WHEN AND HOW SHOULD WE "TRUST THE SCIENCE"?

By Duane Blackburn

The COVID-19 pandemic, paradigm-shifting technological advancements, and (unfortunately) politics have placed a bright spotlight on the role of science within public policy. Science and technology (S&T) "facts" are alternatively used to argue multiple sides of debates, cherry-picked or taken out of context to "prove" an individual's desired outcomes, or dismissed as being irrelevant if they do not support political objectives.

Scientific and technological advancements have long been the bedrock of our nation's security and prosperity. Public confidence in the science community is higher than many other institutions and has been fairly constant for decades,[1] but events of 2020 seem to have fractured that confidence as partisan divides and personal biases increased in prominence. Because of this increase, combined with ubiquitous use of social media for millions to share their opinions, public discussion of S&T policy issues is rapidly growing and devolving into self-reinforcing (and opposing) camps.

> **It is just extraordinary what we've been through in the past year. If you were to write a script about how to destroy the credibility of science, we just saw it.**
>
> Eric Topol, founder and director of the Scripps Research Translational Institute[2]

It is clear we must work to help policymakers, the press, and the public better understand when and how they should "trust the science". This paper provides an initial primer along two important lanes to aid with their understanding: the science itself and how the science is being explained.

How Solid Is the Science?

Many everyday conversations and influencer campaigns use the term *science* interchangeably, and inaccurately, with *facts*. Let's be perfectly clear: science is not automatically infallible. The very nature of scientific discovery is a series of hits and misses, then arguing about those hits and misses until the learned community coalesces around a solidly proven idea. Sometimes, though not very often, that proven consensus ends up being disproven decades later! One notable example is that for centuries the world's greatest scientific minds were convinced that the sun revolved around the earth. The point to remember is that scientists always analyze and reanalyze what we think to be true—science is never completely settled.

Scientists have wide discretion when selecting inputs and parameters of their research, which can unintentionally skew results. They are also human. Like everyone else, they have biases based on past experiences and a tendency to promote positive results while hiding negative results (or promoting the finding minus necessary discussion about the uncertainty in that finding).

Fortunately, the Scientific Method helps scientists work through these issues and mitigate the influence of error and bias. It can also help nonscientists better understand scientific activities. Used for centuries, the Scientific Method drives the development of scientific understanding.[3]

For each research question, scientists create a hypothesis, then design experiments to see if the hypothesis is correct. Because most hypotheses are incorrect—at least not completely, or in all conditions—scientists report what they've learned, then leverage new insight and data to adjust their hypothesis. This process repeats itself until scientists find a proven hypothesis.[4] After they share this information, other scientists run similar experiments to prove the claim. Once enough additional scientists have also confirmed the hypothesis—without any others providing similarly rigorous experimental results that contrast the hypothesis—the community considers the hypothesis to be proven.

