

Kate Bubacz / BuzzFeed News

SLICED & DICED

SCIENCE

Here's How Cornell Scientist Brian Wansink Turned Shoddy Data Into Viral Studies About How We Eat

Brian Wansink won fame, funding, and influence for his science-backed advice on healthy eating. Now, emails show how the Cornell professor and his colleagues have hacked and massaged low-quality data into headline-friendly studies to “go virally big time.”



Stephanie M. Lee
BuzzFeed News Reporter

Posted on February 25, 2018, at 8:45 p.m. ET

In the summer of 2013, Özge Siğirci, a young scientist in Turkey, had not yet arrived at Cornell University for her new research stint. But she already had an assignment from her future boss, Brian Wansink: Find something interesting about all-you-can-eat buffets.

As the head of Cornell's prestigious food psychology research unit, the Food and Brand Lab,

Wansink was a social science star. His dozens of studies about why and how we eat received mainstream attention everywhere from O, the Oprah Magazine to the Today show to the New York Times. At the heart of his work was an accessible, inspiring message: Weight loss is possible for anyone willing to make a few small changes to their environment, without need for strict diets or intense exercise.



Brian Wansink in 2013.
Gallo Images / Getty Images

When Siğirci started working with him, she was assigned to analyze a dataset from an experiment that had been carried out at an Italian restaurant. Some customers paid \$8 for the buffet, others half price. Afterward, they all filled out a questionnaire about who they were and how they felt about what they'd eaten.

Somewhere in those survey results, the professor was convinced, there *had* to be a meaningful relationship between the discount and the diners. But he wasn't satisfied by Siğirci's initial review of the data.

"I don't think I've ever done an interesting study where the data 'came out' the first time I looked at it," he told her over email.

More than three years later, Wansink would publicly praise Siğirci for being "the grad student who never said 'no.'" The unpaid visiting scholar from Turkey was dogged, Wansink wrote on his blog in November 2016. Initially given a "failed study" with "null results," Siğirci analyzed the data over and over until she began "discovering solutions that held up," he wrote. Her

tenacity ultimately turned the buffet experiment into four published studies about pizza eating, all cowritten with Wansink and widely covered in the press.



BuzzFeed News

But that’s not how science is supposed to work. Ideally, statisticians say, researchers should set out to prove a specific hypothesis before a study begins. Wansink, in contrast, was retroactively creating hypotheses to fit data patterns that emerged after an experiment was over.

Wansink couldn’t have known that his blog post would ignite a firestorm of criticism that now threatens the future of his three-decade career. Over the last 14 months, critics the world over have pored through more than 50 of his old studies and compiled “the Wansink Dossier,” a list of errors and inconsistencies that suggests he aggressively manipulated data. Cornell, after initially clearing him of misconduct, has opened an investigation. And he’s had five papers retracted and 14 corrected, the latest just this month.

Now, interviews with a former lab member and a trove of previously undisclosed emails show that, year after year, Wansink and his collaborators at the Cornell Food and Brand Lab have turned shoddy data into headline-friendly eating lessons that they could feed to the masses.

In correspondence between 2008 and 2016, the renowned Cornell scientist and his team discussed and even joked about exhaustively mining datasets for impressive-looking results. They strategized how to publish subpar studies, sometimes targeting journals with low standards. And they often framed their findings in the hopes of stirring up media coverage to, as Wansink once put it, “go virally big time.”

The correspondence shows, for example, how Wansink coached Siğirci to knead the pizza data.

First, he wrote, she should break up the diners into all kinds of groups: “males, females, lunch goers, dinner goers, people sitting alone, people eating with groups of 2, people eating in groups of 2+, people who order alcohol, people who order soft drinks, people who sit close to buffet, people who sit far away, and so on...”

Then she should dig for statistical relationships between those groups and the rest of the data: “# pieces of pizza, # trips, fill level of plate, did they get dessert, did they order a drink, and so on...”

