Validating Computer Security Methods

Meta-methodology for an Adversarial Science

Antonio Roque
MIT Lincoln Laboratory
Lexington, MA, USA

Abstract—Recent explorations on the science or theory of computer security have been hindered by its unique properties. We confront this by precisely defining those properties: that computer security is adversarial and engineered, and that because of this it is contextual. We use these definitions to address the practical question of how we can justify the validity of our methods. To answer this meta-methodological question, we develop a taxonomy of methods, and consider the components of a well-constructed methodological validation. We use strategic theory to derive one such validation, and discuss the uses and properties of validations.

I. INTRODUCTION

A tremendous amount of work is done on computer security, and researchers occasionally seek a science or a theory to guide them [46][50][51]. Unfortunately, achieving this goal has been “elusive” [35]. As engineers and mathematicians, this may seem like an overly abstract topic, perhaps best suited to scholars or bureaucrats. So we need to be specific about what we need from a science or theory, and how it will help our technological efforts. This paper will be guided by the specific need of answering the question: how do we know if we’re using the right methods? In other words, if we’re developing a “theory, design, implementation, analysis, verification, or empirical evaluation and measurement of secure systems,” how can we be sure that we are in fact increasing our security? These are meta-methodological questions, which have a long history of being studied in computer security [29] [64] [79, pg. v].

Ultimately we will argue that computer security involves technical methods that are vulnerable to adversarial and contextual invalidation, and should be validated by their operational context. Operational methods are also vulnerable and are evaluated by analytical methods that can be characterized as theoretical in a general sense of the word, meaning philosophical, strategic, and what is often indicated by “scientific.” These theoretical methods are validated by other theoretical methods, which creates a regress, and which are partially validated by ongoing interdisciplinary study of the adversarial struggle.

But for this answer to make sense we need to define our terms, and to carefully examine the nature of the adversaries we’re securing against, and the engineered systems we’re working with. So in Section II we will define several fundamental terms related to the unique properties of computer security. And in Section III we will be specific about the types of methods we are discussing and how they are related. So taxonomies are an essential part of this paper. Computer security has a laudable history of using such classifications and definitions [4] [43] [45].

From here on we will use the term cyber, which has been used in terms such as cyberspace and cyber security, to refer to computers, electronic communication systems, and related technical infrastructure systems [15, pg. 60]. Although this term is relatively new, it has the advantage of being technology-neutral. Over the coming decades, our technical environment may move away from traditional computers and networks, while keeping the fundamental attributes of controlling, communicating machines. We seek answers that will remain relevant beyond our immediate technological environment.

II. PROPERTIES OF CYBER SECURITY

As with all disciplines, cyber security possesses a set of distinctive properties. Herley and van Oorschot [35] warn against using these properties as an excuse to avoid identifying a coherent scientific approach. However, this paper is not making such an excuse; instead, it is using cyber security’s unique properties to suggest and constrain meta-methodology.

As we will describe, cyber security is:

1) Adversarial
2) Engineered
3) Contextual

A. Adversarial

By definition, the existence of an intelligent adversary is inherent to cyber security.

There are numerous definitions of cyber security and they usually justify this assertion implicitly. For example, the definition “Cyber security refers generally to the ability to control access to the networked systems and the information they contain,” [6, pg. 1] implies the existence of something capable of accessing the systems and information against our wishes. Likewise, “Security is a composite of the attributes of confidentiality, integrity, and availability” [4, pg. 13] implies an entity which could break that confidentiality, integrity, and availability. Section III-C and Section IV describe reasons for relying on foundational definitions in this way, and some of the dangers of doing so. For now, of the three properties we will spend the most time on the adversarial, as it is foundational.
To explore the adversarial property, we will use as a point of reference the work of contemporary science historian Peter Galison, who used several theoretical ideas from post-World War 2 control-theorist Norbert Wiener. Galison presented his ideas to the JASON government advisory group for their 2010 report on ‘Science of Cyber-Security’ [39], and also as a keynote to a 2012 science of security workshop [25].

Galison adapted terminology from Wiener to identify an adversarial science as one that involves an “active oppositional intelligence” [24, pg 232] who in Wiener’s words “is determined on victory and will use any trick of craftiness or dissimulation to obtain this victory” [24, pg 231]. Galison suggests cyber security is one such adversarial science [25], along with game theory, control theory, and military operations research [24, pg 232]. In contrast, non-adversarial sciences such as Biology and Chemistry involve overcoming “chance and disorder” [24, pg 231] including a mere lack of knowledge.

We expand upon their basic distinction to consider three descriptions of adversarial disciplines. A discipline practitioner can interact with their adversary, each trying to disrupt the other’s mission, but a practitioner can also work to accomplish their mission in the presence of an adversary without directly attacking their adversary’s mission, or merely theorize about adversaries without the theorizing being part of an adversarial interaction.

First, the term adversary interactive describes a discipline in which the discipline practitioner and their adversary each attempt to disrupt the other’s mission. An example from cyber security is in attack-defense “capture the flag” exercises. The nature of these events vary, but a typical one is described by Cowan et al. [12]: participating teams share a computer network and score points by maintaining a server that provides a service which is known to be vulnerable to exploitation. Participating teams can also score points by exploiting that service while it is being run on their adversaries’ servers. So all participating teams are adversaries of each other.

A non-cyber example of adversary interaction is in “active defense” criminal trials, in which defending and prosecuting lawyers each attempt to establish narratives regarding the defendant while attacking the validity of the other lawyer’s narrative.

Second, the term adversary present describes a discipline in which an adversary attempts to disrupt a discipline practitioner’s mission. The practitioner does not attack their adversary.

