Challenges in cybersecurity: Lessons from biological defense systems

Edward Schrom (a), Ann Kinzig (b), Stephanie Forrest (c,d,e), Andrea L. Graham (a,e), Simon A. Levin (a*), Carl T.Bergstrom (f), Carlos Castillo-Chavez (g), James P. Collins (b), Rob J. de Boer (h), Adam Doupée (i), Roya Ensafi (j), Stuart Feldman (k), Bryan T. Grenfell (a,l), J. Alex Halderman (j,m), Silvie Huijben (b), Carlo Maley (o,c), Melanie Moses (q,r,e), Alan S. Perelson (s,e), Charles Perrings (b), Joshua Plotkin (t), Jennifer Rexford (u), Mohit Tiwari (v)

a Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544,
b School of Life Sciences, Arizona State University, Tempe, AZ 85287,
c Biodesign Center for Biocomputation, Security and Society, Arizona State University, Tempe, AZ 85287,
d School of Computing, Informatics and Decision Sciences Engineering, Arizona State University, Tempe, AZ 85287,
e Santa Fe Institute, Santa Fe, NM 87501,
f Department of Biology, University of Washington, Seattle, WA 98195,
g Retired Professor, Arizona State University, Tempe, AZ 85287,
h Theoretical Biology and Bioinformatics, Utrecht University, 3584 CH Utrecht, The Netherlands,
i Center for Cybersecurity and Digital Forensics, Global Security Initiative, Arizona State University, Tempe, AZ 85287,
j Department of Electrical Engineering and Computer Science, Computer Science and Engineering Division, University of Michigan, Ann Arbor, MI 48109,
k Schmidt Futures, New York, NY 10011,
l Princeton School of Public and International Affairs, Princeton University, Princeton, NJ 08544,
m Center for Computer Security and Society, University of Michigan, Ann Arbor, MI 48109,
n Center for Evolution and Medicine, School of Life Sciences, Arizona State University, Tempe, AZ 85287,
o Arizona Cancer Evolution Center, Arizona State University, Tempe, AZ 85287,
p Department of Computer Science, University of New Mexico, Albuquerque, NM 87131,
q Department of Computer Science, University of New Mexico, Albuquerque, NM 87131,
r Department of Biology, University of New Mexico, Albuquerque, NM 87131,
s Theoretical Biology and Biophysics Group, Los Alamos National Laboratory, Los Alamos, NM 87545,
t Department of Biology, University of Pennsylvania, Philadelphia, PA 19104,
u Department of Computer Science, Princeton University, Princeton, NJ 08540,
v Department of Electrical and Computer Engineering, University of Texas, Austin, TX 78712
*Corresponding Author: Simon A. Levin, Department of Ecology and Evolutionary Biology, 106A Guyot Hall, Princeton University, Princeton, NJ 08544, USA, (609) 268-6880, slevin@princeton.edu

Simon A. Levin
ORCID No.: 0000-0002-8216-5639
Carl T. Bergstrom
ORCID No.: 0000-0002-2070-385X
Carlos Castillo-Chavez
ORCID No.: 0000-0002-1046-3901
James P. Collins
ORCID No.: 0000-0003-1022-9952
Rob J. De Boer
ORCID No.: 0000-0002-2130-691X

Bryan Grenfell
ORCID No.: 0000-0003-3227-5909
Stephanie Forrest
ORCID No.: 0000-0002-5904-1646
Andrea L. Graham
ORCID No.: 0000-0002-6580-2755
Silvie Huijben
ORCID No.: 0000-0002-3537-1915
Ann Kinzig
ORCID No.: 0000-0001-7265-0289
Carlo Maley
ORCID No.: 0000-0002-0745-7076
Melanie Moses
ORCID No.: 0000-0002-8848-9554
Alan S. Perelson
ORCID No.: 0000-0002-2455-0002
Charles Perrings
ORCID No.: 0000-0002-4580-3697
Joshua Plotkin
ORCID No.: 0000-0001-6232-7613
Edward Schrom
ORCID No.: 0000-0002-1793-6433

Abstract
Defending against novel, repeated, or unpredictable attacks, while avoiding attacks on the ‘self’, are the central problems of both mammalian immune systems and computer systems. Both systems have been studied in great detail, but with little exchange of information across the different disciplines. Here, we present a conceptual framework for structured comparisons across the fields of biological immunity and cybersecurity, by framing the context of defense, considering different (combinations of) defensive strategies, and evaluating defensive performance. Throughout this paper, we pose open questions for further exploration. We hope to spark the interdisciplinary discovery of general principles of optimal defense, which can be understood and applied in biological immunity, cybersecurity, and other defensive realms..