“This is really important to try and find as many things here as possible *before* you come,” Wansink wrote to Siğirci. Doing so would not only help her impress the lab, he said, but “it would be the highest likelihood of you getting something publishable out of your visit.”

He concluded on an encouraging note: “Work hard, squeeze some blood out of this rock, and we’ll see you soon.”

Siğirci was game. “I will try to dig out the data in the way you described.”

All four of the pizza papers were eventually retracted or corrected. But the newly uncovered emails — obtained through records requests to New Mexico State University, which employs Wansink’s longtime collaborator Collin Payne — reveal two published studies that were based on shoddy data and have so far received no public scrutiny.

Still, Wansink defends his work.

“I stand by and am immensely proud of the work done here at the Lab,” he told BuzzFeed News by email, in response to a detailed list of allegations made in this story. “The Food and Brand Lab does not use ‘low-quality data,’ nor does it seek to publish ‘subpar studies’”

He pointed out that an independent lab confirmed the basic findings of the pizza papers. “That is, even where there has been unintentional error, the conclusions and impacts of the studies have not changed,” he wrote.

Siğirci and Payne did not respond to requests to comment for this story.

“I am sorry to say that it is difficult to read these emails and avoid a conclusion of research misconduct.”

Wansink’s practices are part of a troubling pattern of strategic data-crunching across the entire field of social science. Even so, several independent statisticians and psychology researchers are appalled at the extent of Wansink’s data manipulation.

“I am sorry to say that it is difficult to read these emails and avoid a conclusion of research misconduct,” Brian Nosek, a psychologist at the University of Virginia, told BuzzFeed News. As

executive director of the Center for Open Science, Nosek is one of his field's most outspoken reformers and spearheaded a massive project to try to reproduce prominent discoveries.

Based on the emails, Nosek said, “this is not science, it is storytelling.”



Wansink in the Food and Brand Lab in 2006.

David Duprey | Associated Press

The so-called replication crisis has punctured some of the world's most famous psychology research, from Amy Cuddy's work suggesting that “power poses” cause hormonal changes associated with feeling powerful, to Diederik Stapel's fabricated claims that messy environments lead to discrimination. In an influential 2015 report, Nosek's team attempted to repeat 100 psychology experiments, and reproduced less than one-half of the original findings.

One reason for the discrepancy is “p-hacking,” the taboo practice of slicing and dicing a dataset for an impressive-looking pattern. It can take various forms, from tweaking variables to show a desired result, to pretending that a finding proves an original hypothesis — in other words, uncovering an answer to a question that was only asked after the fact.

In psychology research, a result is usually considered statistically significant when a calculation called a p-value is less than or equal to 0.05. But excessive data massaging can wind up with a p-value lower than 0.05 just by random chance, making a hypothesis seem valid when it's actually a fluke.

Wansink said his lab's data is “heavily scrutinized,” and that's “what exploratory research is all

about.”

But for years, Wansink’s inbox has been filled with chatter that, according to independent statisticians, is blatant p-hacking.

“Pattern doesn’t look good,” Payne of New Mexico State wrote to Wansink and David Just, another Cornell professor, in April 2009, after what Payne called a “marathon” data-crunching session for an experiment about eating and TV-watching.

“I also ran — i am not kidding — 400 strategic mediation analyses to no avail...” Payne wrote. In other words, testing 400 variables to find one that might explain the relationship between the experiment and the outcomes. “The last thing to try — but I shutter to think of it — is trying to mess around with the mood variables. Ideas...suggestions?”

Two days later, Payne was back with promising news: By focusing on the relationship between two variables in particular, he wrote, “we get exactly what we need.” (The study does not appear to have been published.)

“That’s p-hacking on steroids.”

“That’s p-hacking on steroids,” said Kristin Sainani, an associate professor of health research and policy at Stanford University. “They’re running every possible combination of variables, essentially, to see if anything will come up significant.”

In a conversation about another study in August 2015, Wansink mentioned a series of experiments that “were chasing interactions that were hard to find.” He apparently hoped that they would all arrive at the same conclusion, which is “bad science,” said Susan Wei, an assistant professor of biostatistics at the University of Minnesota.

