all this. But first, I think it's important to give some context.

Carl Sagan, even 20 years after his death, is a towering figure in science communication. If you look hard enough, the way lots and lots of people perceive science can be traced back to his books and this documentary. 36 years after it was first broadcast, Cosmos remains the most watched PBS series in the world. That's a huge achievement. And I think Sagan should absolutely be commended for being a scientist set on popularizing, communicating, and advocating for science publically.

But. Wow. In retrospect, I have some problems with how things are presented in this episode.

First. While I think the tour around the universe in a ship of the imagination works as a poetic expression of the vastness of the universe, I don't think it's a particularly meaningful way of communicating actual astronomy. It also peters dangerously close to science fiction. This is a documentary that is explicitly dedicated to celebrating scientific fact and in the first half of the first hour we visit a world with an advanced alien civilization. I think that's really confusing. I get the idea, but the execution doesn't work for me.

I also really object to the depiction of science in the library of Alexandria section. Putting aside the fact that I started this podcast explicitly to undermine the stereotype of the lone scientific genius- which is on ample display as we talk about Ptolemy and Euclid and friends, but as a scientist turned librarian, I know something about libraries. And I happen to know that the story of the library of Alexandria is far less straightforward than is presented here. I don't think an extended conversation about budget problems and the necessity of backing up your collection of knowledge would make a particularly compelling science documentary, but I think it confuses the situation a bit when you're a myth as fact in a science documentary, even if you're only doing rhetorically so as a way to talk about something else.

Also- and maybe this won't be true in future episodes- but Sagan is literally the only person on screen for the whole hour. There are no expert interviews. For an episode that ends with a quote about the "The long, collective enterprise of science", we really only get one voice.

So a mixed bag so far.

I think the music and visuals are really awesome, even 35 years later.

But so far a lot of the presentation is almost antithetical to how science is done in the lab. Maybe that's just because it's the first episode, maybe that because this documentary is so influential that I'm having a hard time seeing it for what it is, but so far I think the emphasis on showing the majestic wonder of science sort of undercuts how hard it is for the people that actually do it.

Anyway, I'm going to keep with it.

Next time:

Episode 2: One Voice in the Cosmic Fugue
Episode 3: Harmony of the Worlds

Outro

Media	"Cabalista" by Wild Flag

Thanks for listening,

For links and show notes, visit our new website: BoldSignals.Com

You can also follow the podcast on social media: Facebook.com/BoldSignalsPod or Twitter.com/BoldSignalsPod

Email us comments at BoldSignals@Gmail

You can listen on Soundcloud, Itunes, Google Play, our your podcast app of choice.

If you like the podcast and want to help us out, please write us a review or leave us a rating on iTunes. The more we have, the more featured we are. We also wouldn't mind if you recommended the podcast to a friend.

Thanks again! Talk soon.