Introduction

Securing cyber-systems is one of the central challenges of the 21st century. Within the past five years, cyber attacks have disrupted U.S. oil supplies, leaked personal data of 50 million cell phone users, and rerouted Ukrainian Internet traffic through Russian communication infrastructure, just to name a few examples. Future consequences could be even more catastrophic, from severe disruption of financial markets to the demise of democratic governments to inadvertent nuclear war. Although cybersecurity experts have made tremendous progress enhancing the security of computers[1] and networks over the course of decades, attackers often appear to be one step ahead, rapidly deploying innovative methods to overcome the latest defensive strategies, and we are still devising piecemeal solutions. Continued creative inspiration for new principles and designs of defensive systems is both timely and likely to be valuable.

A promising source of this inspiration is the study of biological immune systems. As National Medalist of Technology Carver Mead notes, "As engineers, we would be foolish to ignore the

[1] We use 'computer' as a shorthand for computing devices, including routers, servers, smartphones, etc.

lessons of a billion years of evolution." Indeed, the deep history of coevolution between parasites and vertebrate hosts produced a fully distributed immune system that deploys a remarkable diversity of defenses against an equally remarkable diversity of parasitic attackers, from viruses less than 10nm long to tapeworms exceeding 10m (Jackson et al 2009). The challenges faced by the vertebrate immune system share many key similarities with those faced in cybersecurity: both systems must recognize attackers that are diverse, dynamic, and evolving; both must root out these attackers without excessive waste or damage to self; both must handle uncertainties about when, where, and how attacks will occur; and both must be effective at the scales ranging from individuals (e.g. a single human or a single computer) to populations (groups of humans, networks of computers).

Both biological immunity and cybersecurity are examples of complex adaptive systems (CASs), in which patterns at high levels of organization emerge from localized interactions and selection processes operating on diverse agents at lower levels of organization, and feed back to affect those lower-level processes (Levin 1998). In immunity, it is the self-organized interactions of numerous cells and molecules that collectively dictate organism- and population-scale infection outcomes. In cybersecurity, analogous interactions of hardware, code, and human users collectively dictate security successes and failures, even at national and global scales. Moreover, while the ability to freely engineer computer systems appears to contrast with the constraints of evolutionary processes that occur across generations, engineering and evolution may ultimately share more similar dynamics and outcomes than most observers would expect [Box 1]. This suggests that strategies for defense that have been optimized by billions of years of evolution may also succeed in the engineered context of cybersecurity.

Looking to immunology for insights into computing security is not a new idea (Forrest et al. 1994; Kephart et al. 1995; Forrest et al. 1996; Hofmeyr et al. 2000, Wooley and Lin 2005). For example, intrusion detection systems (IDSs) originally monitored computers for malicious activity using a process called signature detection, in which patterns of system activity were compared to a database of known intrusive patterns. In 1996 an anomaly-detection system inspired by vertebrate immunity (Forrest et al. 1996) was created to instead automatically learn normal system behavior via direct observation and to respond adaptively to unfamiliar patterns, eliminating reliance on a database of predetermined patterns. Subsequently, to further lower the rate of damaging false positives, process Homeostasis (pH) was invented (Somayaji and Forrest 2000) which mimics costimulatory receptors on T cells—an important mechanism to prevent false-positive immune responses. As IDS were designed for entire networks, further immune-inspired features were incorporated, such as negative selection of detectors (for flexible distributed execution), a secondary response (to respond to previously seen attacks more quickly), diversity of pattern presentation (to avoid single points of failure), and avidity (to further control false positives) (Hofmeyr and Forrest 2000).

Nonetheless, these success stories are decades old: the many developments in both computing and biotechnology since then warrant a fresh look at how insights from immunology could be leveraged to better protect computer systems. For example, cloud-based computing applications are increasingly built from self-sufficient containers, which can interact, reproduce, and be destroyed, just like biological immune cells. This suggests that other aspects of the evolved immune system could be replicated in cybersecurity settings. Furthermore, the medical utility of burgeoning biotechnologies from CRISPR/Cas9 to mRNA vaccines are revealing new principles of immune action and suggest new interventions more generally in CASs.

Therefore, our goal is to renew interest in the following question: How can biological immunity reveal general principles of optimal defense, which might be successfully applied to provide

CASs, including cybersecurity systems, a decisive upper hand against attackers? In the subsequent sections, we provide a framework for studying connections between cyber and immune defense. Each section contains several topics to guide cross-system comparisons, along with related questions to spur future research. These questions are intended to transcend mere comparisons of systems and inspire general principles, in pursuit of a unified and broadly applicable theory of defense across biological and computing systems.

