Science of Security: Combining Theory and Measurement to Reflect the Observable

Cormac Herley
Microsoft Research, Redmond, WA, USA
cormac@microsoft.com

P.C. van Oorschot
Carleton University, Ottawa, ON, Canada
paulv@scs.carleton.ca

***Abstract*—What would a "Science of Security" look like? This question has received considerable attention over the past ten years. No one argues against the desirability of making security research more "scientific". But how would one would go about that? We argue that making progress on this requires clarifying what "scientific" means in the context of computer security, and that has received too little attention. We pursue this based on a review of literature in the history and philosophy of science and a belief that work under the theme "Science and Security" should align with and ideally, benefit from what has been learned over a few hundred years in science. We offer observations and insights, with a view that the security community can benefit from better leveraging past lessons and common practices well-accepted by consensus in the mainstream scientific community—but which appear little recognized in the security community.**

I. Introduction

The past decade has seen a significant effort to develop a Science of Security (SoS).[1] The US government sponsored a study to identify fundamental principles, scientific methods and whatever else might facilitate a more scientific approach to security research towards creating a "science of cyber-security". The resulting 2010 JASON report [2] mainly discussed formal approaches, but also highlighted an under-development of empirical work. Other government-led initiatives promoting Science in Security included a kick-off workshop in 2008 sponsored by the NSA, NSF and IARPA [3]; an NSA-sponsored Science of Security best paper prize since 2013; funded "lablets" on this theme at four universities and related workshops in 2012, 2014-2017 (originally a Science of Security Community Meeting, now called HotSoS); a workshop series called *LASER* (Learning from Authoritative Security Experiment Results) conceived in 2011, partially NSF-funded, and focused on repeatable experiments as a path to more scientific security work; the DETER Cybersecurity Project, which since 2004 has supported scientific experimentation in security; a UK academic Research Institute in the Science of Cybersecurity; and a very recent National Academies report on Foundational Cybersecurity Research [4].

We consider these efforts in light of historical literature in the Philosophy of Science. (Background on history and philosophy of science is available in introductory books, e.g., see Chalmers [5] and Godfrey-Smith [6].) We wish to distinguish at the outset two interpretations of "SoS". The first has focus on *scientific methods*—approaches that are, by consensus from other fields, important to scientific pursuits. The second is *externally-motivated* SoS research, work resulting from external promotion of an agenda by the name "Science of Security". Arguing neither directly for nor against SoS-labelled work, we focus on scientific methods that we believe can aid pursuit of an agenda under this name. We find that aspects on which many other scientific communities have reached consensus are surprisingly little used in security, and offer suggestions to drive security research in a more scientific fashion.

Even without a universally agreed definition of "Science", many aspects and approaches of science appear worth following. While we see great potential benefit from following "scientific methods", it is not our goal to argue that all of security must be based on rigidly scientific principles, nor that "science" is always necessarily the best or only way forward for security research; defending science is not our mission. Complementary approaches have much to offer—for example, a significant component of security is engineering [7], which despite not having as clearly articulated methods as science, shares a defining trait of traditional scientific endeavors: regular contact with, and feedback from, the observational world. We do however highlight that approaches that dismiss contact with the observable world and remain purely in the deductive realm—thus failing to bridge the *deductive-inductive split* (Fig.1, as explained below)—risk both detachment from the observable world and producing results independent of it.

Based on relevant literature from the history of science and more recent security literature, we offer observations as suggestions to advance an agenda of pursuing security research more scientifically. Our work highlights literature that may help reduce confusion in security research, and identify areas where the security community has failed to adopt accepted lessons from the broader scientific community.

II. Inductive-Deductive Split

Probably the most significant settled point in the Philosophy of Science is that inductive and deductive statements constitute different types of knowledge claims. That is, we draw conclusions about the empirical world using observations and inferences from those observations, and these are fundamentally different from mathematical or deductive statements

[1] Author's personal version: 14 Nov 2017. Official version at: http://doi.ieeecomputersociety.org/10.1109/MSP.2018.1331028, *IEEE Security&Privacy* 16(1):12-22, Jan/Feb 2018. A full original version appeared May 2017 [1].

derived from axioms. (We recommend Chalmers [5, Ch.4] for a discussion of induction vs. deduction.) For example, after many observations we may infer rules which account not just for the things observed, but things not-yet-observed (e.g., "all swans are white"). These can always turn out to be wrong, if a future observation violates a rule we have inferred (e.g., we observe a black swan).

Deduction, by contrast, produces statements that follow with certainty from a self-consistent set of axioms. An example is Euclidean geometry; e.g., Pythagoras' theorem follows from the axioms, and there is no possibility of observing anything that violates it. The separateness of these real and ideal realms has been a foundation of Philosophy and Science at least since the work of Hume [8] and Kant [9] in the 1700's. The modern description relies heavily on the work of the *logical positivists*, an influential group active in Vienna in the early 1900's [10].

While deductive statements are certain consequences of the premises, inductive statements are always subject to error. Considerable Philosophy of Science literature examines the question of when we can rely on an inductive statement. Hume's "problem of induction" [8] is that the logical basis for believing inductive claims is weak in comparison with the certainty of deductive ones. No amount of corroborating evidence can establish the truth of a generalization. That all the swans we have seen were white is no guarantee against encountering a black one. That induction has proved a reliable guide to knowledge before is itself an inductive argument, and thus circular.

