Thinking About Thinking: Exploring Bias in Cybersecurity with Insights from Cognitive Science

FORCEPOINT

Table of Contents

Exploring Bias in Cybersecurity

Imagine today's news is full of the latest privacy breach executed by foreign hackers, unleashing social security numbers to passwords for hundreds of thousands of individuals. When evaluating new cybersecurity threats your company should counter, these news reports could drive you or your leaders to focus on outside attacks. But, if you dig deeper, you may find that these reports are really about a different industry, and the biggest threat to your organization could be better countered by focusing on the behavior of your own employees. Relying on what's top of mind is a common human decision-making tool, but can lead to faulty conclusions.

When situations are less than clear cut, our initial reactions and decisions can be driven by unconscious biases like the "availability bias" described above. In Part II of our series on cognitive science in cybersecurity, we're taking a closer look at how human cognitive biases and reasoning errors impact decisions and business outcomes in information security. Whereas Part I of this series, *Exploring the Gray Space of Cybersecurity with Insights from Cognitive Science*, discussed specific strengths of human cognition and development that could be leveraged to help technology deal with ambiguity, Part II draws attention to the predictable and, at times, preventable errors linked to human susceptibility to cognitive bias.

By improving our understanding of biases, it becomes easier to identify and mitigate the impact of flawed reasoning and decision-making conventions. Our efforts to build harmony between the best characteristics of humans and the best characteristics of technology to tackle cybersecurity challenges depend on understanding and overcoming bias.

Our efforts to build harmony between the best characteristics of humans and the best characteristics of technology to tackle cybersecurity challenges depend on understanding and overcoming bias.

The Psychology of Bias

Bias is the tendency for people to favor one group, person, or thing over another, while unfairly discriminating against the remainder of the choices. For better or worse, bias is an inescapable feature of the human experience. We are shaped by a combination of our environment, our genetics, and our cognitive ability to process and make sense of our world. This means that our decisions, behaviors, and experiences are influenced by the experiences of the past and the present.

To understand bias, it is helpful to categorize human thought into a framework, called the Dual Process Theory. Dual Process Theory splits human cognition into two modes:[1]

System 1	System 2
Intuition	Reasoning
Automatic	Effortful
Implicit	Explicit
Fast	Slow
Metaphorical	Exact

1 For a full overview of Dual Processing Theory (System 1/System 2) and behavioral economics (including bias), refer to the work of Daniel Kahneman (academic articles, or for an accessible book, refer to "Thinking, Fast & Slow")

Both System 1 and System 2 are required to keep humans running smoothly through their lives. Just as our automatic processes allow us to tie our shoelaces without thinking about it, our effortful processes allow us to systematically think through various pros and cons associated with difficult career or financial decisions.

An exceptional human trait is that we are able to think about thinking, which means that we have the ability to consciously switch from System 1 thinking to System 2 thinking.

Take a look at Figure 1. On the top, the two horizontal lines look like they are different lengths.[2] On the bottom, the image appears to show equilateral triangles.[3] System 1 is responsible for your initial perception; your ability to automatically use contextual cues to estimate the sizes of objects, and your ability to "fill in the blanks" by establishing patterns.

However, we can also engage System 2 when looking at these images. If you measure the two horizontal lines from Image A, you'll see and logically understand that the lines are the same length, but this won't necessarily stop you from perceiving them as two different lengths. When you take a closer look at Image B, you'll notice that none of the shapes are actually triangles, but you will continue to see triangles in the image. Ultimately, we are not able to block these perceptual illusions from occurring. This is not a problem in situations where the illusion has no impact on our performance, or on our decisions. However, when faced with a critical decision, depending on faulty impressions or gut feelings can result in errors in reasoning and poor decision-making.

The concept that people engage in different types of thinking is not new. The System 1 and System 2 paradigm aligns with psychological theories that pre-date our current knowledge of cognitive and neuropsychology. Sigmund Freud, for instance, believed that all human behavior was driven by unconscious urges and that conscious, observable human behavior represented an extremely small fraction of our individual identities. Freud believed that conscious reality, or the world that we are actively aware of, makes up a very small piece of human existence. Rather, our unconscious mind has a much larger impact over our experiences and our behavior—even if we aren't aware of what is occurring beneath the surface. While Freudian theories may not be quite as popular as they once were, we do know that people spend an overwhelming percentage of their life guided by and engaged in automatic thinking.