“It does very much seem like this Brian Wansink investigator is a consistent and repeated offender of statistics,” Wei added. “He’s so brazen about it, I can’t tell if he’s just bad at statistical thinking, or he knows that what he’s doing is scientifically unsound but he goes ahead anyway.”

In 2012, Wansink, Payne, and Just published one of their most famous studies, which revealed an easy way of nudging kids into healthy eating choices. By decorating apples with stickers of Elmo from *Sesame Street*, they claimed, elementary school students could be swayed to pick the fruit over cookies at lunchtime.

But back in September 2008, when Payne was looking over the data soon after it had been collected, he found no strong apples-and-Elmo link — at least not yet.

“I have attached some initial results of the kid study to this message for your report,” Payne wrote to his collaborators. “Do not despair. It looks like stickers on fruit may work (with a bit

more wizardry).”

Wansink also acknowledged the paper was weak as he was preparing to submit it to journals. The p-value was 0.06, just shy of the gold standard cutoff of 0.05. It was a “sticking point,” as he put it in a Jan. 7, 2012, email.

“It seems to me it should be lower,” he wrote, attaching a draft. “Do you want to take a look at it and see what you think. If you can get the data, and it needs some tweaking, it would be good to get that one value below .05.”

From: [Brian Wansink](#)
To: [David Just](#)
Cc: [Collin Payne](#); [Sandra Cuellar](#)
Subject: Can Branding Improve School Lunches?
Date: Saturday, January 7, 2012 7:17:42 AM
Attachments: [Elmo Icon-AJPH - 1-7-12.doc](#)
[ATT00001.htm](#)

Hi David,

Here's the Elmo study we are going to spin off and submit. I think we start with the AJPH as a Brief (80 word abstract and 800 word paper), and go from there. I'll give Sandra a list of the journals and the priority order we should consider. Let's consider these two first:

Brief -- American Journal of Public Health

Research Letter -- Archives of Pediatric and Adolescent Medicine

One sticking point is that although the stickers increase apple selection by 71%, for some reason this is a p value of .06. It seems to me it should be lower. Do you want to take a look at it and see what you think. If you can get the data, and it needs some tweaking, it would be good to get that one value below .05.

Best,

Brian

BuzzFeed News

Later in 2012, the study appeared in the prestigious JAMA Pediatrics, the 0.06 p-value intact. But in September 2017, it was retracted and replaced with a version that listed a p-value of 0.02. And a month later, it was retracted yet again for an entirely different reason: Wansink admitted that the experiment had not been done on 8- to 11-year-olds, as he'd originally claimed, but on preschoolers.

Scientists are under a lot of pressure to attain the 0.05 p-value, said Wei of the University of Minnesota, even though it's an arbitrary cutoff. “It's an unfortunate state of being in the research community, in the publishing world.”

Still, the Food and Brand Lab appears to approach science far more flagrantly than most other scientists who face the same pressures, Nosek said.

“It’s a cartoon of how someone in the most extreme form might p-hack data,” he said of the emails as a whole. “There was the explicit goal of ‘Let’s just get something out of the data, use the data as a device to find something, anything, that’s interesting.’”

Back in March of last year, shortly after his pizza papers were called into question, Wansink was interviewed by the Chronicle of Higher Education. He told the outlet that before all the hubbub over his studies, he’d never heard of the term “p-hacking” or the replication crisis. “Science is messy in a lot of ways,” he said.



Mike Groll / AP

The emails reviewed by BuzzFeed News point to potential problems with two of Wansink’s studies that haven’t received any public criticism.

In July 2009, Wansink wrote to collaborators about a study in progress. Mall shoppers had been asked to read a pamphlet that described one of two kinds of walks — one that focused on listening to music and the other on exercising. At the end, researchers offered them salty and sweet snacks as a thank you, and recorded how much participants served themselves.

“What’s neat about this is that it shows that just thinking about walking makes people eat more,” Wansink wrote. “But we should be able to get much more from this.” He added, “I think

it would be good to mine it for significance and a good story.”