Thus induction is inherently fallible. It is inferior to deduction in this respect. However, while pure deduction offers certainty, it is incapable of describing anything in the real world. Deduction always starts from axioms (or assumptions). Axioms constrain abstractions rather than real-world objects; the only requirement of a set of axioms is that it be self-consistent. Thus, it is not meaningful to ask whether an axiom (e.g., that parallel lines meet at infinity) is true or not. Insofar as a set of axioms is free of contradiction, it is true [10]. Assumptions, by contrast, posit that real-world things are constrained in certain ways (e.g., to obey Newton's Laws); if the assumptions are true then any deductions from them are also true. However, whether any set of assumptions match reality is not something that can be demonstrated with pure deduction. For example, if $a^2 + b^2 \neq c^2$ for the dimensions of a door, this is not evidence that Pythagoras' theorem is wrong. Rather, it is evidence that the particular door does not satisfy the assumptions of Euclidean Geometry (e.g., the corners are not exactly right angles). Whether a real-world system meets the assumptions of a deductive model is an empirical claim that cannot be established formally.

This is now well-established, and recognized by scientists from diverse fields [5]. Serious scientists do not claim certainty for their statements about the world, and do not claim to deduce facts about the world that weren't implicit in the assumptions. To quote Einstein:

> *As far as the laws of Mathematics refer to reality they are not certain, and as far as they are certain they do not refer to reality.*

Note that deduction simply reveals the implications of the assumptions that we started with. It says nothing about whether those assumptions match reality. Axioms and definitions are not real-world facts, so deduction starting there can say nothing definitive about the world. On the other hand, deduction that begins with assumptions or inductive inferences can explore their real-world implications. Thus, deductions that start from Newton's laws allow conclusions about real-world observations, but deductions from Euclid's postulates do not.

Thus, pure deduction offers no route to reliable knowledge about the world. This leaves us with induction. Much of the effort in Philosophy of Science is devoted to determining when we can trust conclusions based on induction. Clearly, there is a significant difference between fields like Physics and Medicine, which we regard as scientific, and ones like Homeopathy, which we do not. Can this difference be expressed in a manner that will allow us to separate more generally fields and claims and techniques that are scientific from ones that are not? The most broadly accepted answer to this question is due to Popper, which is that scientific theories should be falsifiable [11]:

> *A theory which is not refutable by any conceivable event is non-scientific. Irrefutability is not a virtue of a theory (as people often think) but a vice.*

In Popper's view, to count as scientific a theory must "stick its neck out" and make predictions or statements that run at least some risk of being contradicted by empirical observation. A theory not running such a risk is compatible with every possible observation, and thus, can't help make statements about things not-yet-observed.

Falsification places Physics in the scientific camp and Homeopathy in the non-scientific. However, it also places Mathematics (and all deductive reasoning) in the non-scientific camp. This seems more troubling to some. Does this mean that Math is worthless as a tool of scientific investigation? Since much of the ancestry of Computer Science (and Security) is mathematical it's worth trying to understand this clearly.

We trust a deductive statement only when it has been proved rigorously from a self-consistent set of axioms. We trust an inductive (or real-world statement) only when it is supported by many confirming observations, is falsifiable, and has resisted severe attempts at refutation [5, pp.184-187]. A mathematical model that assumes certain things (e.g., attacker capabilities) involves both inductive and deductive reasoning. We can trust conclusions based on the model only if both the deductions are rigorous and the assumptions hold true. However, whether the assumptions hold true is an inductive claim; we can trust that they do only when we have many confirming examples, the assumptions are falsifiable, and have resisted severe attempts at refutation. It is worth emphasizing that arguing, for example, that a set of assumptions is reasonable, is not a substitute for this test. We trust Newton's laws because they have resisted severe tests, not because it is reasonable to assume that the atmosphere is a vacuum.

Limits of formal approaches (Sidebar)

The split between inductive and deductive realms has important implications for the limitations on what can be established formally about a real-world system. An idealization is shown in Figure 1: formal and real-world systems live in the deductive and inductive realms of this depiction respectively. We can sometimes establish rigorously that a mathematically-defined system $A^{'}$ is immune to certain types of attack, under specified assumptions, e.g., based on what attackers can or cannot do, or the hardness of certain underlying problems (e.g., factoring products of large primes). However, the formal system $A^{'}$ is an abstraction. It has neither power-supply nor keyboard; nobody can ever actually use it or enjoy any strong guarantees that it is proved to offer. Instead we must use a real-world system A. Now A may be based on $A^{'}$, by which it is implied the assumptions under which $A^{'}$ was proved secure, hold also for A. However, as we have previously noted, whether any set of assumptions match reality is something that cannot be established formally (i.e., purely by deduction). Only if the assumptions are well-defined, falsifiable, and have withstood severe attempts at refutation, can we have some confidence that A enjoys the security properties proven of $A^{'}$. Even then there is the residual uncertainty that accompanies any inductive statement [8], and the difficulty of ensuring that the list of assumptions is complete [12]. If we have not attempted to refute, or cannot explicitly state the assumptions, then the security of A rests entirely on assumption. That is, as a real-world system, the security of A can never be proved; but it can't be trusted as an inductive claim either, unless the assumptions have survived concerted attempts at refutation. An argument that a real-world system is secure because the assumptions are reasonable falls short on both fronts: not only does a proof plus assumptions not establish a claim deductively, it doesn't establish it inductively either.