People spend the vast majority of their life immersed in System 1 thinking because brains are built for efficiency. Brains require approximately 20% of the human body's energy,[4] even at rest, which creates a need to prioritize saving mental time and energy over engaging in resource-heavy analytic thought. Psychologists often refer to our natural inclination towards conserving mental energy as being "cognitive misers." Misers avoid spending their assets, and similarly, humans avoid spending mental effort. The major difference is that financial misers conserve resources on purpose, but cognitive misers conserve resources subconsciously. In most cases,

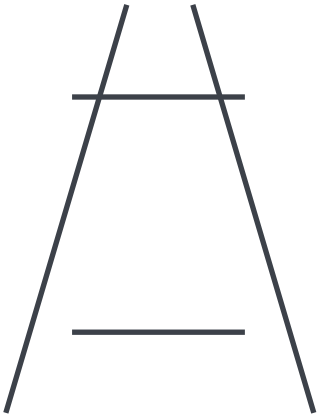


Image A

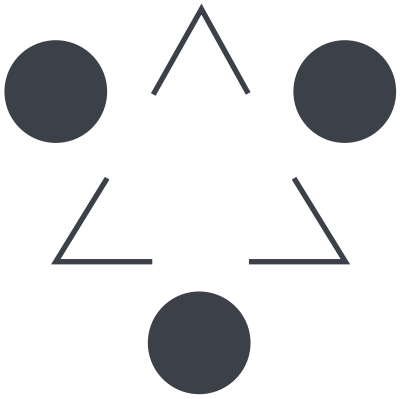


Image B

Figure 1. Image A is a geometrical-optical illusion known as the Ponzo Illusion. Image B is the Kanizsa Triangle, demonstrating the concept of modal completion.

2 Ponzo illusion. Mario Ponzo, 1911.
3 Kanizsa Triangle Illusion. Gaetano Kanizsa, 1955.
4 Richardson, M. W. (2019). How much energy does the brain use? (https://www.brainfacts.org/Brain-Anatomy-and-Function/Anatomy/2019/How-Much-Energy-Does-the-Brain-Use-020119)

being a cognitive miser is an advantage. If we constantly analyzed each detail in our environment, we would not have the energy to engage in higher-level System 2 executive functions such as impulse control, planning, and deliberate reasoning. On the other hand, being a cognitive miser sometimes results in bias, or in making incorrect decisions based on mental shortcuts called heuristics.

Consider the following question:[5]

> **Jack is looking at Anne, but Anne is looking at George. Jack is married, but George is not. Is a married person looking at an unmarried person?**
>
> **a) Yes**
> **b) No**
> **c) Cannot be determined**

Responses to this question vary, but up to 80% of respondents will select "C." However, the correct answer is A.

When taking a deeper look at the options, you can see that it does not matter whether or not we know if Anne is married. If she isn't married, then Jack, a married person, is looking at Anne, an unmarried person. If she is married, then Anne, a married person, is looking at George, an unmarried person.

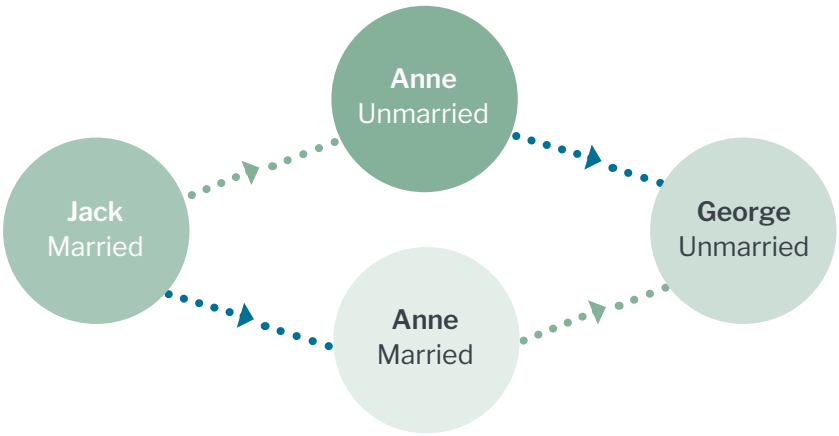


Figure 2. Whether Anne is married or unmarried, among the group, a married person is looking at an unmarried person.