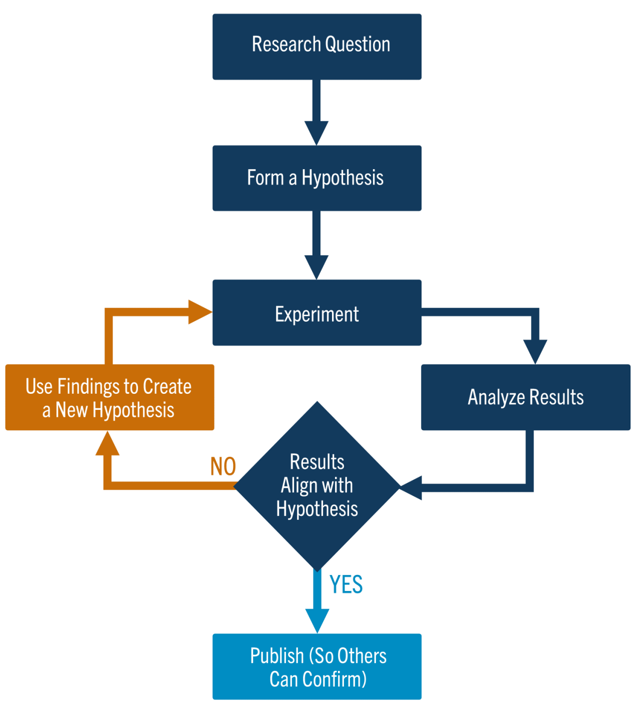

Figure 1 - Scientific Method

Another aspect of scientific progress involves scientists publishing results of their experiments, which enables the scientific community to critically analyze the experiment and its results to both ensure their validity and to serve as building blocks for subsequent research. A hallmark of S&T publishing is the concept of peer review, wherein experts in the field review draft papers before they are published to ensure scientific integrity and to determine the highest quality papers for inclusion in a scientific journal. The most respected scientific journals will only publish papers that have successfully passed rigorous peer review, but there are also lower-tier journals with noticeably lower thresholds for peer review, quality, and impact. (Unfortunately, there also seems to be a recent increase in the number of peer-reviewed papers with results that cannot be replicated, a so-called "replication crisis" within the community.)[5] There is also a growing trend of scientists sharing results online and via press releases rather than in peer-reviewed journals. Doing so enables much faster and wider dissemination of their results, which is positive, but the lack of peer review means that the quality and trustworthiness of these alternatively published papers will vary significantly.

In short, the scientific community never initially accepts a new scientific discovery or statement as factual. Indeed, the publishing of a paper with novel discoveries unleashes a torrent of additional research seeking to prove or disprove it, as well as tangential investigations to explore the conditions under which it remains accurate and what the results actually mean.

Consider a 1998 article in the medical journal *The Lancet* describing experimental results seemingly showing autism-like syndromes could be associated with the measles-mumps-rubella (MMR) vaccine. This rather shocking outcome led other scientists to seek to prove the authors' results. They quickly discovered multiple issues with the original investigation and its results—the least shocking of which was an extremely small sample size of just twelve subjects. Other issues included inaccurate medical histories on the subjects and findings that some of the supposedly autistic subjects didn't have autism at all. Over the years, numerous additional studies could not find any association between the MMR vaccine and autism.[6]

In this case, the results in the original paper were proven false—so false, in fact, that most of the co-authors of the original paper issued a letter retracting their results of the paper, and *The Lancet* publication formally retracted the paper.[7] Additional investigations into the original paper identified sloppy data practices, medical ethics issues, and failure to disclose financial conflicts of interest. The lead author's medical license was eventually revoked because of these issues. This example shows that the Scientific Method works.

There's no way that those of us not trained in the study of vaccines could analyze the numerous scientific reports and accurately understand what they were individually and collectively stating. That's actually true for most complicated and nuanced S&T topics. However, we can leverage what we know about the Scientific Method to help us understand how much trust to place in what we read.

The following chart groups message sources into seven tiers of trust. The higher the tier number, the more one should trust what is being presented.