As an example from cyber security, the JASON group ultimately understood cyber security to be “science in the presence of adversaries” [39, pg 15], a formulation that seems to be shaped by cryptography [39, pg 2], which is about “communication in the presence of adversaries” [62, §2]. For example, communications that aspire to privacy define an adversary and their capabilities. The adversary may observe encrypted messages, or observe example encrypted messages with their unencrypted versions, or the adversary may have access to a version of the cryptographic system [62, §6.3.1]. The goals of protecting against the adversary can then be used to produce a formal and provable characterization of security. The issue of importance here is that the adversary is attacking the discipline practitioner’s privacy, but the practitioner is not attacking the adversary back.

As a non-cyber security example, in academic peer review ideas are “attacked” to seek flaws, and while the scholar being reviewed may provide defenses and justifications, they are generally not expected to attack the reviewers back.

Third, an adversary modeling discipline involves developing a representation or theory about an adversarial relationship. The theorist does not participate in that relationship.

As a non-cyber example, consider an ecologist studying lynx-hare population dynamics in the wild [57]. Clearly such a predator-prey system is adversarial, but while the lynxes and hares are adversaries of each other, they are not adversaries of the ecologist.

As a cyber example, consider Nguyen et al.’s application of Stackelberg games to security problems [52]. This work is motivated by particular specified use cases, but the focus is on developing general methods and principles. At one extreme, game theoretic efforts such as von Neumann’s seminal paper [78] may have no identifiable applications to a specific adversary. At the other extreme, even the theoretical papers can be used in adversary interactive contexts. We need a way to distinguish between adversary modeling and adversary interactive descriptions.

So we have defined several adversarial descriptions. But how can we determine the right description for a method?

We begin by considering how Wiener describes adversarial methods:

“The distinction between the passive resistance of nature and the active resistance of an opponent suggests a distinction between the research scientist and the warrior or the game player. The research physicist has all the time in the world to carry out his experiments, and he need not fear that nature will in time discover his tricks and method and change her policy. [80, pg 36]

In Wiener’s non-adversarial setting, nature is not attempting to disrupt the research physicist’s mission, so the physicist’s task is easier. They have “all the time in the world” to carry out their experiments and, presumably, to validate their experimental methods. But in Wiener’s adversarial setting, the adversary will undermine the methods of the warriors and
game players. This adversary will do so by observing the “tricks and methods” used and intentionally invalidating them.

We call this the vulnerability property of methods: the extent to which an adversary can intentionally invalidate a method’s ability to support a mission, for a defined perspective and scope. We can use this to determine an adversarial description, as shown in Figure 1. First, we define a vulnerability criterion: is the method vulnerable to an adversary? If the vulnerability criterion is met, we then ask if the method is part of an effort against that adversary who is invalidating it. If so, then the method is described as adversary interactive; if not, then it is adversary present.

If the vulnerability criterion is not met, then we ask if the subject of the method’s effort is an adversarial interaction. If so, then the method is described as adversary modeling; if not, then it is non-adversarial.

The vulnerability property is defined “for a defined perspective and scope” because adversarial descriptions are relative to their circumstance of use.

In terms of perspective, consider a terrorist who mails explosive devices to university scholars. Any scholar could become the target of such an attacker, but it does not follow that all scholarly methods are inherently adversarial. Rather, this introduces the subjective aspect of adversariality: that an adversarial characteristic is limited by perspective. From the terrorist’s perspective, the scholars may indeed all be part of a political or cultural conflict. From the scholars’ perspectives, this may not be the case. Therefore there are at least two subjective interpretations of the adversarial situation. The terrorist’s perspective may seem more comprehensive, but consider a third-party intelligence operative who is observing this activity, unbeknown to the terrorist and scholar. This third party’s perspective is even more comprehensive.

In terms of scope, consider a truck that is speeding through a combat zone, occasionally swerving to avoid any possible enemies. An enemy aircraft is tracking the truck unbeknown to it, waiting for the right shot. In this encounter-level scope, the truck is adversary present. But if the truck is carrying munitions to an anti-aircraft weapon that is tracking the enemy aircraft, then in this operation-level scope the truck is adversary interactive.

We conclude this subsection with some additional examples.

Example 1: adversarial shift. Consider an ecologist studying lynx-hare interactions for purposes of maintaining stable populations in a wildlife preserve. In this ecologist’s context, their studies are adversary modeling. If a cyber security researcher read the ecologist’s studies and applied its theoretical insights to an active cyber conflict, then this use would be adversary interactive. The opponent targeted by this method might also read the ecologist’s studies and change their policy to invalidate their use, but this would only invalidate the ecological studies’ adversarial application, not their original use. This is because the opponent is only targeting the adversarial adaptation of ecological methods, not the original ecological use of the methods.

Fancifully, the opponent may even decide to disrupt the ecologist’s studies, perhaps secretly corrupting the ecologist’s data to ultimately mislead the security researcher. Such a disruption would be in the scope of the adversarial use of the ecological methods. The ecologist, focused on their scope of lynxes and hares, may nevertheless be confused by the unexpected data. This confusion is due to their subjective perspective, and is a result of the adversarial shift that occurred when their methods moved from a non-adversarial to an adversary interactive use. The security researcher might thus inform the ecologist that their methods are being used adversarially, and that validating a vulnerable method involves considering its invalidation by an intelligent adversary.

At this point, the ecologist could reject the cyber security context, and try to return to the scope of their non-adversarial studies. But even then, they would be wise to change their methodology to protect their data from any further interference, in the same way that they might protect their credit card data for example. A well-constructed methodology in this case would be adversary present, since it would be on guard against any possible adversary. Alternately, the ecologist might become enraged by the threat to their research and dedicate themselves to the conflict, focusing their efforts not only on protecting their data but on producing studies that assist the cyber security researcher in their conflict. In this case, the ecologist would have adversary interactive methods.
Example 2: *intentionality and the vulnerability property.* A botanist’s studies of flora in a wildlife park are adversary present, because to complete their studies the botanists need to maintain their safety in the presence of predators such as tigers. The botanist may also need to maintain their safety during any floods that may happen, but the floods will not invalidate their efforts intentionally in the same way that a tiger will. So if the botanist stays within a radius of high ground and shelter, the flood will not seek some way of overcoming this. But if the botanist stays within a radius of a safe truck, the tiger may learn to stealthily creep between the botanist and the truck before attacking.