Meanwhile, one of his coauthors, former Cornell visiting scholar Carolina Werle, was trying — in vain — to find a link between how much time the participants spent in the mall and other variables she’d tested. “There is no interaction with the experimental conditions,” she wrote, adding that “nothing works in terms of interaction.”

Werle did not respond to a request to comment for this story. The study wound up in the journal Appetite two years later, claiming that just the act of imagining physical activity prompts people to take more snacks.

In 2013, Werle and Wansink were discussing a different study about whether describing a walk as fun, such as by framing it as a scenic stroll rather than a form of exercise, influenced how much the walkers would want to eat afterward.

The scientists emailed back and forth about two facts that the subsequent paper did not disclose: that the “exercise” group was much smaller than the “fun” group, and that there was some missing data.

In another message, he mentioned that “there’s been a lot of data torturing with this cool data set.”

“Why are you hiding that? That seems like pretty pertinent information,” Sainani of Stanford said.

Wansink himself acknowledged that the data was weak. At one point, he told Werle that he didn’t think the study was good enough to submit to a prestigious journal. In another message, he mentioned that “there’s been a lot of data torturing with this cool data set.” The journal Marketing Letters eventually published the paper in May 2014.

Journals should scrutinize these two papers due to questions raised by the emails, said Nick Brown, an independent graduate student at the University of Groningen in the Netherlands who has spent more than a year dissecting Wansink’s work.

The paper that told walkers to think of their exercise as fun, for example, reports p-values of 0.04 and less than 0.05. But by Brown’s calculation, they appear to have been cut in half using an unjustified statistical method. (Wei said the group’s method may have been acceptable.) And one of Werle’s emails to Wansink indicated that one of the experiments had 47 people, but the paper reports 46.

Brown remains befuddled by the breadth of errors that he and other critics have unearthed in the Food and Brand Lab’s research, from data impossibilities to sample sizes that don’t add up.

“They’re doing the p-hacking and they’re getting other stuff wrong, badly wrong,” he said. “The level of incompetence that would be required for that is just staggering.”



Wansink demonstrates his "bottomless bowl of soup" after he was awarded a 2007 Ig Nobel Prize.

Stan Honda / AFP / Getty Images

These papers are only a fraction of the 250-plus that Wansink has produced since the early 1990s. One of the driving factors behind his prolific output, his emails suggest, is a hard truth that all academic scientists face: The more papers you publish, the more likely you are to be rewarded in promotions, funding, and fame.

In August 2012, Wansink relayed some sad news to more than a dozen colleagues: One of his former grad students had just been denied tenure.

“Whenever I called her, it always seemed like she was working on a ton of stuff,” he wrote. Yet he was surprised to see that she’d only published one article in a minor journal in seven years.

This unnamed scholar was working on lots of research that never saw the light of the day, Wansink wrote: “Too much inventory; not enough shipments.” Ideally, he mused, a science lab would function like a tech company. Tim Cook, for instance, was renowned for getting products out of Apple’s warehouses faster and pumping up profits. “As Steve Jobs said, ‘Geniuses ship,’” Wansink wrote.

So, he proposed, the lab should adopt a system of strict deadlines for submitting and

resubmitting research until it landed somewhere. “A lot of these papers are laying around on each of our desktops and they’re like inventory that isn’t working for us,” he told the team. “We’ve got so much huge momentum going. This could make our productivity legendary.”

Kirsikka Kaipainen knows firsthand how the Food and Brand Lab manages to churn out so many papers. She did research there in 2012 as a graduate student visiting from Finland.

Wansink was “warm and generous and energetic, somebody who would take you in with open arms,” she recalled.

Soon after she arrived, Wansink handed her a dataset from an online weight-loss program run by the lab. At a brainstorming session, he led the charge in proposing a half dozen or so papers that could come out of the data, Kaipainen said.

But when Kaipainen drilled down, there wasn’t enough information to support the conclusions he wanted. “I felt I was a failure,” she said. “I realized over time it wasn’t my wrongdoing — it was just that there was nothing there to be found.”