Tier 0: Everyday social media posts	In terms of sheer numbers, these make up most modern-day scientific-sounding statements. They are also the least trustworthy–in fact, they probably should not be trusted at all. Your favorite celebrity or a friend from high school may have the best of intentions but most likely does not have the training or experience necessary for you to trust and repeat what they say. The sheer volume of these posts and velocity in which they spread unfortunately means that these snippets of information, no matter how accurate or inaccurate they may be, are what the population mostly sees.
Tier 1: Statements by outcome-oriented advocates	These entities tend not to perform research via the Scientific Method, but rather research to find snippets of information that seems to support their beliefs or desired outcomes. There may or may not be some accuracy to what they are saying, but it is almost certainly skewed.
Tier 2: Statements made by a well-intentioned professional without supporting evidence	Here, someone with scientific training discusses an issue of which they do not have direct training and experience. We've seen this extensively over the past year, with medical professionals trained to *treat individuals* offering analysis on COVID-19 policy decisions, despite not having training or experience dealing with pandemics at a national scale.
Tier 3: Paper or posting by the individual that has made a new discovery	Assuming the individual followed the Scientific Method properly, artifacts from that research should be taken seriously–for *initial consideration*, as they have not yet been proven by the scientific community. S&T papers first published on websites typically fall in this tier.
Tier 4: Peer-reviewed journal paper by an individual that has made a new discovery	Peer review takes a new scientific paper to the next step by adding a review by other experts to ensure the scientific design, analysis, and conclusions were properly developed. This step still does not guarantee all aspects of a paper are accurate–recall the prior discussion on varying thresholds of peer review and quality across different journals. Note as well that Linus Pauling received two Nobel prizes, but even he also authored technical papers that were later widely discredited. (That's how science works!)
Tier 5: Peer-reviewed papers by additional entities that have independently proven prior discoveries, or have added clarity on the circumstances or limitations of when the findings are accurate	At this point, we are starting to reach scientific consensus, and we can usually trust what is being conveyed.
Tier 6: Statements by recognized (subject-specific) experts about the community's consensus findings	These individuals have studied all available works on the topic and are providing the community's collective insights of currently proven knowledge on the subject. One example is a "consensus statement" from professional societies, governments, or international organizations that aim to communicate a summary of the science from the expert community to others.

How Is the Science Being Explained?

Now let's discuss how to understand the variety of ways even established S&T facts can be misrepresented. This happens quite often—in many cases unintentionally. But the purposeful misrepresentation of scientific facts to support influence operations[8] also seems to be growing, and entities that do so have become quite skilled at it. (Hint: Falsehoods can be crafted to be simple to understand, whereas S&T truths are often quite complicated and nuanced.)

Word and Number Confusion

The first example within this issue category is a translation issue. For many of us, reviewing a technical journal article or draft legislation is an utterly confusing endeavor. It looks somewhat like English, but a derivative form that would have developed over time by a community isolated for generations.

That's basically what happened as these specific communities developed verbiage specific to that community, even though some terms may have

THE ISSUE OF BIAS WITHIN "BIAS"

The word "bias" is loaded with different meanings that are often incorrectly intermingled in public dialogue on S&T topics. This creates misunderstandings that significantly cloud public debate. It's therefore a good example of a typical "translation issue."

Within S&T circles, "bias" is a systematic error that requires major consideration when developing experiments and when analyzing their results so we can understand what those results mean. Scientists are constantly looking to find and overcome these biases as a part of the learning process. There are many forms, including:

- Sampling bias, to ensure sample data accurately represents the range of data a system would actually encounter
- Time interval bias, caused by incorrect sampling of data (e.g., multiplying the number of emails sent between 0100 and 0200 by 24 does not provide a good estimate of how many emails are sent each day)
- Confirmation bias, which is a tendency to see information that favors what we re hoping to find rather than what is actually occurring
- Survivorship bias, where we tend to focus on results from experiments that made it all the way through while ignoring those that failed in the middle (e.g., 99% of completed experiments may be accurate, but if only 100 out of 1,000 experiments reached that point, the system isn't delivering accurate results 99% of the time)
- Omitted variable bias, where necessary information isn't included (e.g., a shiny sportscar on sale for $10k is a deal; a shiny sportscar whose engine has thrown a rod on sale for $10k is not)
- Demographic bias, where one set of individuals would receive varying outcomes based on qualities such as race, gender, or age
- Sponsor bias or scientific pride bias, where results are propped up to appear more successful than reality

Within operations, the big concern is the impact of operator bias. Everyone using the system will have their own unique background, training, and experiences (which is collectively that operator s bias) impacting their decision-making. These biases occur in all applications, with all users. This can lead to differing system results, even with identical inputs. (e.g. A security guard on their first shift will investigate a motion sensor alarm more studiously than their experienced peer who knows the neighbor puts out their cat at that time every night.)

Within everyday public conversations, "bias" typically refers to a purposeful inclination or prejudice in favor (or against) a person or group, usually in a way that is closeminded or unfair.

