



ARMS RACE

Going to College

If you sit down to dinner with friends in certain cities—San Francisco and Portland, to name two—you'll likely find that sharing plates is an impossibility. No two people can eat the same things. They're all on different diets. These range from vegan to various strains of Paleo, and people swear by them (if only for a month or two). Now imagine if one of those regimes, say the caveman diet, became the national standard: if 330 million people all followed its dictates.

The effects would be dramatic. For starters, a single national diet would put the agricultural economy through the wringer. Demand for the approved meats and cheeses would skyrocket,

pushing prices up. Meanwhile, the diet's no-no sectors, like soybeans and potatoes, would go begging. Diversity would shrivel. Suffering bean farmers would turn over their fields to cows and pigs, even on land unsuited for it. The additional livestock would slurp up immense quantities of water. And needless to say, a single diet would make many of us extremely unhappy.

What does a single national diet have to do with WMDs? Scale. A formula, whether it's a diet or a tax code, might be perfectly innocuous in theory. But if it grows to become a national or global standard, it creates its own distorted and dystopian economy. This is what has happened in higher education.

The story starts in 1983. That was the year a struggling news-magazine, *U.S. News & World Report*, decided to undertake an ambitious project. It would evaluate 1,800 colleges and universities throughout the United States and rank them for excellence. This would be a useful tool that, if successful, would help guide millions of young people through their first big life decision. For many, that single choice would set them on a career path and introduce them to lifelong friends, often including a spouse. What's more, a college-ranking issue, editors hoped, might turn into a newsstand sensation. Perhaps for that one week, *U.S. News* could match its giant rivals, *Time* and *Newsweek*.

But what information would feed this new ranking? In the beginning, the staff at *U.S. News* based its scores entirely on the results of opinion surveys it sent to university presidents. Stanford came out as the top national university, and Amherst as the best liberal arts college. While popular with readers, the ratings drove many college administrators crazy. Complaints poured into the magazine that the rankings were unfair. Many college presidents, students, and alumni insisted that they deserved a higher ranking. All the magazine had to do was look at the *data*.

In the following years, editors at *U.S. News* tried to figure out

what they could measure. This is how many models start out, with a series of hunches. The process is not scientific and has scant grounding in statistical analysis. In this case, it was just people wondering what matters most in education, then figuring out which of those variables they could count, and finally deciding how much weight to give each of them in the formula.

In most disciplines, the analysis feeding a model would demand far more rigor. In agronomy, for example, researchers might compare the inputs—the soil, the sunshine, and fertilizer—and the outputs, which would be specific traits in the resulting crops. They could then experiment and optimize according to their objectives, whether price, taste, or nutritional value. This is not to say that agronomists cannot create WMDs. They can and do (especially when they neglect to consider long-term and wide-ranging effects of pesticides). But because their models, for the most part, are tightly focused on clear outcomes, they are ideal for scientific experimentation.

The journalists at *U.S. News*, though, were grappling with “educational excellence,” a much squishier value than the cost of corn or the micrograms of protein in each kernel. They had no direct way to quantify how a four-year process affected one single student, much less tens of millions of them. They couldn’t measure learning, happiness, confidence, friendships, or other aspects of a student’s four-year experience. President Lyndon Johnson’s ideal for higher education—“a way to deeper personal fulfillment, greater personal productivity and increased personal reward”—didn’t fit into their model.

Instead they picked proxies that seemed to correlate with success. They looked at SAT scores, student-teacher ratios, and acceptance rates. They analyzed the percentage of incoming freshmen who made it to sophomore year and the percentage of those who graduated. They calculated the percentage of living alumni who

contributed money to their alma mater, surmising that if they gave a college money there was a good chance they appreciated the education there. Three-quarters of the ranking would be produced by an algorithm—an opinion formalized in code—that incorporated these proxies. In the other quarter, they would factor in the subjective views of college officials throughout the country.

U.S. News’s first data-driven ranking came out in 1988, and the results seemed sensible. However, as the ranking grew into a national standard, a vicious feedback loop materialized. The trouble was that the rankings were self-reinforcing. If a college fared badly in *U.S. News*, its reputation would suffer, and conditions would deteriorate. Top students would avoid it, as would top professors. Alumni would howl and cut back on contributions. The ranking would tumble further. The ranking, in short, was destiny.

In the past, college administrators had had all sorts of ways to gauge their success, many of them anecdotal. Students raved about certain professors. Some graduates went on to illustrious careers as diplomats or entrepreneurs. Others published award-winning novels. This all led to good word of mouth, which boosted a college’s reputation. But was Macalester better than Reed, or Iowa better than Illinois? It was hard to say. Colleges were like different types of music, or different diets. There was room for varying opinions, with good arguments on both sides. Now the vast reputational ecosystem of colleges and universities was overshadowed by a single column of numbers.