This is the first type of intentionality: the purposeful, reactive nature of the attack, which might be formally represented with a Belief-Desire-Intention [67, §4.2.1] or similar model. This type of intentionality is part of what makes the adversary intelligent, and is part of the vulnerability property of methods. It is represented in the first question in Figure 1.

The second type of intentionality relates to whether the tiger intends to stop the botanist from examining the flora. In other words, was the vulnerable mission in question, the study of the flower in this case, the intended target of invalidation? In fact, the tiger attacks the botanist out of hunger rather than a fierce determination to save the flora from examination. The tiger does not care about the botanist’s mission, any more than a network intruder cares about the many missions of people whose traffic they are sniffing (and confidentiality they are violating) while searching for a particular target. This second type of intentionality is not a necessary part of the vulnerability property of methods, although it characterizes the adversarial relationship in terms of the particular missions that are in conflict.

Example 3: *subjective effects on adversarial characterizations.* Consider a network intruder who is illicitly sniffing traffic for financial information. If a legitimate user of that network decides to encrypt their email out of a nonspecific wish for privacy, they are not attacking the mission of that specific intruder, but rather, of any such intruder. Likewise, that intruder is not attacking the confidentiality-preserving mission of that specific user, but of any such user. So they are in an adversary present relationship with each other.

If the intruder decides to target that user and bypasses their encryption implementation by installing a keylogger on the user’s workstation, then the intruder has entered an adversary interactive relation with the user, while the user is still adversary present.

The intruder may assume that the network’s administrator is logging various types of activity and scanning various files for malware. The intruder and the admin are attacking the missions of any such adversary. So they are in an adversary present relationship with each other.

Imagine the admin detects the keylogger and begins to carefully monitor their network traffic. The admin is reacting to the intruder’s actions, so the admin is adversary interactive with the intruder, though the intruder is still adversary present.

The admin notifies the user about the keylogger, so the user begins using a different workstation and email account; that user is now also adversary interactive, because they are reacting to that specific intruder’s actions.

Eventually the admin and user may believe the intruder’s software has been removed from their machines, so they once again assume they are adversary present. The intruder, having noticed the keylogger removal and the email account change, is now adversary interactive and is planning their next move.

In this way the adversarial relationship iteratively changes, and the engineered context changes with it. The next two sections describe this.

B. Engineered

Cyber security involves both engineered and physical aspects, but is dominated by the engineered.

In this context engineered and physical are defined in terms of the mechanistic approach in philosophy of science, as applied to computer security by Hatleback and Spring [34] [69].

Mechanisms are organized entities and activities that produce regular changes in a process [34, pg 443] [48, cited]; a definition that supports a philosophical perspective on the sciences. Hatleback and Spring use this to distinguish between engineered and physical mechanisms in computer science, and in computer security. Engineered mechanisms are artificial, such as DNS schemes and DNS cache poisoning attacks, and physical mechanisms are natural, such as the cognitive aspects of human-computer interaction [34, pg 444-446]. This is contrasted with an engineering vs. scientific distinction [65], which precludes the possibility of considering a scientific approach to its engineering aspect and which lacks the clear connection to the mechanisms involved [34, pg 445-446]. A similar argument can be made against distinctions such as research vs. development [3] or applied research vs. basic research [14].

The engineered mechanisms are those which can be changed [34, pg 445], and both sides of the adversarial interaction will change the engineered environment to further their mission and impede the other’s mission.

For cyber security, the engineered environment is the set of all implemented mechanisms. It is constantly changing because of the security struggle as well as due to the evolution of technology. A struggle takes place in the engineered environment, each participant seeking to gain control of the environment and to modify it to their benefit. Although the physical mechanisms cannot be changed, they act as constraints in which the struggle takes place, and the adversary that understands them best will be best positioned to maneuver around them.

Generally speaking, if two participants are in conflict in an engineered environment, one participant will modify the environment until it is suitable enough for their needs, at which point the other participant will do the same, and so on. This creates an endless adversarial cycle. In cyber security this has been called an “arms race” [17] [71] [53], a “cat and mouse game” [18] [23], or a game of “whack-a-mole” [2] [13] [66]. The cyber security methods that are relevant to one iteration
of the cycle may not be relevant to the following cycle. In fact, a competent adversary will ensure that this is the case. 

C. Contextual

Cyber-security is contextual: characterized by the state of the adversarial conflict, and the state of the engineered environment, for a particular time and place.

How is the context defined? By the adversarial state: what each participant is doing, what their plans are, what they hope to achieve. Also by the engineering state: what the combatants are struggling for control over, who currently controls what aspect of the environment, and how much each side understands of the situation.

Because this context is adversarial, its exact state cannot be known with certainty. Consider a security practitioner who wants to know if a USB drive contains malware. Perhaps they study it with a software tool. But how do they know if the tool, or the operating system it runs on, is not infected in such a way as to lie to them? Perhaps they install the operating system directly from the manufacturer, and compile the tool from source code without connecting to the internet. But how do they know if the compiler, operating system, or even hardware were not compromised with malware by the manufacturer, by an insider threat perhaps? Even if the practitioner were able to construct their own hardware from sand and plastic and write their own tool in binary, they still might make a design or programming mistake enabling a malicious exploit. Of course practically speaking at some point the practitioner is satisfied enough with the relative safety of their tools to proceed with their analysis, but the fact remains that their sense of security has no absolute foundation. In Section III-C we will see similar epistemological regress problems.