Thinking through the possible options on marital status and directional gaze takes more effort than quickly identifying that Anne does not have a marital status. The missing information about Anne's marital status quickly registers as "missing information" for many readers, and the miserly mind connects the missing information to the "cannot be determined" answer. If you answered this question correctly, it is possible that due to the context of this paper you assumed that the question would be tricky, and therefore engaged in purposeful critical thinking. Bias lurks within System 1, as snap judgments, stereotyping, and rules of thumb allow us to take shortcuts to conserve mental energy—just like the shortcut that many readers made when answering the question above.

> Psychologists often refer to our natural inclination towards conserving mental energy as being "cognitive misers."

5 Hector Levasque, as cited by Keith Stanovich, "Rational and Irrational Thought: The thinking that IQ Tests Miss"

Six Biases Skewing your Security Strategies

Sometimes people are aware of the biases that they have. For instance, maybe they are aware that they only read news written from one political viewpoint—even though they know they'd have a more balanced perception of real-world events if they diversified their news sources. However, in many situations people are unaware of their own biases and how such biases impact their decisions and actions. This is because biases dwell beneath the surface of our awareness as automatic thought processes. Building awareness of cognitive biases can help us move beyond biased decision making, and more importantly, help us avoid designing systems that perpetuate our own biases in technology. To achieve this type of awareness, we have to challenge ourselves to think about thinking. Thinking about, and understanding, how we think and reason is especially beneficial when we identify situations where bias is likely to have a significant negative impact on our choices or behaviors.

In cybersecurity, understanding and overcoming security-related perceptual and decision-making biases is critical, as biases impact resource allocation and threat analysis. The following represent a small subset of known decision-making biases that are meaningful to cybersecurity professionals. They are described and contextualized in an effort to raise awareness of how each bias functions, and how biases can impact our understanding of the cyber landscape, our perception of risks, and ultimately, our perceptions about each other.

Building awareness of cognitive biases can help us move beyond biased decision making, and more importantly, help us avoid designing systems that perpetuate our own biases in technology.

Aggregate Bias

Aggregate bias, or ecological inference fallacies, happens when we infer something about an individual using data that describes trends for the broader population. This results in bias because information used to understand groups of people cannot be assumed to be accurate at the individual level, as individuals often have many other confounding variables that impact their behavior.

For example, older people are frequently characterized as riskier users based on their supposed lack of familiarity with new technologies. However, this characterization is not necessarily true at the individual level. For example, recent studies show that older adults are less likely to share passwords than younger people. Approximately 35% of millennials (ages from the mid-20s to mid-30s) share passwords for streaming services like Netflix and Hulu, with 19% of Generation X users sharing passwords, and only 13% of Baby Boomers sharing passwords. While there may be some differences in the prevalence of using streaming services across these age groups, the younger generation's willingness to share sensitive information, such as passwords, is far riskier than older users' habits.

This is especially important, as additional research shows that people frequently reuse identical passwords across domains (up to 40%) and when considering partial password reuse, the number jumps to 80-90%.[6] This means that when a person shares one password with a friend, for something as seemingly innocuous as a video streaming service, the person may in fact be sharing their banking password.

6 https://www.forcepoint.com/blog/security-labs/passwords-passphrases-or-%E2%80%9Ci%E2%80%99ll-pass%E2%80%9D-nist%E2%80%99s-digital-identity-guidelines

While it is not possible to know the actual rate of password reuse, especially for cross-sectional domains such as streaming services versus banking services, the willingness of younger adults to share their passwords at a much higher rate illustrates a potential misconception about which users of technology are riskiest. The trope of making sure our grandparents do not send money to a Nigerian prince may be far less important to our overarching security than identifying ways to decrease credential sharing among a younger generation that perceives account details and privacy through a different (and seemingly more lenient) lens.