Each individual's experiences will dictate which meaning of bias they automatically consider when they read the word, and that meaning may or may not correctly align with the author's intended usage. Technical reports that discuss scientific and operational biases are regularly, and incorrectly, described in news articles and social media posts as having shown evidence of prejudicial biases.

Note that we each also have trust biases that impact our decision-making. Most of us trust GPS navigation systems, even though they don't always guide us optimally. If we don't like a politician, scientist, or news personality, we tend to not trust what they say (and vice versa). Sometimes we even trust those we feel close to rather than experts, such as grandma's home remedy vs. a new-to-us doctor's recommendation.

different (or multiple) meanings externally. It is rare to find individuals steeped in both scientific and policy jargon that can accurately translate between the two, and much rarer for someone to be able to explain both to the general population. Incorrect translations lead to inaccurate understandings of S&T findings, which in turn produce both a diminishing of trust in the findings and nonproductive policy debates.

For example, scientists are trained to incorporate uncertainties in their writing. Consider this sentence from a few paragraphs ago: "Over the years numerous additional studies could not find any association between the MMR vaccine and autism." To those that understand how science works, this is a very definitive statement: There is no evidence linking the two (but recognize that it is always possible for new findings to be discovered). To the public and policymaking community, who desire clear-cut statements such as "the MMR vaccine does not cause autism," the considered sentence could be viewed as sufficiently squishy for them to believe that scientists don't really know.

Studies of scientific papers have also found that somewhere between 10 and 20 percent of citations are used to support claims that conflict with the findings of the original paper.[9] While unintentional, these mistakes perpetuate false understandings of S&T findings with an air of scientific propriety.

No one likes to look at pages of raw data, so placing that data into a chart makes it easy for readers to quickly understand what the data is telling us. However, it is also quite easy for individuals to play with the charts, while still using accurate data, to craft vastly different stories about what the data tells us. Common issues include only presenting desirable snippets of the data, playing with the scale or range of the chart axes (for example, not starting at zero or zooming so far out that you can't see any changes), or using different axes for different data elements on the same chart. Everyone must pay close attention to chart elements to avoid being duped; studious analysts always need to go back and study the original data itself.

A somewhat related chart and graph issue is improperly intermingling correlation (there's an overlapping pattern) with causation (one event causes another). To better understand, visit this humorous website that shows visible data correlations that are obviously not causations: https://www.tylervigen.com/spurious-correlations.

Lastly, S&T capabilities that are obviously different to subject matter experts may seem the same to the rest of us. This easily leads to the exchange of misinformation and faulty analysis. Recent examples include COVID-19 masks (not all mask types have the same protective properties) and the confusion between issues associated with facial analytics tasks (which analyze a facial photo to estimate the subject's age, gender, and ethnicity, or to identify medical conditions) and face recognition tasks (a biometric to confirm or determine identities).

Issues in Context

Context is also important in scientific discussions. Recall from earlier that S&T truths are often quite complicated and nuanced. What is true for one situation may be completely inaccurate for a second situation, even if the two are very similar. This occurs often in operational contexts, where even a slight adjustment in operational parameters can produce wildly different outcomes or issues.

There is also a vast difference between performance matters of a specific S&T capability in a laboratory experiment and an operational system that leverages that capability. In most cases, taking a S&T capability out of a highly controlled environment and using it "in the wild" produces much different (often worse) results. But it's also possible for concerning results found in a laboratory experiment (such as a demographic bias) to not manifest in operations as the S&T capability (tested individually in the lab) is one of several components within a fielded system, with the collection of components designed to work together to minimize those concerning aspects from appearing in the overall system's output.[10] Analyses that fail to recognize these variances often produce inaccurate conclusions.

S&T EVALUATIONS TYPICALLY FALL INTO ONE OF THREE TYPES:

Technology Evaluation, which tests a core S&T component in a controlled manner using common data. Results are quite useful to researchers as they seek to improve the component. With learned analysis, these results can directionally point to potential operational considerations, but do not prove (or disprove) their existence.