As an example of deductive-only reasoning failing in security, consider side-channel attacks on cryptosystems. Many deductive arguments have proven attacks on given systems to be equivalent to solving hard math problems. This reasoning is logically sound at one level, but the deductive model has often failed to consider various side-channel attacks which recover private keys without breaking any hard problems. The real-world system in which attacks actually occur simply contains threat vectors beyond those considered in the abstract system of deductive reasoning. Theorists may argue that "everyone" understands such limitations; of course, the many who do not are rarely technically qualified to explain their confusion. There is of course no possibility whatsoever of *proving* that a real-world system is immune to all attacks; it can be argued that some specific attacks, or classes thereof, can be stopped (sometimes even these arguments fail, e.g., if assumptions and environmental conditions at hand differ from those assumed).

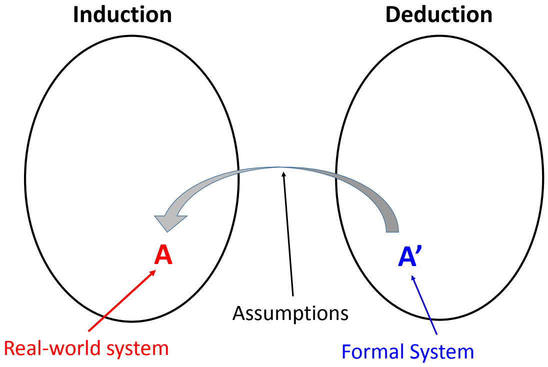

Fig. 1: The inductive/deductive split. A formal system $A^{'}$ can be proven secure, subject to assumptions, e.g., on attacker capabilities. We can't conclude that a real-world system A enjoys any $A^{'}$ properties unless all assumptions and axioms of the formal system are satisfied by A. Whether these are indeed satisfied is an empirical claim, not one that can be established by proof.

III. A PATH TO MORE SCIENCE IN SECURITY

Feynman gave a description of Science that makes it sound easy (quoted in [1]): guess laws, compare predictions to experiment and what disagrees with experiment is wrong. This simple recipe gets complex on considering the details. If we lack even guesses at fundamental laws in security, how are we to proceed? A good place to start is with possible definitions of Science itself. This is our first "take-away" (T) point.

Clearly define desired aspects of Science

T1: *Pushes for "more science" in security, that rule nothing in or out, are too ambiguous to be effective. Many insights and methods from philosophy of science remain largely unexplored in security research.*

What exactly would be desired of a Science of Security? How is Science defined in this context—would it mean research that can be captured by equations, or quantized with hard numbers, or illustrated by graphs? Are repeatable experiments the magic ingredient? Perhaps rigour or mathematical proofs? Or hypothesis testing through controlled-variable experiments? What role should human factors tests play in security research?

A search of the security literature for guiding descriptions of a Science of Security finds many attempts which simply re-use the term, and circular exhortations to do things more scientifically. There is naturally some tension about defining Science: nobody wishes to have their work declared on the wrong side of a line laid down by someone else. But confusion as to what is (or is not) wanted allows us to go on as before and enables every researcher to cast their current work as precisely what is needed, or retrofit definitions so that favorite areas feature prominently. Thus, we find it unhelpful to promote a "Science of Security" movement without a serious attempt to explicitly define or openly discuss more precise objectives. Any strong push for more "science" in security research and practice should be supported by explicit discussion of what we mean. We argue that those pursuing security research would be well-served by this, and those promoting a "Science of Security" should encourage this and themselves add clarity. Hiding behind vague language doesn't help; asking for something vaguely defined is ineffective.

Rather than offer our own definition of a Science of Security we seek to leverage the centuries of accumulated wisdom in other disciplines. How have top-tier scientists characterized Science? By consensus, when theory conflicts with observation, the theory is wrong. Theory must be brought into contact with the observable world—to make conflict with observation possible. Claims must be falsifiable (see Popper above; and T3 below). Claims consistent with every possible observation are not scientific—nothing observable depends on such a claim. Confidence in a claim increases the more it is scrutinized, and the more (failed) attempts to falsify it. Refutation has been called the engine of scientific progress, allowing self-correction and iterative improvement. Claims must also be precise; as Feynman famously said, "You can't prove a vague theory wrong." Whether those encouraging a Science of Security seek these, or other characteristics, should be clarified.

Acknowledge the inductive-deductive split

T2: *Ignoring the sharp distinction between inductive and deductive statements is a consistent source of confusion in security.*

The importance of this divide, and of being clear which type of statement is being made, is recognized in most branches of Science. To be unequivocal on this point, note that there is no possibility whatsoever of proving rigorously that a real-world system is "secure" in the commonly interpreted sense of: invulnerable to (all) attacks. This is not simply because of the possibility of flawed implementation. Formal methods can deduce certain guarantees only if the assumptions are met. However, we repeat once more, whether a real-world system meets any set of assumptions is an empirical claim, not something that can be established formally. The combination of a rigorous deductive statement and a (necessarily) less-than-rigorous empirical one can never yield a rigorous guarantee. A claim that a real-world system enjoys the same security guarantees as mathematically proven is logically unsound.

Thus, while formal approaches are valuable, the hope that they offer a way of avoiding the messiness and fallibility of empirical statements is simply illusory. Unfortunately, much work on formal approaches appears to imply that avoiding this messiness is actually possible. For example, Shoup (quoted in [13]) suggests that with provable security "we essentially rule out all possible shortcuts, *even ones we have not yet even imagined*. The only way to attack the cryptosystem is a full-frontal attack on the underlying hard problem. Period." Another formalist claims that [14] "The only way to" get evidence of the security of a design "is to develop a formal mathematical model and language in which to reason about such schemes." We claim that these statements, and others like them that assert an inherent general advantage to formal approaches, are incorrect. As noted earlier, we can trust that a specific real-world system inherits proved properties of a formal one only if the assumptions have been *severely tested*. Or course this is only possible if the assumptions are specified. We can make specific claims of specific systems by testing their assumptions, but any claim of general superiority (i.e., independent of the assumptions) is insupportable.