If you look at this development from the perspective of a university president, it’s actually quite sad. Most of these people no doubt cherished their own college experience—that’s part of what motivated them to climb the academic ladder. Yet here they were at the summit of their careers dedicating enormous energy toward boosting performance in fifteen areas defined by a group of journalists at a second-tier newsmagazine. They were almost like

students again, angling for good grades from a taskmaster. In fact, they were trapped by a rigid model, a WMD.

If the *U.S. News* list had turned into a moderate success, there would be no trouble. But instead it grew into a titan, quickly establishing itself as a national standard. It has been tying our education system into knots ever since, establishing a rigid to-do list for college administrators and students alike. The *U.S. News* college ranking has great scale, inflicts widespread damage, and generates an almost endless spiral of destructive feedback loops. While it's not as opaque as many other models, it is still a bona fide WMD.

Some administrators have gone to desperate lengths to drive up their rank. Baylor University paid the fee for admitted students to *retake* the SAT, hoping another try would boost their scores—and Baylor's ranking. Elite small schools, including Bucknell University in Pennsylvania and California's Claremont McKenna, sent false data to *U.S. News*, inflating the SAT scores of their incoming freshmen. And Iona College, in New York, acknowledged in 2011 that its employees had fudged numbers about nearly everything: test scores, acceptance and graduation rates, freshman retention, student-faculty ratio, and alumni giving. The lying paid off, at least for a while. *U.S. News* estimated that the false data had lifted Iona from fiftieth to thirtieth place among regional colleges in the Northeast.

The great majority of college administrators looked for less egregious ways to improve their rankings. Instead of cheating, they worked hard to improve each of the metrics that went into their score. They could argue that this was the most efficient use of resources. After all, if they worked to satisfy the *U.S. News* algorithm, they'd raise more money, attract brighter students and professors, and keep rising on the list. Was there really any choice?

Robert Morse, who has worked at the company since 1976 and heads up the college rankings, argued in interviews that the rankings pushed the colleges to set meaningful goals. If they could im-

prove graduation rates or put students in smaller classes, that was a good thing. Education benefited from the focus. He admitted that the most relevant data—what the students had learned at each school—was inaccessible. But the *U.S. News* model, constructed from proxies, was the next best thing.

However, when you create a model from proxies, it is far simpler for people to game it. This is because proxies are easier to manipulate than the complicated reality they represent. Here's an example. Let's say a website is looking to hire a social media maven. Many people apply for the job, and they send information about the various marketing campaigns they've run. But it takes way too much time to track down and evaluate all of their work. So the hiring manager settles on a proxy. She gives strong consideration to applicants with the most followers on Twitter. That's a sign of social media engagement, isn't it?

Well, it's a reasonable enough proxy. But what happens when word leaks out, as it surely will, that assembling a crowd on Twitter is key for getting a job at this company? Candidates soon do everything they can to ratchet up their Twitter numbers. Some pay \$19.95 for a service that populates their feed with thousands of followers, most of them generated by robots. As people game the system, the proxy loses its effectiveness. Cheaters wind up as false positives.

In the case of the *U.S. News* rankings, everyone from prospective students to alumni to human resources departments quickly accepted the score as a measurement of educational quality. So the colleges played along. They pushed to improve in each of the areas the rankings measured. Many, in fact, were most frustrated by the 25 percent of the ranking they had no control over—the reputational score, which came from the questionnaires filled out by college presidents and provosts.

This part of the analysis, like any collection of human opinion, was sure to include old-fashioned prejudice and ignorance. It

tended to protect the famous schools at the top of the list, because they were the ones people knew about. And it made it harder for up-and-comers.

In 2008, Texas Christian University in Fort Worth, Texas, was tumbling in the *U.S. News* ranking. Its score, which had been 97 three years earlier, had fallen to 105, 108, and now 113. This agitated alumni and boosters and put the chancellor, Victor Boschini, in the hot seat. "The whole thing is very frustrating to me," Boschini told the campus news site, *TCU 360*. He insisted that TCU was advancing in every indicator. "Our retention rate is improving, our fundraising, all the things they go on."

There were two problems with Boschini's analysis. First, the *U.S. News* ranking model didn't judge the colleges in isolation. Even schools that improved their numbers would fall behind if others advanced faster. To put it in academic terms, the *U.S. News* model graded colleges on a curve. And that fed what amounted to a growing arms race.

The other problem was the reputational score, the 25 percent TCU couldn't control. Raymond Brown, the dean of admissions, noted that reputation was the most heavily weighted variable, "which is absurd because it is entirely subjective." Wes Waggoner, director of freshman admissions, added that colleges marketed themselves to each other to boost their reputational score. "I get stuff in the mail from other colleges trying to convince [us] that they're a good school," Waggoner said.