Kaipainen eventually steered the paper to a place she was comfortable with, a straightforward review of how the program worked. Wansink was fine with it, she said, since it wound up in a notable journal. But even though his name is on it, he had little to do with writing it — and his few contributions made her uneasy, she said.

“That’s when I started feeling like, this is not the kind of research I want to do.”

“Like he hadn’t really looked at the results critically and he was trying to make the paper say something that wasn’t true,” she said. “That’s when I started feeling like, this is not the kind of research I want to do.” (In 2014, Wansink would cite this paper in a Kickstarter campaign that raised \$10,000 for a weight-loss service that never launched.)

As months went by, Kaipainen began to notice that her experience fit into a disturbing pattern in the lab as a whole: Collect data first, form hypotheses later. She isn’t the only former lab member who felt troubled by how the place was run. Other students have told the Cornell Daily Sun that Wansink and his team often chased down findings that were intuitive-sounding yet overly broad and unreliable.

Kaipainen was also troubled by lab members’ frequent remarks that a finding “would make a good story or a good article” or “be interesting for media.” At one point, the lab brainstormed questions for an outside party’s market research survey. “That was weird also,” she recalled, “to come up with some questions not based on any theory, just ‘What would be cool to ask?’, ‘What cool headlines could we get if we got some associations?’”

Even when writing manuscript drafts, Wansink was thinking about how to sell his

inspirational results to the public.

“We want this to go virally big time,” he wrote to Payne, Just, and another Cornell colleague on March 27, 2012, explaining why he wanted to make the chart labels for the Elmo and apples study sound “more generalizable.” Another time, he discussed playing up “the quotable point” of the study.

The study about mall shoppers thinking about exercising, he once wrote, “would make quite a media splash.” Another time, he proposed: “Let’s think of renaming the paper to something more shameless. Maybe something like “Thinking about exercise makes me hungry.”” (Its published title was “Just thinking about exercise makes me serve more food: physical activity and calorie compensation.”)

Wansink told BuzzFeed News that attracting media coverage is a core part of the Food and Brand Lab’s mission “to be accessible and impactful.”

“The reality is that, by and large, the public does not have access to academic articles,” he said. “Framing our work in a way that can be accessed and used by the public is something all social scientists should consider.”

These days, Kaipainen leads an online mental health startup. She said she stands by the four papers she wrote with Wansink. And she believes her old boss has good intentions at his core. “He wants to help people eat better. I kind of admire him for it,” she said. “But ends don’t justify the means.”

To Nosek, the relentless pursuit of fawning press is just one element of a backward rewards system in science. When a university is deciding who to hire, or a funding institution is selecting a grantee, it’s easy to pick those who are publishing a lot, he said.

“And scientists,” Nosek added, “have to do those dysfunctional behaviors — do everything you can to seek an exciting, sexy story, even if it means reducing the credibility of the research.”

Scientists often rely on peer reviewers — anonymous experts — to weed out errors in papers before they go to press. But journals didn't, or couldn't, catch every inaccuracy from the Food and Brand Lab.

For example, reviewers were generally positive about what ended up being a controversial 2012 study in Preventive Medicine. It reported that schoolchildren ate more vegetables at lunch when they had catchy names like "X-Ray Vision Carrots."

"This is a well-written manuscript that addressed critical questions pertinent to vegetable consumption among school-aged children," one reviewer wrote. Another praised the experiments because they "use real-world data and show changes in actual behavior."

Last fall, Wansink admitted that the lunchtime observation part of the carrots study had actually been done on preschoolers, not the reported 8- to 11-year-olds (just like the Elmo and apples article). The paper was corrected on Feb. 1, shocking critics who had called for a full retraction. In their defense, the researchers noted that a sub-experiment mentioned in the study did involve elementary schoolchildren. "These mistakes and omissions do not change the general conclusion of the paper," they wrote.