In summary, a proof can deliver guarantees only about a mathematical model or formal system, not a real-world one. Since it is real-world systems that we ultimately use, the choice is not between one approach which offers immunity to attack and another which does not. Rather, the question is to what degree properties proven about a mathematical system can be translated into useful properties of a real-world one.

Stop relying on unfalsifiable claims

T3: *Unfalsifiable claims are common in security—and they, along with circular arguments, are used to justify many defensive measures in place of evidence of efficacy.*

There is also considerable failure to avoid unfalsifiable claims and statements. There is an inherent asymmetry in computer security that makes large classes of claims unfalsifiable [15]. We can observe that something is insecure (by observing a failure) but no observation allows us to determine empirically that something is secure. It follows that we can't ever show that something isn't necessary for security. For example, to falsify "in order to be secure you must do X" we would have to observe something secure that doesn't do X. If we interpret "secure" as a real-world property, such as the avoidance of future harm, then observing it requires knowing the future. On the other hand, if "secure" is interpreted formally, while we can now identify mathematically secure systems, we can make no deductions about real-world events (e.g., that harm will be avoided). In summary, claims of necessary conditions for real-world security are unfalsifiable. Claims of necessary conditions for formally-defined security are tautological restatements of the assumptions.

An illustrative example may clarify. To falsify the claim "a password must have at least 8 characters and contain letters, digits and special characters to be secure" we would have to find a secure non-compliant password. However, we can't find a secure password, since there is also no way to show that a password is safe against not-yet-known attacks. The alternative is to formally define security of a password as having a certain structure, or resisting a certain number of guesses, etc. We can of course find necessary conditions if security is defined formally, but these are just restatements of the definition (e.g., a password that withstands 10^{14} guesses is secure if security means withstanding that number of guesses). To relate the formal (e.g., password has certain structure) and real-world (password can resist guessing attack) notions of security we must make assumptions about what an attacker can and cannot do (e.g., attacker can get access to the hashed password file but cannot execute more than 10^{14} guesses).

Assumptions that attackers cannot do something (e.g., exceed 10^{14} guesses) are unverifiable. By symmetry, assumptions that they can do something (e.g., get the hashed password file) are unfalsifiable. A significant number of defensive measures are justified by unfalsifiable assumptions about what attackers can do.

Popper identifies falsification as the engine of self-correction [11]. Security measures that are justified with unfalsifiable claims are immune to corrective feedback. This can result in accumulation of countermeasures as there is no mechanism for rejecting unnecessary measures or declaring them no longer required. Thus a failure to avoid unfalsifiable claims may account for the security overload users complain of experiencing.

Self-correction requires that observable evidence both for and against X be possible. This requires speaking of outcomes rather than security, since the latter is inherently unobservable. If we can say when we would accept that X isn't doing any good (e.g., no observed difference under specified circumstances) as well as when it is (e.g., observed difference) then self-correction is restored.

Stop using the "security is special" excuse

T4: *Claims that unique aspects of security exempt it from practices ubiquitous elsewhere in science are unhelpful and divert attention from identifying scientific approaches that advance security research.*

It is sometimes argued that security has special difficulties and unique challenges that preclude placing the field on a more scientific footing. It faces an adaptive, intelligent adversary [16]; while bridge-builders must address hostile conditions, nature does not devise new types of storms when stronger bridges appear. Also, security depends largely on human artifacts (computer software and hardware), involves human factors, and the relevant environment and threats evolve rapidly. As such, *"cyber-security is an artificially constructed environment... only weakly tied to the physical universe... There are no intrinsic "laws of nature" for cyber-security as there are, for example, in physics, chemistry or biology"* [2].

Despite such arguments, we reject the idea that security is so special that scientific methods should be abandoned. Definitions of Science are intentionally independent of discipline-specific details. Popper and later philosophers sought a demarcation criterion of use whether investigating General Relativity, heredity in fruit flies, Marxist theory or phrenology—without specific pre-conditions, e.g., that a discipline have invariant laws or be free of active adversaries. Indeed it was argued that a scientific approach was simply the most reliable way of investigating matters of fact—we might be unhappy with the constraints it imposes or the strength of statements it allows, but no clearly superior alternative is available.

While security certainly faces major challenges, so do other fields. In Astronomy, the paths of planets and stars are not easily controlled as independent variables, but observational experiments prove invaluable. Quantum Physics research continues despite the inability to directly observe subatomic particles. Biological and military systems also face adversaries [17]; for example, the evolution of pathogens changes underlying landscapes. In scientific fields where landscapes evolve (e.g., flu viruses—or computer viruses or software ecosystems), repeated measurements are needed over time, to re-test theories against current environments. In that many of its findings can be elegantly captured as time-invariant mathematical laws, Physics is an exception rather than the rule [5]; *"most biology has little use for the concept of a law of Nature, but that does not make it less scientific"* [6].

Numerous branches of science have overcome difficulties that once seemed unique and insuperable. Pleading uniqueness to avoid being held to scientific approaches is common in unscientific fields, and would place Security in poor company.

Negative statements lacking alternatives don't aid progress. Suggesting that a scientific approach is a poor fit for security in no way helps unless we suggest a better alternative. The broadly accepted outlines of scientific method, having evolved over much time and great scrutiny, are by consensus view the best way to figure things out. Such methods should be used wherever they help advance security research and practice.