I. Framing the Context of Defense

When drawing analogies between the defensive systems of biological immunity and cybersecurity, the context in which defense occurs must be carefully considered. This context includes the goals of defense, the goals of attack, and the environment in which attacks and defense occur.

The Goals of Defense. Here we primarily consider the vertebrate immune system as a model of biological defense. Having evolved by natural selection, the vertebrate immune system has only one "goal" or function, in the broadest sense: to enhance the lifetime reproductive output of the host organism. There is no direct selection towards other goals. This explains several seemingly disadvantageous aspects of biological immunity. For example, the lack of selection for post-reproductive survival may explain immunosenescence, i.e. the gradual dysregulation and dysfunction of immunity in old age (Peters et al 2019). The lack of selection for host comfort may explain why some parasites are tolerated, i.e. allowed to persist as chronic infections with their negative impacts only partially mitigated (Roy and Kirchner 2000; Medzhitov et al 2012).

At face value, cybersecurity defenses might appear to address a broader array of goals. The devices and software that protect both individual computers as well as entire networks must not only prevent infection[2], but also limit costs on several other fronts. For example, the monetary expense of installation, upgrades, and operation must not be too high, and efficient run-times of legitimate applications must not be sacrificed.

However, in reality, the apparent contrast in the breadth of goals between cyber and biological defenses is not nearly so sharp. By limiting costs to individual and institutional users, cyber defense systems are ultimately designed to attract more users and/or to enable an institution to persist successfully through time. In other words, a cybersecurity system's broad array of proximate goals largely serves the ultimate goal of ensuring continued representation in the future, analogous to natural selection. Following the same logic in the opposite direction, in order to maximize lifetime reproductive output, immune systems must meet a wide variety of proximate goals. Just as cybersecurity systems must limit costs, immune systems must not consume too much caloric energy or limiting nutrients. Just as cybersecurity systems must not hinder legitimate applications, immune systems must not interfere with other crucial biological functions. Indeed, the vertebrate immune system not only defends against parasites but participates in other biological functions, including wound repair (Ellis et al 2018), cognitive behavior (Salvador et al 2021), and more. Thus, we argue that the structure of goals is quite similar between biological and cyber defense.

Open Questions: Can parallel ultimate and proximate goals between cyber and immune defenses be measured to suggest the relative utility of analogies in specific cases?

[2] Although not all cyber-attacks involve spreading malware, we use 'infection' in the cyber context to mean any attack that compromises all or part of a cyber system.

The Goals of Attack. As with defense, evolution by natural selection in biological systems ultimately selects for the reproductive capacity of attackers. We consider as biological attackers all parasites and pathogens that infect vertebrate hosts, encompassing an enormous diversity of viruses, bacteria, fungi, protozoa, nematodes, and other organisms. Once again, the ultimate determinants of fitness (e.g. growth in the current host and transmission to new hosts) are served by a variety of proximate goals, which vary from parasite to parasite: stealing host resources (e.g. hookworm consumption of host blood (Periago & Bethony 2012)); triggering host physiological mechanisms that facilitate transmission (e.g. induction of coughing by rhinoviruses (Atkinson et al 2016) or induction of vomiting by noroviruses (Booth 2014)); manipulating host behavior (e.g. rabies driving host aggression and biting (Jackson 2016)); or even killing the host (e.g. Ebola causing hemorrhaging and death (Kucharski & Edmunds 2014)).

Cyber attacks, in contrast, can and do have a broader array of goals. These goals include stealing data or credentials, stealing or extorting money, triggering system failures, manipulating human behavior (e.g. through phishing or the spread of misinformation), or even seeking the collapse of corporations or governments. While the numerous proximate goals of cyber attacks do not always serve an ultimate goal of persistence into the future, diversity in the nature and intent of cyber attacks nonetheless mirrors diversity in the strategies of biological parasites.

Unlike defensive systems, which typically balance many goals, individual attackers often pursue only a small subset of the numerous possible goals. Thus, analogies between cyber and biological attackers must be drawn carefully. For example, a spyware cyber attack, which aims to maximize the amount of data siphoned from a computer over an extended period of time, where fitness might be measured as the volume of leaked data, may be more accurately compared to herpesviruses than to Ebola virus. Whereas the human herpesvirus Cytomegalovirus maximizes its total reproduction by escaping detection during intermittent periods of latency (Chaturvedi et al 2020), Ebola virus maximizes its total reproduction (spread) by destroying its host quickly and horrifically (with blood loss aiding in infection of others) (Sofonea et al 2018). The latter strategy would undermine the intent of spyware, which cannot continue to gather data from an incapacitated computing system.