Scenario Evaluation, which tests a system that uses the S&T component in a mock environment designed to match its operational use. Results are useful to operational planners as they showcase accuracy and anticipated issues for that specific operational application.

Operational Evaluation, which tests a system as it is being used in practice. What can be tested in this type of evaluation is often limited (due to lack of ground truth within the experimental protocol), but what is tested provides the clearest indicator of system performance for that specific application.

We also must be aware that our view of the usefulness and issues associated with an S&T capability can vary based on the objectives and alternative approaches we compare it against. The fields of artificial intelligence and automation struggle with this considerably, with errors brightly highlighted as reasons for them to not be used. From this perspective, based on desiring 100% accuracy before being used, this argument makes perfect sense. On the other hand, if one instead compares the new capability to the existing approach (such as human performance), the new tools usually reduce errors considerably, while also performing their tasks faster and more cheaply. From this second perspective, an argument to implement them makes perfect sense. We must ensure we're making the proper comparison for each individual application before making any decision.

The nation's lack of probability and statistics training in high schools also contributes to these phenomena, as does a natural human difficulty of grasping scales of numbers that we don't normally run across (try conceptualizing what $1 trillion really is!). At the time of this writing, the Centers for Disease Control and Prevention (CDC) placed a pause on one of the COVID-19 vaccines because six women developed a blood-clotting disorder after being vaccinated. People are understandably concerned and the CDC's actions to ensure safety are commendable, but we must also recognize that six adverse reactions out of nearly seven million vaccinations is not a large number—the odds of randomly dying on a given day are much higher.

Nonscientific Objectives

The scientific information that news sources provide us are usually more sensationalized and one-sided than what the scientific community is focusing on, leaving us with distorted views of the state of the science. Journalists like to focus on what is new or what contradicts existing thinking, so naturally they tend to focus on sources that are lower on the Tiers of Trust chart. Their editors further sensationalize the title to generate mouse clicks, sometimes messing up the message in the process. Social media algorithms are even more focused on generating mouse clicks, and they know the best way to do that is to feed us stories that are similar to others we previously interacted with—so we end up with a stream of increasingly inaccurate messages rather than the variety of nuanced explanations we require. And there's certainly little, if any, effort to recant stories that we've previously read when contradictory studies become available!

Far too commonly, outcome-focused advocacy entities cherry-pick snippets of scientific data, remove the contexts in which those results apply, and then merge them together to paint their desired picture. These arguments often seem compelling but in reality are often rife with misinformation.

Anything that seems to prove an author's desired outcome by only using snippets of scientific "evidence" should be viewed with a non-trusting eye.

Proper analyses always need to start with explaining the S&T experiment and outcomes themselves, then determining their meaning and the contexts in which they apply. It is only with this background, properly developed, that an analysis can begin to be considered accurate.

> **Any idiot can create more bullshit than you could ever hope to refute.**
>
> **The amount of energy needed to refute bullshit is an order of magnitude bigger than what is needed to produce it.**
>
> Carl Bergstrom & Jevin West, *Calling Bullshit: The Art of Skepticism*

Consider the prior example regarding the MMR vaccine and its alleged correlation with autism. Even though the experiment and resulting analysis within the original paper was completely discredited by the scientific community, those with an innate stance against vaccinations latched onto the paper, widely promoting it as evidence not to trust vaccines. They even dismissed the scientific community's subsequent experiments to prove or disprove the original paper, a required step within the Scientific Method, as a Big Brother-type conspiracy! The disproven message that vaccines cause autism unfortunately continues today, spread by misinformed individuals, bots, and Russian trolls promoting discord in the United States[11]—along with a measurable increase in vaccine-preventable diseases as more individuals refuse to accept vaccinations.

Upon coming across an analysis of an S&T topic, readers must first ask themselves, "Who is the author, and what is their objective?" In other words, does the author seem to be driven toward a predetermined outcome, or are they genuinely attempting to explain a result as unbiasedly as they can? Most of the time the former should be ignored, and you should review the latter with a critical eye to ensure they've started with an analysis of the data and experimental protocol so they understand the context of the results.