Physics is not a role model for all of Science

T5: *Physics-envy is counterproductive; seeking "laws of cybersecurity" similar to physics is likely to be a fruitless search.*

This observation is not new but warrants explicit discussion. The accomplishments of physics over the last 150 years may be the most successful scientific research program ever conducted. However, most sciences do not look like physics (nor crypto, below), and we should not pre-judge what a Science of Security will look like. Large sub-areas of security might be better compared to the life sciences [17]. Caution should be exercised that a desire for quantification does not disadvantage the applied or systems research, or impose mandatory quantitative metrics where no such meaningful metrics are known. Admitting the possibility of there being no formal laws to find leaves other paths open.

One reason physics is the envy of other sciences is that its results are often quantitative and of high precision, aside from being reproducible. Quantitative metrics are often positioned as a gaping hole in security; advances in security metrics may even be viewed as essential for progression to a science. Yet progress has been slow in efforts to define and advance security metrics. Pfleeger [18] suggests that an important step forward is to "stop insisting that quantitative is better than qualitative; both types of measurement are useful." Verendel [19] finds that despite significant work, little evidence supports the hypothesis "security can correctly be represented with quantitative information", and notes that "many assumptions in formal treatments are not empirically well-supported in operational security". The JASON report notes that "things that are not observed such as new attack approaches are not going to contribute to metrics. It is not possible to definitively measure a level of security" [2, p.4]. We should also be careful to ask for metrics only where they are meaningful—to measure things which provide indications useful for security, rather than just to produce numbers.

Crypto is not a role model for all of Security

T6: *Crypto-envy is counterproductive; many areas of security, including those involving empirical research, are less amenable to formal treatment or mathematical role models.*

Cryptography has a special hold on the minds of security researchers, and without the accomplishments of cryptography many of the security technologies we take for granted might not exist. However, despite many citing crypto as role-model for a Science of Security, its methods are less suitable for numerous areas, e.g., operating systems, security architecture and software engineering. Crypto's rigorous mathematical foundations are in sharp contrast to, for example, "messy" systems security, and areas dealing with human factors. Crypto does not typically involve the type of scientific experimentation found in empirical sciences generally, nor systems security in particular.

Some have asserted, for example, that *provable security* has transitioned crypto to a science [20]; counter-arguments claim that crypto is far from science [21]. Those who position crypto as a role model should beware: (1) its heavy reliance on formal proofs (see Inductive-Deductive Split); (2) its poor fit for many empirical sub-areas, as just noted; and (3) its poor track record of misleading language, such as "provable security" and "proofs of security". We expand now on (3).

"Provable security" involves proofs showing that breaking a target cryptosystem allows solving a believed-hard problem in not much further effort; it can be used to rule out important classes of attacks. But as has been noted [21], the terms "proof" and "theorem" have historically implied 100% certainty (statements accepted unconditionally), whereas provable security results are conditioned on the degree to which a model matches reality. This is entirely avoidable confusion.

Another area of security in which cryptographic research has seen controvery has involved *side-channel attacks* (see Sidebar, above), a well-known collection of powerful methods for extracting cryptographic keys without defeating crypto-algorithms themselves. In a higly accessible exposition of "what can go wrong when systems that have been proven secure in theory are implemented and deployed in real environments", Degabriele et al. [20] observe:

> *Practitioners might think provable security results provide an absolute statement of security, especially if they're presented in such a manner. When they later discover that a scheme is insecure because of an attack outside the security model, this might damage their confidence in the whole enterprise of provable security.*

They also explain other practical side-channel attacks despite provable security proofs on MAC-then-encrypt constructions, including on an SSL/TLS mechanism exploiting observable timing differences caused by padding errors.

Overall, cryptography has been one of the most successful areas of security research. However, its particular methods do not make it a role model for a Science of Security any more than physics is a role model for life sciences research.

Insist on bringing theory into contact with observation

T7: *Both theory and measurement are needed to make progress across the diverse set of problems in security research.*

The history of science offers many examples of misguided beliefs caused by a failure to bring theory into contact with observation. Equally, however, indiscriminate measurement offers fewer opportunities for discovery than experiments that deliberately set out to refute or refine existing theory.

Recall that a scientific model is judged on the accuracy of its predictions; lack of data or difficulty in making measurements does not justify trusting a model on the sole basis of its assumptions appearing reasonable. But this is often done in security research.

Consider for example the long-accepted wisdom that passwords are made stronger by the inclusion of upper-case letters, digits and special characters, recommended by Morris and Thompson [22] to address the observed problem of users choosing English words as passwords. This has for years been widely mandated, supported by long-standing authentication guidelines including Appendix 1 of NIST SP 800-63 (June 2004) [23]. It was assumed that including digits and special characters would push users to choose random-like strings. Originally, this may have appeared a reasonable assumption (even if false); the strength of users' preference for simple passwords and ingenuity in circumventing security measures was not obvious in 1978. However, storing passwords as salted hashes (a second major recommendation [22]) precluded easily measuring whether mandates on character composition were having the predicted effect. Recent empirical work shows that they do not [24], [25], [31]. For three decades after 1978, not only were there few apparent attempts to check the accuracy of the prediction, but great effort was devoted to having users follow misguided means to improve password security. Community actions were based on the assumed truth of something that depended critically on an untested assumption. We return to this, with a positive outcome, in our concluding remarks.

Insist results be put in context with full solutions

T8: *More security research of benefit to society may result if researchers give precise context on how their work fits into full solutions—to avoid naive claims of providing key components, while major gaps mean full-stack solutions never emerge.*

That Security research should aim to benefit society is generally accepted, especially when given targeted funding on the grounds of practical importance to society. Maughan [26] discusses the challenges of translating security research from academic lab to real world, in the context of government-funded security projects. To trigger useful community discussion, we encourage considering the question: Who is responsible for the overall roadmap for emergence of full-stack solutions addressing important problems?