Not only is context changed by the adversary, but it also evolves under the pressure of consumer demand, business product creation, software and hardware innovations by entrepreneurs, and so on. So the context changes in unpredictable ways. As time goes on, the amount of change increases. And a method that was valid at one point in time may not be valid in another. That likelihood of contextual invalidation increases as time goes on.

To consider ways of managing this, we will consider the types of cyber security methods involved.

III. METHODS IN CYBER SECURITY

Our focus is on validating methods, so we need to consider the three types of methods involved in cyber security.

1) Technical
2) Operational
3) Theoretical

Because the engineered environment of any particular problem varies, one set of methods relates to solving problems related to the current engineered environment. These are the technical methods.

And because the adversary and the state of the adversarial struggle vary, one set of methods relates to deploying those technical resources to achieve goals in the current adversarial context. These are the operational methods.

Finally, confronted by seemingly endless adversarial cycles, cyber security practitioners occasionally see the need to systematize their discipline, to generalize problems across contexts. So one set of methods relates to exploring the theory and the science of the discipline. These are the theoretical methods.

In Section III-A we describe ways of validating technical methods. This includes considering operational invalidation, so we describe operational methods in Section III-B. However, the technical validation methods are meta-methods, which themselves need to be validated. Some basic concepts to do so are described in Section III-C, and validating such meta-methods is described in Section IV.

A. Technical Methods and their Validation

The continually changing engineered environment provides an endless source of technical problems to be solved by security practitioners. Although this creates an exciting amount of interesting work to perform, it also is inherently problematic: having implemented a technical method and having tested for technical invalidation (e.g. by performing unit tests, processing sample data, etc.) the security practitioner is often encouraged to hurry on to the next technical problem, without systematically validating the method against vulnerabilities.

This includes a vulnerability to contextual invalidation: that the engineered environment, either through natural technological innovation or through the efforts of an adversary, changes in a way that the technical method is no longer usable. How should this be dealt with? The security practitioner might consider the ways in which contextual invalidation could happen: for example, by stating what developments in the engineered environment would invalidate the assumptions that the technical methods depend on. The practitioner might also state ways to detect when such an invalidation has occurred, and how they would react upon detecting this.

The security practitioner should also realize that because of the adversarial shift described in Example 1 of Section II-A, their technical methods are now part of an adversarial struggle. Practitioners may also describe their technical method in terms of managing the adversarial cycle that is part of the danger of contextual invalidation. To what extent does the method “slow” the iterating of this class of adversarial cycle?

| TABLE I: VALIDATING TECHNICAL METHODS |
|----------------------------------------|
| Vulnerability                        | Validation                  |
|----------------------------------------|
| contextual invalidation               | considering, detecting, reacting |
| operational invalidation              | managing adversarial cycle   |
|                                       | considering                  |

A second type of vulnerability is suggested by our previous use of the term invalidation. The vulnerability property in Section II-A referred to the invalidation of a method’s ability “to support a mission.” A method that is useful for one mission
Operational invalidation occurs when a mission is not conducive to the effective use of a particular method. The security practitioner might manage this by considering the operational parameters required by the method, such as assumptions, use cases, and potential misuses.

B. Operational Methods

The previous section argued that technical methods in cyber security are ultimately used to support operations. So we should be specific about the types of operations we are talking about. We will use a taxonomy derived from Rid [61], with the following categories:

\[
\begin{align*}
\text{Criminal} & \quad \text{Espionage} \\
\text{Political Violence} & \quad \text{Sabotage} \\
\text{Military} & \quad \text{Information Operations}
\end{align*}
\]

We will consider methods related to all participants in the adversarial struggle: for example, both criminal methods and law enforcement methods. Of course we do not condone criminal methods, but to understand the conflict we must understand the crime, and be able to categorize a level of threat.

1) Criminal: This category of operation is the one most people are likely to encounter, as it involves illegal activity between individuals and organizations, not nations or insurgent groups which wish to defeat nations. Political violence or military activity may break international norms or even international law, and therefore be considered crimes, but for our purposes we will consider an activity criminal if it is breaking a nation’s laws and it primarily involves victims within that nation.

Criminal methods include ransomware, identity theft, and other cases of stealing from businesses [75] and individuals. Methods against criminal activities include “computer safety” approaches, as well as security procedures both by organizations and by home computer users [36]. This can range from help desk scripts (which might begin with, for example, “did you try rebooting?”) to government guidelines and suggestions for businesses [73], to digital forensic methods used when bringing criminals to trial [10].

2) Political Violence: This involves conflict between nations or insurgent groups, as opposed to individuals or organizations. Political violence includes espionage, sabotage, and information operations.

The goal of cyber espionage is “to capture data of the opposing force” and includes “exfiltration, monitoring, and theft of digital information.” [68, pg 114-115] These operations may be “advanced persistent threats” by highly-skilled teams whose work is difficult to detect [49].

Cyber sabotage operations involve, for example, attacks on national infrastructure such as industrial control systems and power grids [68, ch. 11-12].

Finally, the understanding of information operations varies somewhat by country [28]. In the US its purpose is to “influence, disrupt, corrupt, or usurp the decision-making of adversaries” [15, pg. 115] but is sometimes known as “influence campaigns” [44], “strategic messaging” [40, pg. 24], or “information warfare” [31]. Information Operations have been carried out by Russia [55] based on well-developed theories [22], as well as by China [8] [11, pgs 23-25], Israel and their antagonists Hezbollah and Hamas [68, Chapter 4], and ISIS [41].