Aggregate bias can also impact security investigations, in which an analyst wrongly focuses on an individual due to the individual's group membership (e.g., highly technical person with a lot of access) rather than the facts or forensic information that accurately describes the individual and their behavior. Focusing on an individual due to a misapplication of characteristics can prompt analysts to fish for answers and reasons to support their assumptions, which can delay identification of the true source of security issues.

Overcoming aggregate bias through understanding of individual human behavior is critical to security solutions that want to address human error and/or human risk factors in protecting data. To achieve this goal, and to move beyond attributing or misattributing behavioral characteristics to individuals, advanced behavioral analytics that allow for self-to-self, self-to-peer, and self-to-global comparisons can help provide context for understanding complex individual behaviors.

Overcoming aggregate bias through understanding of individual human behavior is critical to security solutions that want to address human error and/or human risk factors in protecting data.

Anchoring Bias

Anchoring occurs when a person locks onto a specific salient feature or set of features of information early in the decision-making process. This frequently occurs with numbers, such as in sales, when one party in a negotiation proposes a price point. Once a price point is set, the number serves as an anchor for additional negotiations (which may be too high, too low, or even accurate).

Anchoring is simple to demonstrate if you're willing do a little experiment. Ask two separate groups of people (say, five or six people per group) to estimate the number of phishing attempts a large company copes with on a weekly basis.

- Ask the first group, how many phishing attempts do you think we get every week, 3,000?
- Ask the second group, how many phishing attempts do you think we get every week, 300,000?

You will likely find that the estimates for the group with the lower anchor (3,000) are much lower than the estimates from the group with the higher anchor (300,000). Of course, if your groups are comprised of people who deal with phishing in a professional capacity, their answers may be anchored by their real-life experience!

At an individual level, anchoring influences cybersecurity when an analyst latches on to a specific value during early phases of detection or investigation, and then fails to move away from the preliminary "anchor" even when the solution to the issue requires a complete deviation from the initial salient information point. When analysts' attention is drawn to a specific feature, they may miss or erroneously

discount other influential information associated with the threat. At a broader level, if a high-level person within an organization such as a CISO provides information about potential threats or quantifies the potential impact of a threat, the CISO's words prime and anchor employees lower on the organizational chart to focus (sometimes incorrectly) on specific threats.

Overcoming anchoring is particularly challenging, as awareness of the anchoring phenomenon does not necessarily negate its effects. Recalculating estimates is not something that humans are particularly good at, especially when there are multiple (or complex) factors at play. This is one specific type of bias where humans can, or should, depend more heavily on statistical analysis techniques that can decrease the impact of overly weighted early judgments in favor of balancing the impact of new and critical information into their decision paradigm.

Availability Bias

Memory plays a large role in availability bias. The more frequently a person encounters specific types of information, the more readily accessible the information is in their memory. The availability of information, where certain types of information are encountered more frequently, can impact how humans perceive how likely an event is to occur (Figure 3).

In information security, news cycles that focus on ransomware or specific types of threats can influence analysts' perceptions of what's risky and can influence their approaches towards security. Hot topics that continuously rise to the top of an analyst's mind can be distracting and bias the diagnosis of system issues that are deemed "less likely" simply because they are less available (i.e., less active) in the analyst's memory. At an organizational level, availability bias can influence the allocation of resources and can lead to a misinterpretation of risk.

Luckily, analysts are in the business of thoroughly exploring data. While they are, at times, susceptible to overestimating the probability of an event occurring, their job is to consistently challenge their reasoning strategies and to consistently seek the unexpected. Organizationally, leadership—who may be more or less technical, and more or less privy to in-depth information that analysts see on a daily basis—is more likely to be swayed by availability bias. This means that organizational cultures that undervalue, or ignore, data that accurately represents the probability of specific types of threat events may seek out or invest in solutions that are built to cope with problems that carry very low likelihood of occurring.

Understanding the probabilities in the information security threat landscape requires working with our technology, and our data, to better and more accurately represent the state of the threat world so that decisions are not made based on news cycles that potentially inflate or misrepresent the probability of certain types of threats. This means that coping with availability bias requires both humans and technology. Humans are required to create an organizational culture and communication strategy that values the expertise of security personnel. In addition, technology can help provide more accurate probabilities of various types of threats.