Science has seen many instances of difficult problems involving complex, multi-part solutions. Sometimes the responsibility of ensuring delivery of all parts to a complete solution has been taken on single-handedly by one scientist spanning the full spectrum from fundamental research to a fully-engineered solution; an exemplar is Pasteur, as set out in a research approach called *Pasteur's Quadrant* [27] formulated by former Princeton University dean Donald Stokes. Rather than seeing basic research (whether theoretical or experimental) and applied research as competing against each other, Stokes identified numerous examples from historical science where their combination led to direct societal benefit. His model, advocating for *use-inspired basic research* intersecting fundamental and applied research, calls for research problems that seek both fundamental understandings and considerations of use. Besides Pasteur, his exemplars of this approach include Kelvin (physics), Irving Langmuir (physical chemistry of surfaces), molecular biologists researching genetic codes, Manhattan Project scientists, and the large collection of advances that enabled successful cardiac surgery.

Stokes notes a historical bias dating to Greek elites and philosophers favoring pure inquiry, leaving manual labor and practical arts to lower classes. Outside of medical practice, society benefited little from elite knowledge until Greek science hit western Europe; Bacon and contemporaries combined science and manual service, including to improve technology where previously this was considered the realm of laborers. Supporting this approach, note that improved technology has often enabled scientific advances rather than vice versa, e.g., telescopic lenses aiding astronomy,

Turing Award winner Herbert Simon [28] also had strong views on historical tensions between pure and applied sciences, and on what he called *sciences of the artificial*—involving human-made artifacts including computer hardware and software. Many security results depend directly on such artifacts. This raises the risk of results positioned as general (beyond such artifacts) when they are not; generalization is typically inductive. Independent of this, security results may fail to enjoy the long-term relevance of results in the hard sciences—most of the physical world is stable over time, while software and adveraries can change daily. Both the generality and longevity of results impact their societal value. This increases the importance of emphasizing full-stack solutions.

Insist that assertions be supported by evidence

T9: *Conflating unsupported assertions, and argument-by-authority, with evidence-supported statements, is an avoidable error especially costly in security.*

An authoritarian statement is an assertion that is made forcefully, often by someone in a position of power or influence, but not supported by evidence. It was common to rely on authoritarian statements before the rise of the scientific method; in contrast, Science establishes facts by observation and experiment, and the status of a scientific statement derives from the evidence supporting it, rather than the authority of the person making the statement.

If a security policy is based on authoritarian statements, and it is both unaccompanied by supporting evidence and obtaining empirical data that might contradict it is difficult, then overturning the policy is difficult. Complicating factors include vague claims being hard to refute (see earlier), and the impossibility of establishing that a defense is *not* necessary (see T3). Such errors are costly since self-correction [11] is now lost.

Landwehr notes [7]:

> *"Before the underlying science is developed, engineers often invent rules of thumb and best practices that have proven useful, but may not always work."*

They may also be confused with authoritarian statements. 'Rules-of-thumb' are not called 'laws' for many reasons, including that they have not been as rigorously tested, nor as precisely stated; similarly for security 'principles'. The utility of both derives from the evidence supporting them, and their predictive ability; for both, we must be careful that they are supported not only by convincing evidence, but that their relevance is continually re-challenged, e.g., as computer systems and their environments evolve. Scientific statements stand or fall on how they agree with evidence. Calling something a principle, best-practice, rule-of-thumb, or truism does not remove the burden of providing supporting evidence.

The original NIST authentication guidelines [23] (discussed under T7) acknowledge that many of the password strength measures suggested are based on unverified assumptions, and thus are rough rules-of-thumb; these came to be accepted as principles. Apparently, the difficulty of acquiring empirical data in security—e.g., due to instrumentation challenges, privacy concerns, and commercial forces—extends this problem to many examples in security, including the non-evidence-based security measures noted under T3.

Insist on explicit claims and assumptions

T10: *Despite consensus that assumptions need be carefully detailed, undocumented and implicit assumptions are common in security research.*

When we fail to make assumptions explicit, to subject them to efforts at refutation, or to make proper connections between abstractions and the real world, brilliant and apparently deep results may have little connection with and impact on the observable world. Greater care in explicitly detailing and challenging pre-conditions would better illuminate the breadth or narrowness of results.

Recommending that assumptions be carefully documented seems inadequate. The challenge is not in getting agreement on the importance of doing so, but in establishing why we fall so far short of a goal that few presumably disagree with, and how this might be addressed. One possibility is to find a *forcing function* to make assumptions explicit. As one example (towards a different goal), *Nature* demands that abstracts contain a sentence beginning "Here we show that." Platt [29] recommends answering either "what experiment would disprove your hypothesis" or "what hypothesis does your experiment disprove." By convention, many fields expect explicit hypothesis testing.

As noted under T3, the evidence falsifying a claim is easily described if the claim is precise. If a theory says "X should never happen under assumptions A, B and C" then showing that it does suffices to refute the claim. But when a statement is vague, or assumptions implicit, it is unclear what, if anything, is ruled out. Thus, difficulty articulating what evidence would falsify a claim suggests implicit assumptions or an imprecise theory [5].