Open Questions: What are the analogues of evolutionary fitness that can be used to understand the success of cyber attacks, or their probability of being observed again in the future?

The Role of Third Parties. Both immune systems and cybersecurity systems are embedded in wider 'ecosystems,' where attackers and defenders are not the only relevant actors. In both arenas, the interplay between attack and defense is often mediated by "third parties" - those who are unintentionally or unwittingly exploited to benefit one side or the other by enabling, exacerbating, or mitigating the threat of attack. Classic examples of third parties in human immune defense include disease vectors such as mosquitoes and ticks, as well as animal reservoirs where zoonotic infections evolve independently of human immunity (e.g. pigs and birds for new influenza strains, bats for Covid-19). At wider scales of public health, other third parties may include those who manage land use and wildlife, ship biological materials, develop vaccines, etc, as these activities all impact the risk and/or severity of infection. Third parties are equally diverse in cybersecurity, including those who produce and sell hardware, provide network connectivity, manage storage and application servers, become unwitting participants in distributed denial-of-service attacks, or simply use the internet. For example, in the 2016 U.S. election cycle, those who posted minority political opinions on social media became third parties when Russian hackers amplified their posts to distort public perception of the political climate (Berghel 2017a).

It often seems that third parties are disproportionately exploited by attackers, particularly to increase the number of victims they can target. As such, a better understanding of third-party influences is an opportunity for major improvements in defensive systems, where insights from biological defense might translate to cybersecurity settings.

Consider malaria (the disease caused by the unicellular protozoan *Plasmodium falciparum*) - a widespread cause of morbidity and mortality in many developing countries (WHO 2022). Malaria achieves high rates of transmission among human hosts via the bites of mosquitoes (*Anopheles* species specifically). Here, mosquitoes are arguably a third party exploited by the protozoan attacker of human immune systems.

In direct combat between attack and defense, the human immune system can rarely clear all the infecting *P. falciparum* protozoans from the body. Neither evolution nor drug treatments nor vaccines, even the most promising recent vaccine candidates (e.g. Datoo et al 2022; RTS partnership 2014) have yet resulted in sterilizing immunity against malaria (though there are some signs that vaccine-induced immunity in combination with drugs may come close; e.g., Pasini et al 2022). Worse yet, mosquito-borne transmission causes widespread infection in areas with temperature, precipitation, and land use conducive to mosquito breeding (Greenwood 1989). However, reliance on mosquitoes for transmission also presents unique opportunities for a different defense strategy: decreasing infection risk in the first place. Where individuals use insecticide-treated bed nets to prevent mosquito bites, malaria infection risk is significantly reduced (Galappaththy 2013; Bhatt et al 2015). Further defenses at levels of organization higher than the individual (e.g. regulating the trade of commodities that harbor mosquito eggs, reduction of mosquito habitat via reforestation, etc.) are also effective, and we discuss considerations of system scale in the next section. But bed nets alone mirror a principle that is already well-appreciated in cybersecurity: avoiding contact with malicious code altogether is the best defense. For example, exposure can be minimized by avoiding interfaces with hardware produced by third parties.

Open Questions: Does lack of control over third party behavior inherently favor exploitation by attackers? Or can probabilistic descriptions of third party behavior be leveraged to favor defense?

System Scale. As CASs, both biological immunity and cybersecurity span a range of temporal, spatial, and organizational scales. Immune systems comprise molecules that act within nanoseconds, produced by cells that interact in local tissue zones over minutes, whose interactions lead to emergent outcomes for the whole organism over many days. Similarly, cyber outcomes emerge from the action of individual pieces of code operating at multiple levels. For example, low-level assignment of Internet packet headers combined with high-level Completely-Automated-Public-Turing-Tests-to-Tell-Humans-and-Computers-Apart (CAPTCHAs) improve security in large networks. Such cross-scale activities can sometimes interact to create non-linearities in system behavior which can lead to sudden and/or unexpected outcomes, especially when faced with spatially and temporally heterogeneous attacks.

While evolution by natural selection is expected to tune cross-scale interactions to minimize sudden negative outcomes in a probabilistic sense, uncertainty in the exact nature of attacks means that catastrophe is always possible (Graham et al 2022). For example, during bacterial infection, immune cells are transported by the bloodstream to sites of local infection, where they secrete inflammatory molecular signals to help kill the bacteria. But too many sites of local infection can allow these same molecular signals to accrue in the bloodstream, rapidly expanding their spatial scale and ultimately causing septic shock and death rather than healthy

recovery (Rossaint & Zarbock 2015). In terms of temporal scales, during some respiratory infections, such as Covid-19, immune mechanisms that cause symptoms are evoked after viral shedding has already begun (He et al 2020). While this timing leads to successful recovery of the individual host with minimal tissue pathology, it also leaves the host unable to curb transmission to other sites before it is too late.