> News cycles that focus on ransomware or specific types of threats can influence analysts' perceptions of what is risky.

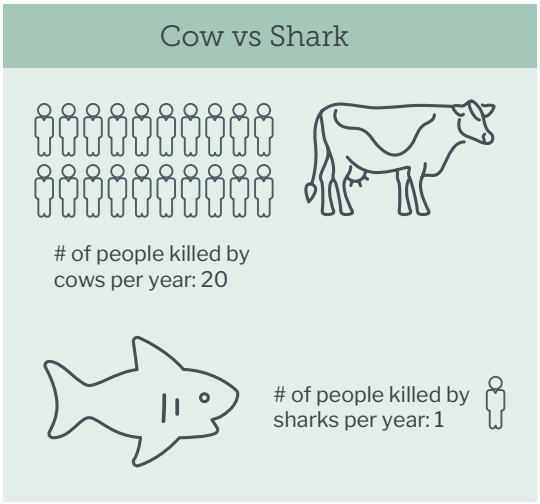


Figure 3. While shark attacks may garner more press coverage, cows actually cause more deaths per year.

Confirmation Bias

In the age of big data, we can almost always find data to support our opinions and ideas. It is often possible to support multiple theories regarding why an incident occurred or what type of risk is present on a network. For example, if you had to argue that the earth is flat, you could find plenty of information to support that claim online. Alternatively, you can find plenty of information to support the claim that the earth is round. When people have a theory to explore when trying to answer a question or support their opinion, they are highly susceptible to confirming their beliefs by searching for (and often finding) support for their hunch. Confirming our own beliefs by searching for and building information around our arguments, while excluding or deemphasizing opposing viewpoints, is called confirmation bias. Confirmation bias not only affects our reasoning strategies, but it also impacts our memory of information. People tend to focus on and remember information that confirms or aligns with their beliefs, while discounting or forgetting information that opposes their viewpoint.

Analysts, with the best of intentions, may find themselves spending a lot of time looking for causes or issues associated with an adverse event by only searching for causes or issues that align with their personal theories or insight. This is particularly relevant for experienced analysts who may "decide" what happened prior to investigating an event. Their expertise and experience, while extraordinarily valuable, can be a weakness if they investigate incidents in a way that only supports their existing belief.

Overcoming confirmation bias requires creative and flexible thinking—in particular, the ability and willingness to look at a situation from different points of view. A company that fosters relationships and teams that are comfortable with pushing each other's beliefs is critical. People, and technology, can facilitate mental exercises such as thinking backwards, role playing, devil's advocacy, and learning from surprising events. [7]

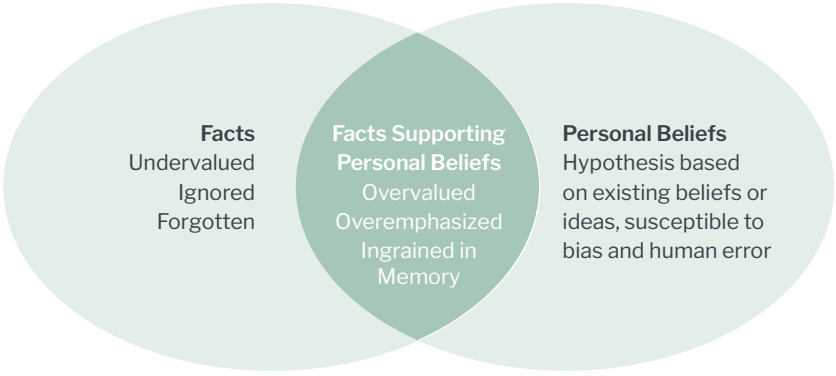


Figure 4. People tend to focus on and remember information that confirms or aligns with their beliefs, while discounting or forgetting information that opposes their viewpoint.

7 For more information on mental toolkits and psychological factors associated with intelligence analysis, see Richards J. Heuer, Jr. (1999). Psychology of Intelligence Analysis.