Consider the large body of work devoted to modifying security-related user behavior. Many large web sites, and governments, devote considerable energy to user education. The bulk of this takes the desirability of the goal as given—e.g., that raising awareness of cyber threats or paying more attention to warnings is inherently beneficial. The assumption that this will improve actual outcomes is often left implicit and unquestioned. Examples in the research literature include defining effectiveness as the fraction of users terminating TLS connections after a security warning, complying with unvalidated advice on detecting phishing attacks, or choosing a password of a certain format. Many efforts to influence users implicitly assume a goal of minimizing risk. But this implies no measure should ever be neglected; a more realistic goal is to minimize the sum of risk *plus* the associated defensive cost [30]. Unstated assumptions too easily escape debate.

Also regarding implicit assumptions: everyone agrees that assumptions should be clearly stated. Precisely and completely enumerating them is rarely argued against, though often it is claimed that relevant assumptions are of course understood by everyone (this means: they are understood by everyone that understands them). But rarely are assumptions explicitly listed, and this is more important in security than in, e.g, traditional physical sciences because whereas the physical world is not capricious, the adversarial world of security and human artifacts is. Given the importance of falsification and refutation (above), note that to refute assumptions, it is necessary to see the list of candidate items to be refuted or falsified. Another reasonable question to ask is: what evidence would be accepted as refuting a claim or assumption? If none can be stated, a claim is unfalsifiable.

IV. Concluding observations

The above observations may leave readers with a negative view of prospects for a Science of Security. While we do not believe science is the exclusive answer to every problem, we suggest that its methods have been underutilized in security. Our more optimistic message is that these observations highlight enormous opportunities to leverage known scientific methods and lessons from other communities. To this end, as a positive example, we return to the earlier discussion of basing password strength on heuristic estimates of Shannon entropy (Appendix 1 of 2004 SP 800-63 [23])—a primary basis for complex password composition policies as well as password expiration policies.

In a 2010 paper [31], this entropy-based approach was directly tested through use of a large empirical dataset of leaked passwords, and shown to very poorly model real-world guessing attacks. The same year, the efficacy of password expiration policies was also explored through an empirical study combined with new algorithmic guessing attacks [32]. This provided strong evidence against expiration policies, finding that newly chosen passwords can often be predicted (guessed) from old passwords with remarkably high success, and therefore once an attacker knows an existing password, forcing an update results in far less benefit than hoped. A 2012 paper [24] detailed a large empirical study of 70 million passwords, finding that typical strength of user-chosen passwords against guessing attacks was the equivalent of 10-20 bits of cryptographic strength, and convincingly argued that *partial guessing metrics* offer a far better measure than Shannon entropy. This research collectively contributed to the 2017 revision of NIST SP 800-63 [23] effectively withdrawing support for both password expiration policies (unless there is evidence of password compromise), and complex password composition policies; the old Appendix 1 no longer appears.

We conclude with a few more overall remarks. A first meta-observation is that the Security community is not only experiencing some problems well-known in other scientific fields, but is also not leveraging history lessons well-known in the mainstream scientific community. We suggest that those who seek, and advocate for, a Science of Security would benefit from being well-versed in science history.

A second meta-observation pertains to those seeing the end-goal of security research being to ultimately improve outcomes in the real world. The failure to validate the mapping of models and assumptions onto environments and systems in the real world has resulted in losing the connections needed to meet this end-goal. A rigorous proof of security of a mathematical system allows guarantees about a real-world system only if the coupling between them is equally rigorous. We have seen repeated failure in poor connections between mathematical systems and real-world ones, and consequent failure of the latter to enjoy properties promised by the former. The scientific value of theory and formal models lies in their ability to make predictions about the real world; experimentation exposes theory to contact with the observable world, opening the door for feedback and model correction. A purely deductive world of axioms, assumptions and their logical implications is not Science. Science requires contact with the observable world.

Longstaff et al. [33] (see also [34]) argue that computer security researchers whose primary background is Computer Science or Mathematics have enjoyed little training in experimental science or scientific methods, and would benefit from better knowledge of these—and that barriers to a Science of Cybersecurity include a community culture favoring quick papers vs. time-consuming efforts typical in experimentally-based fields; a lack of proper scientific training; and culture rewarding novelty and innovative technology over scientific accumulation of knowledge.

Simply wishing for a Science of Security will not result in one. We believe the community needs more active discussion of what would characterize such a science, and of various approaches that might progress it. Whether this goal is attainable, or worthwhile, depends in large part on how it is defined, and on defining it more clearly to begin with.

Acknowledgements. We thank the anonymous referees, all those who provided feedback on this and earlier versions of this work, and guest editors Terry Benzel and Sean Peisert of the special issue of *IEEE Security&Privacy* in which this paper appears, for their support of this work. Van Oorschot is Canada Research Chair in Authentication and Computer Security, and acknowledges NSERC for funding the chair and a Discovery Grant.

References

[1] C. Herley and P. C. van Oorschot, "SoK: Science, Security, and the Elusive Goal of Security as a Scientific Pursuit," in *2017 IEEE Symp. Security and Privacy*, pp. 99–120.

[2] JASON Program Office, "Science of Cyber-security (JASON Report JSR-10-102)," Nov 2010, http://fas.org/irp/agency/dod/jason/cyber.pdf.

[3] D. Evans, "NSF/IARPA/NSA Workshop on the Science of Security," Nov 2008, report, Berkeley, CA (presentation slides available online). http://sos.cs.virginia.edu/report.pdf.

[4] National Academies of Sciences, Engineering, and Medicine, "Foundational Cybersecurity Research: Improving Science, Engineering, and Institutions," 2017, https://doi.org/10.17226/24676.

[5] A. F. Chalmers, *What is this thing called Science? (4th edition)*. Hackett Publishing, 2013.