As these examples demonstrate, it is crucial to understand the potential for propagation of unanticipated effects and interference between defensive strategies operating at different biological scales (Frank 2007). These issues are equally relevant in cybersecurity settings. For example, implementation of two-factor authentication at the institutional level may actually compromise security at the level of individual computers if it lulls individual users into being less vigilant about creating strong passwords (Herley 2009). Shutting down an infected computer may prevent the spread of malware but also cause disruptions in routing across a broader network. Progress toward stronger holistic cybersecurity will require improved understanding of how specific strategies affect higher and lower scales of organization.

While mismatched scales can present a challenge for effective defense, they also present opportunities for new defensive strategies. In particular, heterogeneity at one scale of organization can be leveraged to achieve protection at another, by making the particular defenses an attacker will encounter unpredictable. The many billions of B and T lymphocytes of the vertebrate immune system each expresses a unique receptor, generated by randomized DNA somatic recombination (Burnet 1957). This hinders an evolving parasite population from anticipating which of its peptides is likely visible to immune surveillance. Because such changing or varied defensive structures are not always achievable within the body or lifetime of one host, heterogeneity can also be deployed across a population. Each vertebrate host also expresses several out of hundreds of possible major histocompatibility complex (MHC) alleles, which are responsible for presenting parasite molecular structures to the aforementioned T lymphocytes. As a result of variation across hosts, a mutation which helps a parasite escape detection in one host may actually increase its probability of detection in the next host. Just as heterogeneity among attackers creates uncertainty and makes defense difficult, so too does heterogeneity in defense strategy impose adverse uncertainty on attackers.

The benefits of heterogeneity in defense are appreciated in the realm of cybersecurity as well. Users of uncommon operating systems are serendipitously shielded from attack simply because they are not part of the largest, and therefore most attractive, pools of potential targets (Geer et al 2003). There are several examples of intentionally engineered heterogeneity such as N-variant systems (Cox et al 2006), address-space randomization (Bhatkar 2003), instruction set randomization (Barrantes et al., 2003), and platform diversity (Okhravi et al, 2012). All of these strategies leverage unpredictability, sometimes explicitly mimicking biology by 'genetically' altering each code copy or layout. However, biological immunity appears to deploy heterogeneity as a defense more ubiquitously and spanning more organizational layers than does cybersecurity. In the technology market, economic and logistical factors provide strong incentives for standardization, which can curtail the appeal of heterogeneity. This creates a tradeoff in cybersecurity which would be beneficial to break.

Open Questions: How can unpredictability in defense be generated at multiple organizational scales and which mechanisms are most effective? What dynamic cross-scale feedbacks are required to stabilize such systems? Should engineered heterogeneity be implemented across a wider range of scales in cybersecurity and how much heterogeneity is sufficient?

II. Choosing Defensive Strategies

As CASs, both immunity and cybersecurity comprise many interacting components and mechanisms. While these mechanisms are inextricably linked by their feedbacks and influences on one another, we and others (e.g. Schmid-Hempel 2012) find it useful to assort individual defense mechanisms into 5 general strategic "layers." (Table 1). Which defensive layers are used, how they are implemented, and how they are wired together into a single self-organized system, surely depends on the context of defense, as described above. Here we explore several other crucial factors that influence these choices.

Table 1. General layers of defense.

Layer	**Definition**	**Examples from Biological Immunity**	**Examples from Cybersecurity**
Avoidance	Preventing exposure to attacks	Shunning sick individuals, disgust response toward waste and detritus	Blocking access to blacklisted websites, end-to-end encryption for messages
Blockade	Preventing entry of attack upon exposure	Skin, mucous membranes, anti-viral cell states	Firewalls, passwords, cryptography
Detection	Recognizing an attack upon entry	Toll-like receptors, T cell receptors, immunoglobulin	Virus scanner, intrusion monitoring, anomaly detection.
Alleviation	Reducing the harm caused by an attack	Tissue-repair macrophages, granuloma formation	Slowing down suspicious programs, reinstalling compromised software, changing passwords, replacing infected hardware
Counterattack	Expelling or destroying the attacker	Killer T cells, inflammatory macrophages, neutrophils, B cell antibody secretion, eosinophil toxic granules	Take-down requests for counterfeit websites, censorship, content moderation