3) Military: As explored by Rattray and Healey [60], these include situations that overlap with political violence as well as those that purely support kinetic action. The U.S. military produces a large number of documents on operational methods, and also has a strong tradition in developing technical methods to support operations as well as for acquisitions [27] [58]; these methods are gradually being adapted to cyber security [26] [47].

C. Theoretical Methods

Cyber security as a field is by necessity very focused on its immediate context. Even so, practitioners occasionally ask: how good of a job am I doing at this? and: will I always just be “putting out fires”?

In recent years, security researchers have sought such theoretical methods, hoping to find fundamental principles to drive innovation and to predict the effectiveness of technical tools and operational policies [46][50][51]. Security researchers have looked to physical and social sciences for universal laws and considered whether these laws will follow rather than inspire security solutions [19][42]. Several surveys suggest improvements to methodology [64] and scientific reporting [9], and describe the nature of scientific claims [33] [74], but no consensus has yet developed [35].

This paper is motivated by these types of questions. As a point of approach, we are considering the validation of methods. In colloquial terms, we may ask: how do I know if my security methods are any good? More formally, a meta-methodology is an “investigation into principles of method and their justification” [54, pg. 80]. We use the term validation as a synonym for justification. Section III-A included one such validation, summarized in Table I.

Every method has a meta-methodology; if it is completely unexplored, it may be called “more art than science,” “a black magic,” “artisanal,” or “trade craft,” suggesting that the method is validated by an expert on a case-by-case basis, usually after trial-and-error.

A meta-methodology includes an epistemology, an ontology, and a theory of explanation.

5This is unlike technical methods in non-adversarial sciences, which are used to develop theories – to provide evidence for theories, or to falsify theories, or to compare theories, depending on the epistemological commitments of the practitioners. See Section III-C.

6This formulation is adapted from Bevir’s [7] consideration of the meta-methodology of political science. We do not mean to suggest that using these terms should be mandatory for security professionalism. Rather, these terms are born from well-studied disciplines that can be helpful in resolving difficult questions in theoretical methods.
Epistemology involves “the nature and justification of valid knowledge [and] the procedures by which models or explanations are constructed.” [7, pg. 54] As an example of why this might be useful, consider that to validate a methodology we seek a meta-methodology. But this leads us to ask whether to validate a meta-methodology we should seek a meta-meta-methodology. Which in turn requires a meta-meta-meta-methodology and so on. Informal discourse and popular science often use the imagery of “turtles all the way down” [81] but in philosophy this is an epistemological regress problem [70, §3.3], which has been addressed since the time of ancient Greece. Resolutions include foundationalism, in which some beliefs are assumed (e.g. as with mathematical axioms); coherentism, which requires that assumed beliefs be part of a sufficiently-comprehensive system of beliefs, and infinitism, which accepts the infinite regress [21].

Additional epistemological questions include debates on the existence of a universal scientific method, and whether such a method should allow induction, or focus on validation or falsification [1].

Ontology “seeks to describe the basic categories and relationships of existence thereby providing an account of the types of things there are and the manner (or mode) of their being” [7, pg. 54]. For example, this paper makes ontological commitments regarding the properties of cyber security and has explicitly characterized adversarial interactions with a decision tree.

Explanation is a way of making “claims that go beyond our experiences and purport to discover underlying causes.” [7, pg. 54] This is the aspect that creates understanding and allows prediction. In cyber security, this is difficult due to the vulnerability property of methods: any explanation is subject to invalidation by an adversary or by the evolving engineered environment.

The next section will summarize the challenges faced by cyber security meta-methodologies, and suggest ways of managing those challenges.

IV. Validating Meta-methods

For most security practitioners, the validation of technical methods in Section III-A will suffice. But to be more comprehensive, we also need to consider the validation of meta-methods.

In cyber security, meta-methods are also vulnerable to adversarial and contextual invalidation. An adversary who learns of a given meta-method can in principle seek to modify the engineered environment, and can change their operational plans to invalidate that meta-method. Worse, at any given moment a security practitioner cannot be sure of the extent to which an adversary has already invalidated their methods and meta-methods.

For example, imagine that a researcher builds a system that measures the security of a system. The researcher wonders how they can be sure that their system works well. In other words, their system is a technical method, and they want a theoretical meta-method for validating their system.

Imagine that the researcher decides to use data to test the system, so they collect data from operational systems. Immediately they are confronted with a number of questions: how do they find a “ground truth” about whether the data they collected is in fact secure or not? How do they know if the engineered environment will change in ways that make the data obsolete? How do they know if an adversary will discover their system, and change the engineered environment or their behavior in a way that will circumvent the system without the researcher realizing it? Because if the adversary does so, then they have invalidated the researcher’s meta-method.

So, simply by asking “how do I know if it works well,” the researcher has found themselves in a dizzying theoretical environment, assailed by adversarial invalidation, on the precipice of a regression (what meta-method should validate their meta-methods, etc.), gingerly but inevitably heading towards a never-ending adversarial cycle.

But now we have surveyed the properties and the methods of cyber security. So consider the validation of technical methods described in Section III-A and summarized in Table I. That validation is a meta-method, and thus itself vulnerable to invalidation, or a regress of meta-methods. But now we have three well-studied epistemological approaches to resolving this regress. And we have the operational methods of Section III-B, each with its own tradition, with more or less technological sophistication, of validating methods and making decisions in adversarial environments. And indeed, we can seek further guidance in these adversarial disciplines.

For example, law enforcement forensic sciences are familiar with criminal efforts to invalidate their methods, and have managed questions about the validity of new digital forensics techniques [10] [38].

Another example is strategic theory, which considers war and political violence, as well as the national security impacts of crime [5]. It considers conflict that is adversarial by definition, engineered [30, pgs. 38-40], and contextual [32, pg 43]. It is built on a substantial history of operations and technical methods [58] [27], which serve as material for debates on whether there are universal laws and methods of conflict [32] [56] [76].