The scientists who reviewed the Elmo study in JAMA Pediatrics didn't seem to raise any red flags either, based on Wansink's responses to them. But the journal correspondence does point to a bizarre new discrepancy.

Originally, Wansink made it clear that the researchers had let children pick both apples and cookies if they wanted. “We emphasized why we set up a system that allowed children to a cookie or apple, or both, or neither,” he explained to the journal editor in a March 2012 draft letter. But five years later, when the team retracted the study the first time, the researchers publicly contradicted this statement, writing that “children were offered to take either an apple or a cookie (not both).”

Sometimes, the team’s acceptance into top journals surprised even them. In July 2009, Payne and Just, the Cornell professor, were celebrating a paper that had landed in the Annals of Behavioral Medicine.

“Usually for a top 10 journal I am trying to reinvent statistics.”

“Is it me or was this way too easy for a journal ranked 10th out of 101 journals in its category?” Payne wrote. “Usually for a top 10 journal I am trying to reinvent statistics.” (Just did not respond to a request for comment.)

When their work was rejected, the members of the Food and Brand Lab would often try increasingly lower-quality journals until they succeeded. This practice is in part responsible for the sheer volume of scientific findings that cannot be replicated.

Even so, Wansink and his colleagues were sometimes frank about their desire to put flawed data in publications with low standards.

In March 2009, Wansink asked Payne to look over a manuscript about how people shop in foreign countries where they’re unfamiliar with the brands or language. Wansink had written it with Koert van Ittersum, a professor at the University of Groningen, who did not respond to a request for comment.

“There’s a paper that Koert and I have had shelved for about 5 years that needs to see the journal light of day,” Wansink said over email. It was based on experiments that were “insightful, but not very rigorous.”

He continued, “There’s a bunch of 2nd-tier journals (or lower) we could send it to,” and added, “...we’d kind of like to aim at a pretty quick acceptance.”

Time was crucial. “At this point,” Wansink wrote, “we don’t want to redo any studies, just deal with the ones we have and find an editor that says, ‘Interesting, and good enough.’” So far, none have.

To Brown, the grad student who may know Wansink’s work better than Wansink himself, the lab is like a food-processing company.

“What they’re doing is making a very small amount of science go a very long way when you spread it thinly and you cut it with water and modified starch,” he said. “The product, which is the paper, is designed and marketed before it’s even been built.”

He suspects that more of Wansink’s papers will be corrected or even retracted. (Cornell confirmed to BuzzFeed News that its investigation into Wansink’s research is ongoing, but declined to share any details.)

Brown is less sure, however, whether this saga will lead to any meaningful changes beyond Wansink’s oeuvre. As illuminating as the emails are — he likens them to security footage of a bank robbery — he also thinks that the behavior is “probably quite typical of what goes on in a lot of labs.”

“Science by volume, science by output, by yardage, that has to stop,” Brown said. “The problem is that it’s an awful lot of people’s paychecks and an awful lot of people’s business models.”●

CORRECTION

February 27, 2018, at 10:12 a.m.

A 2012 study in Preventive Medicine reported that schoolchildren ate more vegetables at lunch when they had catchy names like “X-Ray Vision Carrots.” A previous version of this story misstated that name.

CORRECTION

February 28, 2018, at 2:58 p.m.

Amy Cuddy’s research suggesting that “power poses” cause hormonal changes associated with feeling powerful has not been replicated. An earlier version of this story suggested that her research linking “power poses” to feeling confident has not been replicated, but this finding has been replicated.

Emails Show How An Ivy League Prof Tried To Do Damage Control For His Bogus Food Science

buzzfeed.com

Here’s How A Controversial Study About Kids And Cookies Turned Out To Be Wrong — And Wrong Again

buzzfeed.com

This Controversial Ivy League Scientist Left His Kickstarter Donors High And Dry

buzzfeed.com



Stephanie M. Lee is a science reporter for BuzzFeed News and is based in San Francisco.

Contact Stephanie M. Lee at stephanie.lee@buzzfeed.com.

Got a confidential tip? [Submit it here](#).