Resource Costs. Maintaining and deploying any defensive system has costs, i.e. it consumes resources that could have been used for other purposes. For example, antivirus software can increase the run-time of legitimate software, effectively reducing the computing time available for other tasks. Secreting antibodies in response to infection uses proteins that could have been invested in organismal growth or reproduction. While underinvesting in defense leaves a system vulnerable to attacks, overinvesting in defense leaves the user of the system ill-equipped to perform other important tasks. Thus, the cost of a defense strategy should be commensurate with the risks faced. Although this principle is simple, accurately following it is not, due to the difficulty in quantitatively predicting risks posed by inherently unpredictable attacks. Natural selection acting on immune systems uses the evolutionary history of attack risk to calibrate investment in defense (Urban et al 2013; Cressler et al 2015). Even so, ongoing variance or sudden shifts in attack risk often cause hosts to invest incorrectly in specific instances; for example, the mammalian immune system is prone to overproduce inflammatory cytokines, resulting in severe immunopathology (Graham et al 2022). Similarly, attack history can be used to forecast future risk in cybersecurity, but correct calibration of defenses cannot be guaranteed in every case. Infamously, after a period of relative calm and correspondingly low investment in

cyber defense, in 2017 Equifax experienced a security breach that leaked the information of 147 million people (Berghel 2017b).

If the risk of attack is difficult to predict, then detection and counterattack layers that are rapidly induced after an attack occurs may be favored over a blockade layer that is constantly active. This is partially because the former strategy consumes resources less frequently than the latter (Frank 2002; Westra et al 2015). However, other facets of resource limitation may prioritize defensive layers in the opposite order. In many cases, defense is an impure public good: successful defense of one organism benefits others in its population, and successful defense of one computer benefits others in its network. For example, one person in an office who is vaccinated against Covid-19 and regularly updates their software reduces the risk of infection for his/her coworkers in both arenas. One individual detecting and neutralizing attackers can reduce the need for other individuals to do so, reducing their resource costs and creating a 'free-rider' problem that disincentives widespread investment in defense.

Open Questions: What (combinations of) defensive layers minimize aggregate resource costs? How can limited resources best be distributed across multiple defense layers? How are such strategies affected by the broader context of defense?

Sensitivity Tradeoffs. In both organisms and computers, some ingressions are dangerous, but the majority are not (e.g., food, e-mail messages, most software updates). Fighting innocuous ingressions can be as costly as permitting dangerous ingressions. As a result, the sensitivity of defense must be carefully tuned. False positives occur when the immune system attacks an innocuous substance or its own uninfected cells, or when a cybersecurity program denies a user or authorized code legitimate access to data or other resources. False negatives occur when an immune or cybersecurity system fails to respond to a genuine attack . Reducing false positives often increases false negatives, creating a sensitivity tradeoff that constrains the design of defensive systems (Metcalf et al 2017).

Different defensive contexts call for different sensitivity levels. Users of email spam filters typically prefer never to have legitimate mail withheld, even if it means that they are exposed to some junk mail – a balance tipped in favor of false negatives. Meanwhile, managers of servers containing top secret data might prefer multiple checkpoints that slow legitimate accesses, in order to completely block illegitimate attempts – a balance tipped in favor of false positives. Interestingly, the advancement of medical technology and hygiene to treat or prevent infections (e.g. the rapid development of mRNA vaccines (Hogan & Pardi 2022)) has outpaced the treatment of autoimmunity, such that false negatives may be relatively less risky than false positives today than during earlier human evolutionary history. The optimal level of sensitivity may determine which defensive layers are chosen and how they are implemented. Generally, the more layers a defensive system uses, the more sensitive it becomes, because there are more opportunities for an ingression to be blocked, regardless of whether the ingression is harmful or benign.

Ideally, the sensitivity tradeoff could be blunted by simultaneously reducing false positives and false negatives. Several features of the vertebrate immune system accomplish this to some degree, suggesting analogous strategies for cybersecurity. Consider T cells as a detection layer. After exiting the thymus, multiple "peripheral" checkpoints throughout the lifetime of a T cell delete cells that either respond to self or fail to respond to any invader (ElTanbouly & Noelle 2021), reducing both false positives and false negatives. This suggests that ongoing learning based on continually updated signatures of self and attack could be more fully harnessed in cybersecurity. T cells that detect a perceived threat proliferate, while competing with surrounding

regulatory cells for secreted growth signal. The outcome of this competition determines whether or not a full immune response is elicited, and it is a crucial mechanism to prevent spontaneous autoimmunity (Wong et al 2021). This suggests that majority voting processes among multiple autonomous detectors, each biased toward different levels of sensitivity, may outperform a single trained detector.