We are not saying that cyber security should be addressed purely from the perspective of strategic theory, the forensic sciences, or any other operational discipline. But security is interdisciplinary, and we can productively look across disciplines for methodological inspiration. In Section IV-A we will give an example of this, using Clausewitzian critical analysis from strategic theory7. After that in Section IV-B we will describe some of the characteristics of this interdisciplinary methodological endeavor.

7In strategic theory texts it has been noted that “because Clausewitz said something did not necessarily make it true, but it did make it worth considering.” [5, pgs.13].
A. Critical Analysis as a Performative Meta-Method

Clausewitz considered the relation between critical analysis, theory, and war. His basic argument is that while it may be possible to have a doctrine for how to “organize, plan, and conduct an engagement,” we cannot come up with a static set of rules or doctrine to conduct war in general, across time. Instead, theory should study “ends and means” and their related factors. The way this is done is through critical analysis, where the word critical is used in the sense of critique rather than criticism.

Clausewitz describes critical analysis as having three parts.

1) Identification of facts
2) Determination of causes of the facts
3) Investigation of the means used and the intentions behind their use

We will explore this in terms of the parts of a meta-methodology as described in Section III-C.

Epistemologically, knowledge is constructed through a study of past incidents in part 1, and continuing through and analysis of “relationships between phenomena” in parts 2 and 3. The analyst also has to consider “what might have happened” as well as what actually happened.

Determining causality in the second part can be extremely challenging so there are bound to be gaps, but “all a theory demands is that investigation should be resolutely carried on till such a gap is reached. At that point, judgment has to be suspended.”

Clausewitz discusses several additional challenges in determining causes throughout the process. In describing how analysis of causes and intentions work together,

“we can follow a chain of sequential objectives until we reach that one that requires no justification, because its necessity is self-evident. In many cases, particularly those involving great and decisive actions, the analysis must extend to the ultimate objective, which is to bring about peace.”

In this way Clausewitz resolves the regress foundationally.

The second and third parts of the process require a “working theory” to be useful, but the aim of the analysis should not be a “mechanical application of theory.” Instead, the theory should be tools to help the analyst in their judgments. If the analyst is studying a conflict and one of the combatants acts contrary to the “principles, rules, and methods” of the working theory, the analyst should investigate the reasons for this, and the analyst “has no right to appeal to theoretical principles unless these reasons are inadequate.” If that contrary behavior ends in failure, the analyst should not automatically assume that the unexpected behavior was what caused the failure. Likewise, if that contrary behavior ends in success, the analyst should not assume that the working theory was incorrect. The role of the theory is not to provide “laws and standards” but “findings” that assist in the analyst’s “judgment.”

Ontologically, the analysis is made of historical facts (individuals, units, events, actions), the causal chains of those events, and a description of those events in terms of the means that the belligerents used, and the intentions behind those selected means. Furthermore, they should be described in terms of operational language: “the language of criticism should have the same character as thinking must have in wars,” so that the analysis retain its practical value.

The theory of explanation is provided by the final analysis of causes and means which the analyst has constructed.

However, producing the explanatory analysis is not the only product of this process.

“Critical analysis being the application of theoretical truths to actual events, it not only reduces the gap between the two but also accustoms the mind to these truths through their repeated application.”

Not only is theory changed by the critical analysis, but it is meant to prepare the analyst for future, unexpected contexts. “[Theory] is meant to educate the mind of the future commander, or, more accurately, to guide him in his self-education, not to accompany him to the battlefield.” Which is not to claim that all guidelines should be rejected: it is acceptable if “the theorist’s studies automatically result in principles and rules,” but theory is “never to construct an algebraic formula for use on the battlefield. Even these principles and rules are intended to provide a thinking man with a frame of reference... rather than to serve as a guide.”

In other words, critical analysis is performative: the fact that the analyst is performing it is as important as the conclusions they reach, and performing it during an action is its own validation. After all, any pre-action theoretical conclusions may be invalidated by the adversary in the next operation. Because adversarial operations are so context-dependent, rather than simply learning rules about what to do, the combatant should learn how to solve problems in real-time, with theory to provide a guiding “frame of reference.” Critical analysis gives them practice in the skills of producing causal explanations, questioning theories when needed. This is what Clausewitz calls “creative ability” in war.

To describe this in terms we have introduced previously, the methods related to the “engagement” described at the beginning of this subsection are analogous to technical methods in cyber security: when related to “material factors,” this type of military creativity is less necessary. But there is a level at which “the intellect alone is decisive,” and methods at this level are validated through critical analysis: before the action as training, during the action as decision-making, and after the action as preparation for the next event.

B. Interdisciplinary Generative Meta-Methods

However it would be self-contradictory to claim that a strategic-theoretic critical analysis is our only answer. We have described meta-methodology as vulnerable to invalidation by adversarial action or context. We therefore benefit from generative meta-methods: ways of developing additional methods.

---

8Unless otherwise stated, the description and quotes in this subsection are taken from Book 2 Chapters 2-6 of On War [77].
Imagine these two nations become involved in an overt force-on-force cyber-only conflict [60, pg.88,92-95]. The engineered environment, distorted by the conflict and having evolved under the influence of millions of consumers, is unlike that of any year past. The combatant nations’ researchers apply their technical methods in a variety of operations. Their ability to generate effective tactics, to invalidate their adversary’s tactics, to plan effective operations and meta-methodologically estimate their chance of success, etc. are all enabled by the quality of their theories and the performative abilities of their security personnel.

The conflict ends. The rival nations and others examine the conflict, identifying what happened and what caused it to happen that way. Various theories are proposed, those are questioned, which are themselves questioned and serve for new theories, and so on.