The Framing Effect

Another factor that can impact how people make choices is how those choices are worded. People typically prefer knowing that an outcome is a "sure thing" rather than knowing that an outcome has a certain probability of occurring. For example, if you ask someone whether they'd prefer a guaranteed $100, or a 15% chance of getting $1,000, many people will choose the guaranteed $100.

While there are individual differences in how people answer this question (we've all seen the game show participants who are willing to "risk it all" for one more chance to win big), the "sure thing" is often perceived as the best option when it comes to choices associated with gains. When we think about what people choose, and we're talking about a positive outcome, people tend to make the less risky choice by choosing the sure thing.

But what happens when the decision, and the choices associated with the decision, aren't associated with something positive like winning money? What if the choices are presented in a way that highlights a loss, or the chance for a loss? Let's revisit the money question: would you rather definitely lose $100, or have a 15% chance of losing $1,000?

What we see (of course, not perfectly reversed due to those pesky individual differences) is that people are more willing to take the probabilistic (riskier) option when they are faced with a loss.

Security problems are often aggressively worded, and use negative framing strategies to emphasize the potential for loss. This strategy prompts security decision makers to, at times, invest in security solutions that are expensive (or overkill!) to address overly specific and low-probability risk factors. Consider the vendor who promotes that "one out of five small companies never got their data back after a ransomware attack!" The focus on the one company that didn't get data back versus the four that did over-emphasizes the risk for companies.

Companies are unlikely to abandon effective marketing strategies, especially those that benefit from strategic message-framing techniques. However, buyers of security solutions can overcome the impact of framing effects by slowing down and thinking more analytically about the problems they are trying to solve, and the suggested efficacy of the solutions offered. Framing effects are somewhat fragile, and their impact depends on the one-sided nature of the phrasing of a question. When a person knows that they need to pay attention to how questions are phrased, they can overcome their initial knee-jerk reactions for making a choice—reactions that could result in decisions that are far too risky, or far too conservative.

Buyers of security solutions can overcome the impact of framing effects by slowing down and thinking more analytically about the problems they are trying to solve, and the suggested efficacy of the solutions offered.

Fundamental Attribution Error

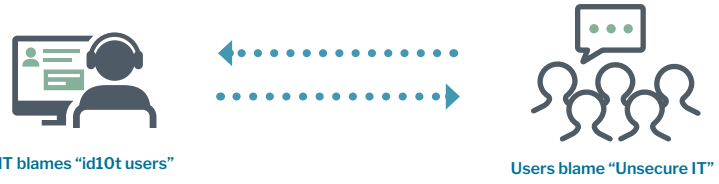
One of the most interesting social and psychological biases that impacts nearly every aspect of human behavior is the fundamental attribution error. This is the tendency to see other people's failures or mistakes as part of their *identity* rather than attributing the failure or mistake to contextual or environmental influences. The most basic example of this type of error is when a person sees another person trip. The observer may think, "Wow, what a clumsy person!" without realizing that there were contextual factors at play, such as an uneven sidewalk. The other side of

the fundamental attribution error is called the self-serving bias, where the individual making the mistake places the "blame" on environmental or contextual factors rather than internalizing the mistake as an internal trait.

Fundamental attribution error impacts multiple areas of cybersecurity. Security analysts and software developers, for instance, often joke about PEBKAC (Problem Exists Between Keyboard and Chair), and "id10t" users creating risks or having issues with technology. Characterizing end users as less capable, less intelligent, and prone to making mistakes out of laziness is a serious form of fundamental attribution error. Shifting of blame in both directions, from IT and engineering to end users, and from end users to IT and engineering, embodies the impact of this bi-directional social bias. For people with high levels of technical expertise, self-serving bias also emerges when they do not recognize their own risky behaviors—or even go so far as to "excuse" their behavior due to their self-perceived technical knowledge and abilities.