Conclusion

This paper provides mental tools for policymakers and other nonscientists to leverage as they critically analyze what they read by:

- Understanding how scientific knowledge progresses and consensus is achieved,
- Leveraging the Scientific Method to assess the trustworthiness of scientific information, and
- Identifying ways that even properly developed scientific facts can be misrepresented.

Continued advancement of S&T is a bedrock of our nation's security and prosperity, and the source of many niceties we enjoy on a daily basis. As important, scientific facts can be invaluable to policymakers and the public as they shape and assess the direction of our nation. But because of the difficulty in understanding scientific-sounding facts when the domain is outside our areas of expertise, those benefits can be greatly delayed—or worse, never come to pass at all.

About the Author

Duane Blackburn leads science and technology policy for MITRE's Center for Data-Driven Policy, which brings objective, evidence-based, nonpartisan insights to government policymaking. Mr. Blackburn previously served for eight years (across two administrations) in the White House Office of Science and Technology Policy (OSTP).

For more information about this paper or the Center for Data-Driven Policy, contact policy@mitre.org.

[1]Cary Funk and Brian Kennedy. Public confidence in scientists has remained stable for decades. 2020. Pew Research Center, https://www.pewresearch.org/fact-tank/2020/08/27/public-confidence-in-scientists-has-remained-stable-for-decades/. Accessed April 13, 2021.

[2]Ryan Cross. Will public trust in science survive the pandemic? 2021, Chemical & Engineering News, https://cen.acs.org/policy/global-health/Will-public-trust-in-science-survive-the-pandemic/99/i3. Accessed April 13, 2021.

[3]This presents a simplified version of the Scientific Method and the development of scientific understanding, focusing on the experimentation stage. Theoretical thinking and "thought experiments" are critical earlier stages in scientific progress, though are not as often seen and discussed by the general public.

[4]Technically, it is very difficult to "prove" a hypothesis, as it can only take one new observation to the contrary to invalidate the initial claim. Instead, scientists use confidence measures to ensure a collection of results meets a statistical significance threshold. For example, in nuclear physics this threshold is 5 sigma (approximately 1 in 3 million); in social sciences there is typically a 5% threshold to ensure results aren't due to random chance.

[5]Jevin West and Carl Bergstrom. Misinformation in and about science. 2021. Proceedings of the National Academy of Sciences of the United States of America, https://www.pnas.org/content/118/15/e1912444117. Accessed April 13, 2021.

[6]Steve Calandrillo. Vanishing Vaccinations: Why Are So Many Americans Opting Out of Vaccinating Their Children? 2004. University of Michigan Journal of Law Reform, https://repository.law.umich.edu/cgi/viewcontent.cgi?article=1378&context=mjlr. Accessed April 13, 2021

[7]Laura Eggertson. Lancet retracts 12-year-old article linking autism to MMR vaccines. 2010. CMAJ, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831678/. Accessed April 13, 2021.

[8]Influence operations are activities designed to foster attitudes, behaviors, or decisions of an audience on a specific topic. The typical "burden of proof" that drives the Scientific Method and personal relationships is not a consideration within these activities.

[9]Laura Eggertson. Lancet retracts 12-year-old article linking autism to MMR vaccines. 2010. CMAJ, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831678/. Accessed April 13, 2021.

[10]This is common in "Emergent Systems," where a system's behavior and outcome are a consequence of the interactions and relationships among its components, rather than the independent behavior of individual elements. Evaluating such a system's operational performance thus requires an end-to-end analysis; evaluating only one component of an emergent system and assuming those results would align with the overall system's results produces incorrect analyses.

[11]David Broniatowski, et al. Weaponized health communication: twitter bots and russian trolls amplify the vaccine debate. 2018. American Journal of Public Health, https://ajph.aphapublications.org/doi/10.2105/AJPH.2018.304567. Accessed April 13, 2021.