This example is highly simplified. Realistically, such a scenario would also involve the influence of activists, international businesses, criminals, knowledgable civilians, and so on, each reacting in their own ways. However, it highlights the interaction between technical, operational, and theoretical methods.

Practically speaking, this type of theorizing can be made part of operations research, the first step of which is to perform an analysis defining the problem and constructing a model before applying the appropriate method [72, pg.8-9]. Wargaming, optimization, capability assessment are all a part of this. At this stage, when the methods are stated, so is a justification for the methods, their appropriateness to the particular context, and situations under which they would no longer be appropriate. The purpose is not to provide unarguable answers. The purpose is to provide possibilities, to support decision-making.

V. An Informal Summary

This paper tried to help us see things a bit clearer in cyber security. We’ve tried to figure out how to tell if your security system’s doing a good job.

Things work differently in security than they do in typical science and engineering because there’s an enemy, and because technology is always changing. You need to think about how the enemy might break what you’re doing, and whether things have changed so much that you’re no longer safe.

Security is not like a science where you do things to make theories. In security, you do things to support operations. So you need to think about how well those operations will work. You can never be sure, so it’s best to have a good team that can think flexibly in unexpected situations.

It might seem bad that there is no easy answer that “solves” cyber security. It might seem bad that cyber science looks more like forensic science or military science and less like a natural science. But technology is just where you work. It’s a good thing that technology isn’t the answer. The answer to security is people.

---

9Of course in doing so, humanist discussion produces material that may be useful in a given context. Philosophy in particular is similar to “pure math,” which explores relations and produces concepts and conclusions that may be useful in operational or scientific contexts, but whose main motivation is exploration, not utility [20].
ACKNOWLEDGMENTS

Thanks to Kevin Bush, Melva James, and Eric Hatleback for their insightful comments, to Jeremy Blackthorne for the conversations, and to John Wilkinson for his support and encouragement.

This material is based upon work supported under Air Force Contract No. FA8721-05-C-0002 and/or FA8702-15-D-0001. Any opinions, findings, conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the U.S. Air Force.

Distribution Statement A. Approved for public release: distribution unlimited.