Despite this grouching, TCU set out to improve the 75 percent of the score it could control. After all, if the university's score rose, its reputation would eventually follow. With time, its peers would note the progress and give it higher numbers. The key was to get things moving in the right direction.

TCU launched a \$250 million fund-raising drive. It far surpassed its goal and brought in \$434 million by 2009. That alone

boosted TCU's ranking, since fund-raising is one of the metrics. The university spent much of the money on campus improvements, including \$100 million on the central mall and a new student union, in an effort to make TCU a more attractive destination for students. While there's nothing wrong with that, it conveniently feeds the *U.S. News* algorithm. The more students apply, the more selective the school can be.

Perhaps more important, TCU built a state-of-the-art sports training facility and pumped resources into its football program. In the following years, TCU's football team, the Horned Frogs, became a national powerhouse. In 2010, they went undefeated, beating Wisconsin in the Rose Bowl.

That success allowed TCU to benefit from what's called "the Flutie effect." In 1984, in one of the most exciting college football games in history, a quarterback at Boston College, Doug Flutie, completed a long last-second "Hail Mary" pass to defeat the University of Miami. Flutie became a legend. Within two years, applications to BC were up by 30 percent. The same boost occurred for Georgetown University when its basketball team, anchored by Patrick Ewing, played in three national championship games. Winning athletic programs, it turns out, are the most effective promotions for some applicants. To legions of athletically oriented high school seniors watching college sports on TV, schools with great teams look appealing. Students are proud to wear the school's name. They paint their faces and celebrate. Applications shoot up. With more students seeking admission, administrators can lift the bar, raising the average test scores of incoming freshmen. That helps the rating. And the more applicants the school rejects, the lower (and, for the ranking, better) its acceptance rate.

TCU's strategy worked. By 2013, it was the second most selective university in Texas, trailing only prestigious Rice University in Houston. That same year, it registered the highest SAT and

ACT scores in its history. Its rank in the *U.S. News* list climbed. In 2015, it finished in seventy-sixth place, a climb of thirty-seven places in just seven years.

Despite my issues with the *U.S. News* model and its status as a WMD, it's important to note that this dramatic climb up the rankings may well have benefited TCU as a university. After all, most of the proxies in the *U.S. News* model reflect a school's overall quality to some degree, just as many dieters thrive by following the caveman regime. The problem isn't the *U.S. News* model but its scale. It forces everyone to shoot for exactly the same goals, which creates a rat race—and lots of harmful unintended consequences.

In the years before the rankings, for example, college-bound students could sleep a bit better knowing that they had applied to a so-called safety school, a college with lower entrance standards. If students didn't get into their top choices, including the long shots (stretch schools) and solid bets (target schools), they'd get a perfectly fine education at the safety school—and maybe transfer to one of their top choices after a year or two.

The concept of a safety school is now largely extinct, thanks in great part to the *U.S. News* ranking. As we saw in the example of TCU, it helps in the rankings to be selective. If an admissions office is flooded with applications, it's a sign that something is going right there. It speaks to the college's reputation. And if a college can reject the vast majority of those candidates, it'll probably end up with a higher caliber of students. Like many of the proxies, this metric seems to make sense. It follows market movements.

But that market can be manipulated. A traditional safety school, for example, can look at historical data and see that only a small fraction of the top applicants ended up going there. Most of them got into their target or stretch schools and didn't need what amounted to an insurance policy. With the objective of boosting its selectivity score, the safety school can now reject the excellent

candidates that, according to its own algorithm, are most likely not to matriculate. This process is far from exact. And the college, despite the work of the data scientists in its admissions office, no doubt loses a certain number of top students who would have chosen to attend. Those are the ones who learn, to their dismay, that so-called safety schools are no longer a sure bet.

The convoluted process does nothing for education. The college suffers. It loses the top students—the stars who enhance the experience for everyone, including the professors. In fact, the former safety school may now have to allocate some precious financial aid to enticing some of those stars to its campus. And that may mean less money for the students who need it the most.

...

It's here that we find the greatest shortcoming of the *U.S. News* college ranking. The proxies the journalists chose for educational excellence make sense, after all. Their spectacular failure comes, instead, from what they chose *not* to count: tuition and fees. Student financing was left out of the model.

This brings us to the crucial question we'll confront time and again. What is the objective of the modeler? In this case, put yourself in the place of the editors at *U.S. News* in 1988. When they were building their first statistical model, how would they know when it worked? Well, it would start out with a lot more credibility if it reflected the established hierarchy. If Harvard, Stanford, Princeton, and Yale came out on top, it would seem to validate their model, replicating the informal models that they and their customers carried in their own heads. To build such a model, they simply had to look at those top universities and count what made them so special. What did they have in common, as opposed to the safety school in the next town? Well, their students had stratospheric SATs and graduated like clockwork. The alumni were rich