[6] P. Godfrey-Smith, *Theory and reality: An introduction to the philosophy of science*. University of Chicago Press, 2009.

[7] C. Landwehr, "Cybersecurity: from engineering to science", *The Next Wave*, 19(2):2–5, 2012, NSA, special issue on: "Developing a Blueprint for a Science of Cybersecurity". https://www.nsa.gov/resources/everyone/digital-media-center/publications/the-next-wave/assets/files/TNW-19-2.pdf.

[8] D. Hume, *An enquiry concerning human understanding: A critical edition*. Oxford University Press (ed. T.L. Beauchamp; original version 1748), 2000.

[9] I. Kant, *Critique of pure reason*. Cambridge University Press (translated by Paul Guyer; original version 1781), 1998.

[10] A. J. Ayer, *Language, Truth and Logic*. Dover Publications, New York (unaltered reproduction of 2/e, 1946), 2014.

[11] K. Popper, *Conjectures and refutations: The growth of scientific knowledge*. Routledge, 1959.

[12] M. T. Dashti and D. A. Basin, "Security testing beyond functional tests," in *Proc. ESSoS 2016 (Engineering Secure Software and Systems)*, ser. Springer LNCS 9639, pp. 1–19.

[13] N. Koblitz and A. Menezes, "Another look at security definitions," *Adv. in Math. of Comm.*, vol. 7, no. 1, pp. 1–38, 2013.

[14] H. Krawczyk, "Letters to the Editor: Koblitz's Arguments Disingenuous," in *Notices of the AMS*. AMS, 2007, p. 1455.

[15] C. Herley, "Unfalsifiability of security claims," *Proc. National Academy of Sciences*, vol. 113, no. 23, pp. 6415–6420, 2016.

[16] D. Geer, "The Science of Security, and the Future (T.S. Kuhn Revisited)," talk at NSF meeting 6-Jan-2015 and RSA 23-Apr-2015, essay at http://geer.tinho.net/geer.nsf.6i15.txt, shorter version "For Good Measure: Paradigm" in *USENIX ;login:* 41(3):80–84, 2016.

[17] S. Forrest, S. A. Hofmeyr, and A. Somayaji, "Computer immunology," *Commun. ACM*, vol. 40, no. 10, pp. 88–96, 1997.

[18] S. L. Pfleeger, "Security Measurement Steps, Missteps, and Next Steps," *IEEE Security & Privacy*, vol. 10, no. 4, pp. 5–9, 2012.

[19] V. Verendel, "Quantified security is a weak hypothesis: a critical survey of results and assumptions," in *Proc. NSPW 2009*. ACM, pp. 37–50.

[20] J. P. Degabriele, K. Paterson, and G. Watson, "Provable security in the real world," *IEEE Security & Privacy*, vol. 3, no. 9, pp. 33–41, 2011.

[21] N. Koblitz and A. Menezes, "Another Look at 'Provable Security'," *J. Cryptology*, vol. 20, no. 1, pp. 3–37, 2007.

[22] R. Morris and K. Thompson, "Password security: A case history," *Commun. ACM*, vol. 22, no. 11, pp. 594–597, 1979.

[23] W. E. Burr, D. F. Dodson W. T. Polk, "NIST SP 800-63: Electronic Authentication Guideline (version 1.0)," Jun 2004, password guidelines revised June 2017 in: NIST SP 800-63B: Digital Identity Guidelines—Authentication and Lifecycle Management, Paul A. Grassi et al.

[24] J. Bonneau, "The science of guessing: analyzing an anonymized corpus of 70 million passwords," in *2012 IEEE Symp. on Security and Privacy*, pp. 538–552.

[25] M. L. Mazurek, S. Komanduri, T. Vidas, L. Bauer, N. Christin, L. F. Cranor, P. G. Kelley, R. Shay, and B. Ur, "Measuring password guessability for an entire university," in *ACM CCS 2013*, pp. 173–186.

[26] D. Maughan, D. Balenson, U. Lindqvist, and Z. Tudor, "Crossing the 'valley of death': Transitioning cybersecurity research into practice," *IEEE Security & Privacy*, vol. 11, no. 2, pp. 14–23, 2013.

[27] D. E. Stokes, *Pasteur's Quadrant: Basic Science and Technological Innovation*. Brookings Institution Press, 1997.

[28] H. A. Simon, *The Sciences of the Artificial*. MIT Press, Cambridge, MA (third edition; originally published 1969), 1996.

[29] J. Platt, "Strong inference," *Science*, vol. 146, no. 3642, pp. 347–353, 1964.

[30] D. Florêncio, C. Herley, and P. C. van Oorschot, "Password portfolios and the finite-effort user: Sustainably managing large numbers of accounts," in *Proc. 2014 USENIX Security Symp.*, pp. 575–590.

[31] M. Weir, S. Aggarwal, M. Collins, and H. Stern, "Testing metrics for password creation policies by attacking large sets of revealed passwords," in *ACM CCS 2010*, pp. 162–175.

[32] Y. Zhang, F. Monrose, and M. K. Reiter, "The security of modern password expiration: An algorithmic framework and empirical analysis," in *ACM CCS 2010*, pp. 176–186.

[33] T. Longstaff, D. Balenson, and M. Matties, "Barriers to science in security," in *Proc. ACSAC 2010*. ACM, pp. 127–129.

[34] R. A. Maxion, T. A. Longstaff, and J. McHugh, "Why is there no science in cyber science? (panel)," in *Proc. NSPW 2010*. ACM, pp. 1–6.