REFERENCES

[1] P. Achinstein. Demarcation problem. In Routledge Encyclopedia of Philosophy, Routledge, 1998.
[2] M. H. Almeshekah and E. H. Spafford. Planning and integrating deception into computer security defenses. In Proceedings of NSPW, 2014.
[3] American Association for the Advancement of Science. Definitions of key terms. https://www.aaas.org/page/definitions-key-terms Last checked August 21, 2017.
[4] A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. IEEE Transactions on Dependable and Secure Computing, 2004.
[5] J. B. Bartholomew Jr. A survey of the theory of strategy. In U.S. Army War College Guide to National Security Issues, Volume I: Theory of War and Strategy. U.S. Army War College, Carlisle Barracks, PA, 2012.
[6] J. L. Bayuk, J. Healey, P. Rohmeyer, M. H. Sachs, J. Schmidt, and J. Weiss. Cyber Security Policy Guidebook. John Wiley & Sons, 2012.
[7] M. Bevir. Metamethodology: Clearing the underbrush. In Metamethodology: Clearing the underbrush. Routledge, 1998.
[8] M. Bristow. China’s internet ‘spin doctors’. Digital Aesthetics and Projects in Speculative Operations, volume 1, pages 120–129. IEEE, 2003.
[9] M. Burcham, M. Al-Zyoud, J. C. Carver, M. Alsaheh, H. Du, F. Filani, L. Williams. Characterizing scientific reporting in security literature: An analysis of ACM CCS and IEEE S&F papers. In Proceedings of the Hot Topics in Security of Science: Symposium and Bootcamp, pages 13–23. ACM, 2017.
[10] E. Casey. Digital Evidence and Computer Crime. Academic Press, 2011.
[11] A. Chang. Warring state: China’s cybersecurity strategy. Technical report, Center for a New American Security, December 2014.
[12] C. Cowan, S. Arnold, S. Beattie, C. Wright, and J. Viega. Defcon Capture the Flag: Defending vulnerable code from intense attack. In Proceedings of DARPA Information Survivability Conference and Expo 2003, volume 1, pages 120–129. IEEE, 2003.
[13] F. de la Cuadra. The genealogy of malware. Digital Evidence and Computer Crime, 2012.
[14] Defense Acquisition University. Research, development, test, and evaluation budget activities. https://dap.dau.mil/glossary/pages/2575.aspx Last checked August 21, 2017.
[15] Department of Defense, Dictionary of military and associated terms. http://www.dtic.mil/doctrine/dod_dictionary/ Last accessed July 11, 2017. June 2017.
[16] J. Drucker. SpecLab: Digital Aesthetics and Projects in Speculative Computing. The University of Chicago Press, 2009.
[17] M. Egele, T. Scholte, E. Kirda, and C. Kruegel. A survey on automated dynamic malware-analysis techniques and tools. ACM Computing Surveys (CSUR), 44(2), 2012.
[18] A. Endeshaw. Internet regulation in China: the never-ending cat and mouse game. Information & Communications Technology Law, 13(1):41–57, 2004.
[19] D. Evans. Workshop report: NSF/AARPA/NSA workshop on the science of security, 2008.
[20] D. Fendel and D. Resek. Foundations of Higher Mathematics: Exploration and Proof. Addison Wesley, 1990.
[21] J. Fieser and B. D. (editors). Internet encyclopedia of philosophy. http://www.iep.utm.edu/ Last checked August 18, 2017.
[22] U. Franke. War by non-military means: Understanding Russian information warfare. Technical Report FOIR-R-4065-SE, Swedish Defence Research Agency (FOI), March 2015.
[23] M. Franz. From fine grained code diversity to JIT-ROP to execute-only memory: The cat and mouse game between attackers and defenders continues. In Proceedings of the Second ACM Workshop on Moving Target Defense, pages 1–1. ACM, 2015.
[24] P. Galison. The ontology of the enemy: Norbert Wiener and the cybernetic vision. Critical Inquiry, 21(1):228–266, 1994.
[25] P. Galison. Augustinian and Manichaean science. Keynote presentation to the 2012 Science of Security (SoS) Community Meeting, available at https://cps-vo.org/file/6418/download/16361, 2012.
[26] M. A. Gallagher and M. Horta. Cyber joint munitions effectiveness manual (JMEM). M&S Journal, 8(2), 2013.
[27] D. L. Giadrosic. Operations research analysis in test and evaluation. American Institute of Aeronautics & Astronautics, 1995.
[28] K. Gales and W. H. 2. Divided by a common language: Cyber definitions in Chinese, Russian and English. In 5th International Conference on Cyber Conflict, Tallinn, Estonia, 2013. NATO CCD COE Publications.
[29] D. I. Good. The foundations of computer security: We need some. Circulated in 1986, available at http://www.ieee-security.org/CFWS/goodessay.html Last checked August 25, 2017.
[30] C. Gray. Making strategic sense of cyber power: Why the sky is not falling. Technical report, Army War College, Strategic Studies Institute, 2013.
[31] T. X. Hammes. Information warfare. In G. J. David and T. R. McKelvey, editors, Ideas as weapons: influence and perception in modern warfare, pages 27–34. Potomac Books, Inc., 2009.
[32] M. I. Handel. Masters of War: Classical Strategic Thought. Routledge, 3rd edition, 2001.
[33] E. Hatleback. The protoscience of cybersecurity. Journal of Defense Modeling and Simulation, forthcoming, 2017.
[34] E. Hatleback and J. M. Spring. Exploring a mechanistic approach to experimentation in computing. Philosophy & Technology, 27(3):441–459, 2014.
[35] C. Herley and P. C. van Oorschot. SoK: Science, security, and the elusive goal of security as a scientific pursuit. In Proceedings of the IEEE Symposium on Security and Privacy, 2017.
[36] A. E. Howe, I. Ray, M. Roberts, M. Urbanska, and Z. Byrne. The psychology of security for the home computer user. In IEEE Symposium on Security and Privacy, pages 209–223, 2012.
[37] IEEE. Call for papers, 39th IEEE symposium on security and privacy. https://www.ieee-security.org/TC/ SP2018/cfpapers.html Last Checked June 29, 2017.
[38] S. H. James and J. J. Nordby, editors. Forensic Science: an introduction to scientific and investigative techniques. CRC Press, 2003.
[39] JASON. Science of cyber-security. Report number JSR-10-102. Technical report, MITRE corporation, November 2010.
[40] M. Johnsson. NATO and the challenge of strategic communication. NDC fellowship monograph, NATO Defence College, Italy, 2011.
[41] P. Kingsley. Who is behind ISIS’s terrifying online propaganda operation? The Guardian, June 23 2014.
[42] C. E. Landwehr. Cybersecurity: From engineering to science. ACM Computing Surveys, 1994.
[43] E. V. Larson, R. E. Darilek, D. Gibran, B. Nichiporuk, A. Richardson, L. H. Schwartz, and C. Q. Thurston. Foundations of effective influence operations: A framework for enhancing army capabilities. Technical report, RAND Arroyo Center, Santa Monica CA, 2009.
[44] U. Lindqvist and E. Jonsson. How to systematically classify computer security needs some. Circulated in 1986, available at http://www.ieee-security.org/CFWS/goodessay.html Last checked August 25, 2017.
[45] T. Longstaff, D. Balenson, and M. Matties. Barriers to science in security. In Proceedings of the 26th Annual Computer Security Applications Conference, pages 127–129. ACM, 2010.
[46] D. MacEslin. Methodology for determining EW JMEM. U.S. Army Combat Capabilities Development Command-Redstone Research, 2012.
[47] T. X. Hammes. Information warfare. In G. J. David and T. R. McKelvey, editors, Ideas as weapons: influence and perception in modern warfare, pages 27–34. Potomac Books, Inc., 2009.
[48] P. Machamer, L. Darden, and C. F. Craver. Thinking about mechanisms. Philosophy of Science, forthcoming, 2017.
[49] J. A. Marquis and J. A. Marquis. Philosophy & Technology, 27(3):441–459, 2014.
[50] M. Johnsson. NATO and the challenge of strategic communication. NDC fellowship monograph, NATO Defence College, Italy, 2011.
[51] P. Kingsley. Who is behind ISIS’s terrifying online propaganda operation? The Guardian, June 23 2014.
[52] C. E. Landwehr. Cybersecurity: From engineering to science. The Next Move, 19(2), 2011.
[53] E. V. Larson, R. E. Darilek, D. Gibran, B. Nichiporuk, A. Richardson, L. H. Schwartz, and C. Q. Thurston. Foundations of effective influence operations: A framework for enhancing army capabilities. Technical report, RAND Arroyo Center, Santa Monica CA, 2009.
[54] U. Lindqvist and E. Jonsson. How to systematically classify computer security needs some. Circulated in 1986, available at http://www.ieee-security.org/CFWS/goodessay.html Last checked August 25, 2017.
[55] T. Longstaff, D. Balenson, and M. Matties. Barriers to science in security. In Proceedings of the 26th Annual Computer Security Applications Conference, pages 127–129. ACM, 2010.
[56] D. MacEslin. Methodology for determining EW JMEM. U.S. Army Combat Capabilities Development Command-Redstone Research, 2012.
[57] P. Machamer, L. Darden, and C. F. Craver. Thinking about mechanisms. Philosophy of Science, forthcoming, 2017.