Open Questions: What algorithms can simultaneously reduce false positives and false negatives? Which defensive strategies are best-suited to implement these algorithms?

Decentralization. A fundamental problem in defense is that attackers have many more frequent opportunities to update their strategies than do defenders. Cyber attackers can privately test many attack strategies before launching the best one, and parasites have much shorter generation times and larger effective population sizes than hosts. This imbalance creates a fundamental asymmetry between attacker and defender. By decentralizing the task of defense to numerous distributed autonomous agents, a defensive system can partially close this gap in evolution rate by allowing the agents to evolve as a single defensive response unfolds. Indeed, the vertebrate immune system is composed of hundreds of lymph nodes and trillions of autonomous cells, many of which (specifically B and T lymphocytes) undergo positive selection during the course of a single infection. Given the growth of large networked enterprise systems and trends toward lightweight container-based processes spread across numerous processors, cybersecurity may also begin to realize the advantages of decentralization (Hofmeyr & Forrest 2000).

With the advantages of decentralization come several challenges, which impact the holistic design of a defense system. For example, coordinated action of distributed autonomous agents requires communication among these agents. The nodes comprising modern computer networks continually exchange packets of information, and immune cells constantly secrete signaling molecules called cytokines that modulate the behavior of surrounding cells. Decentralization vastly increases the number of signaling events, and every signaling event is an opportunity for subversion (Schmid-Hempel 2008), such as spoofing or man-in-the-middle attacks. Defensive systems must expect and preempt such attacks. One approach is to base strategic decisions on the time-integrated sum of many agents' signals, where high stochastic variability is added to the signaling output of each individual agent. As a result, subversion of any individual signaling event is swamped by group-level noise and is less likely to affect the downstream decision. Indeed, parasites often spoof or sequester cytokine signals to promote ineffective immune counterattacks, but high variability in cytokine secretion rates of individual T cells can prevent such mistakes (Schrom et al 2020). Even if an attacker does successfully subvert an entire signaling axis, other approaches can mitigate the consequences. By increasing the number of signaling axes used and the complexity with which they are integrated, defensive systems can create signaling logics that are much more challenging for attackers to manipulate toward a desired outcome (Chastain et al 2012). This might partially explain the vast complexity of cytokine signaling networks (Altan-Bonnet & Mukherjee 2019).

If successful adaptations by decentralized agents are retained after an attack has been cleared and used to improve performance in the future, then the defense system is said to have learned. Whenever a defensive system persists on a longer timescale than the duration of an attack, as in both vertebrate immunity and cybersecurity, learning is desirable (Mayer et al 2016). But the optimal dynamics of learning can vary according to the attack landscape and the goals of defense, among other factors. For example, receptor repertoire updating in the vertebrate immune system follows a Bayesian scheme which optimally balances the weights it assigns to new vs. past attacks according to the sparsity of parasite molecular signatures and the expected

host lifespan (Mayer et al 2019). It is not clear how the current gold standard in machine learning—neural networks— should be optimized for cybersecurity, given the diversity of adversarial strategies that can be used to sabotage performance. For example, in a phenomenon called “catastrophic forgetting,” manipulating the order in which training samples are fed to a neural network can cause predictable downstream failures. Some progress has been made toward overcoming catastrophic forgetting by condensing individual memories into small independent units and then entrenching these units (Hurtado et al 2021). These strategies inadvertently mimic B and T cells, which are the small independent units of immune memory that become entrenched via clonal expansion and differentiation into long-lived subtypes. Further analogies should be explored to improve other facets of adversarial learning in cybersecurity.

Importantly, the examples discussed here are constrained to single layers of defense. Noisy cytokine signaling precipitates a choice between different varieties of alleviation or counterattack layers. Neural network learning has been used in detection of genuine attacks vs. innocuous activity (Sarker 2021). However, the benefits and challenges of decentralized defense certainly span across defensive layers. For example, a memory formed during the detection of a bizarre attack that bears no resemblance to normal activity might be translated into an avoidance heuristic that prevents future contact with such an attack altogether, allowing it to be deleted from the detection memory to free more space for future learning. The amount of damage caused by different attacks might drive learning to tune the balance between alleviation and counterattack in the future.

Open Questions: How can communication across different layers of defense enable holistically optimal decentralized learning? What signaling logics are needed to protect this communication from sabotage?