Coping with fundamental attribution errors, and the self-serving bias, requires personal insight and empathy. It can be extraordinarily difficult to engage in consistent self-assessment to determine when we may be placing blame on a person rather than blame on environmental factors that impacted a person's behavior. It is also difficult to acknowledge when we are responsible, due to our own shortcomings, for adverse events or outcomes. What we can do is practice empathy and build our capacity for giving others the benefit of the doubt. For supervisors and leaders, acknowledging imperfections/failures can help create a more resilient and dynamic culture. For the people designing complex software architectures, consider that your perspective is highly security focused—while your users' motivations may not be—and that their failures are not because they are stupid, but because they're human.

Blame for a Breach



Overcoming Bias with Applied Insight

Coping with and overcoming bias whenever possible to facilitate better decision making requires that we understand that answers fall somewhere in the middle, within the gray space. Human weaknesses and cognitive shortcuts that result in bias require us to foster a sense of intrinsic motivation to address bias, while requiring us to turn towards one another and towards technology to minimize the impact of predictable biases in the cybersecurity community.

Of the biases outlined in this paper, several of them can be addressed directly through the use of improved advanced analytics. The prime example is aggregate bias. As we develop the capabilities to understand individual human behavior, rather than group human behavior, we can get much better at applying policies, rules, and constraints on those individuals who push the boundaries of risky behavior or on those individuals who have the greatest negative security impacts on an organization. The ability to do this, without the application of broad or inflexible rules and restrictions generated for a specific group (say, for older adults or for engineers who create and edit source code), can promote a more resilient workforce that is able to work efficiently and effectively with fewer security-induced roadblocks. Decreasing frustration and friction associated with security protocols is critical, and by understanding individual behavior through advanced behavioral analytics, we are getting closer to an adaptive security framework that benefits users, organizations, and security professionals.

However, there are other biases that require a far more human approach or that do not have an obvious technology-based strategy. One bias that requires human effort is overcoming the impact of the fundamental attribution error. While organizations can raise awareness of this phenomenon, individuals within an organization must take on the responsibility for challenging their own assumptions about themselves and about others. That said, when creating new technologies, use of design thinking techniques and working towards integrating human-centered design methods can help.

As a security professional, take a few moments to walk through the six biases described in this paper:

1. Do you or your colleagues make assumptions about individuals but use group characteristics to form your assumptions?
2. Have you ever been hung up on a forensic detail that you struggled to move away from to identify a new path for exploration?
3. Has the recent news cycle swayed your company's perception of current risks?
4. When you run into the same problem over and over again, do you slow down to think about other possible solutions or answers?
5. When offered new services and products, do you assess the risk (and your risk tolerance) in a balanced way? From multiple perspectives?
6. And finally, does your team take steps to recognize your own responsibility for errors or for engaging in risky behaviors, and give credit to others who may have made an error due to environmental factors?

After taking the time to review our experiences, professional environment, and decision-making habits, we'll all likely find that some of these biases impact us, our teams, or our companies more heavily than others. It's critical, even in today's environment of never-ending alerts and dangers, that cybersecurity teams and professionals slow down and think more deeply and strategically in order to combat these biases. If not, we may find that biases are blinding us to the real threats.

About the Author

Dr. Margaret Cunningham is a Principal Research Scientist for Human Behavior in Forcepoint's X-Labs. Her current passion is identifying critical interactions between humans and technology through a socio-technical systems lens, with a goal of establishing a better human-centric model for improving cybersecurity.

Previously, Margaret supported technology acquisition and integration, research and development, and operational testing and evaluation for the Department of Homeland Security, the U.S. Coast Guard, and other government clients. She also has experience in the healthcare domain, particularly in building metrics that illustrate how integrating new technologies can impact patient safety and care provider performance.

About Forcepoint

Forcepoint is transforming cybersecurity by focusing on what matters most: people's behavior as they interact with critical data and systems. This human-centric approach to cybersecurity frees employees to innovate by understanding the normal rhythm of user behavior and the flow of data in and out of an organization. Forcepoint behavior-based solutions adapt to risk in real time and are delivered via a converged security platform to protect network users and cloud access, prevent confidential data from leaving the corporate network, and eliminate breaches caused by insiders. Based in Austin, Texas, Forcepoint protects the human point for thousands of enterprise and government customers in more than 150 countries.

forcepoint.com/contact

[BIAS IN CYBERSECURITY-GLOBAL-REPORT-US-EN] 200079.052119