Complexity. As CASs, it is no surprise that immune and cybersecurity systems are themselves complex: they contain many intricately interacting mechanisms. The vertebrate immune system includes multiple mechanisms within each of the five broad layers of defense (Table 1), and even the immune systems of much simpler organisms such as corals and bacteria achieve at least four of these layers (Pinzon et al 2014, Westra et al 2015). The repeated evolution of multi-layer defense systems suggests an advantage of complexity in defense. This is echoed by the cyber principle of defense-in-depth, which says more layers and mechanisms lead to fewer successful attacks. The common intuition for this principle is independent redundancy: if one defensive mechanism fails, another can compensate for it. But pure redundancy is rarely an evolutionarily stable outcome, and must be complemented by features like diversity and modularity that provide adaptive capacity (Levin 1999); hence immune mechanisms that appear redundant in any given infection may have partially overlapping but not completely identical uses, more broadly (Nish & Medzhitov 2011). This suggests that multifaceted defense systems could evolve simply because no single mechanism can prevent all attacks, and redundancies across different mechanisms in specific cases are merely serendipitous side effects, rather than adaptative drivers.

In fact, complex interactions among multiple layers of defense could even evolve with no benefits whatsoever. The theory of constructive neutral evolution explains complexity in cell biology as the result of a ratchet. Given multiple proteins, there are more possible mutations that increase than decrease their interdependence, and the former mutations are less likely to be reversible, so random chance inevitably leads to higher degrees of interdependence (Lukes et al 2011). Scaling this argument up, the addition of each new layer or mechanism of defense means that existing layers or mechanisms are underutilized, reducing the selective pressure for

their continued independent functioning. Thus, sophisticated multi-layer defense systems could arise by natural selection without providing long-term advantages over simpler defenses, and perhaps even proving more costly in terms of resources (Frank 2007). Because the engineering of cybersecurity systems may follow similar patterns as biological evolution, deliberately or inadvertently (Box 1), it is important to understand whether defensive complexity evolves due to inherent optimality, constraints on otherwise preferable simpler systems, or the inevitability of runaway complexity.

Open Questions: Is defensive complexity ever advantageous? If so, under what circumstances, and how much is optimal?

III. Evaluating the Performance of Defense

However well designed and adapted a defense system may be, the unpredictability and continual evolution of new attacks means that monitoring and updating defense will always be necessary. In both cybersecurity and immunity, new defenses inspire new attacks and vice versa. Below we consider specific factors that are particularly useful for evaluating defensive systems and predicting their future performance.

Co-evolutionary Patterns. The invention of new attack strategies in response to new defensive strategies and vice versa is a coevolutionary process called an arms race. Both attacks and defense systems gradually become more sophisticated and potent. Arms races can follow a range of trajectories, several extreme cases of which are useful to consider. Improved defensive capability may become so deterrent that the threat of attack largely vanishes. Conversely, attackers may unleash a catastrophic assault that leaves the defensive system overrun and unable to make future updates. Between these extremes, investment in attack and defense may escalate so far that both parties pay wastefully high resource costs that exceed the actual risk and reduce overall fitness. Or attack severity may plateau at a low enough level that alleviation is more cost-effective than counterattack, leading to chronic infections that are simply tolerated by the defender.

The ability to predict which trajectory an arms race is following in real time would be extremely beneficial in the evaluation and preemptive improvement of defensive strategies. Such prediction is sometimes possible using data from vertebrate immunity, thanks to the genomic signatures left by evolving facets of attack and defense. For example, time series of viral and antibody sequences in chronic HIV infections can be used in time-shifted neutralization assays to characterize how well the immune system tracks the evolving virus, which in turn may predict patient prognosis (Nourmohammad et al 2016). Similarly detailed data documenting gradual and reciprocal changes in attack and defense strategy are available in cybersecurity settings such as vulnerability databases, records of software updates, and Internet measurement. Analogous to HIV sequencing data, records of which sites are queried, blocked, and accessed from within China have been used to quantify the performance of the Great Firewall of China, ultimately predicting whether this strategy for national censorship is likely to remain effective (Crandall et al 2007; King et al 2013).

Open Questions: Can coevolutionary patterns predict likely outcomes in cybersecurity: catastrophic attacks, unnecessary expenditure on defense, or prudent tolerance of low-risk ingressions?

Worst-Case Scenarios. Coevolutionary prediction may reveal only the most probable scenarios; however, the magnitude of the worst-case scenario, even if its probability is quite small, is also relevant for the performance of a defense system. A 1% probability of contracting

a common cold or of receiving a spam email may be acceptable, but a 1% probability of fatal systemic infection or of a compromised control system in a nuclear reactor is not. In other words, investment in defense ought to be tailored to the attack risk profile, i.e. the probability distribution across the range of possible attack severities. Experimental evolution studies in the model nematode organism *Caenorhabditis elegans* reveal that prevalent mild infection is not sufficient to warrant hosts paying high costs for defense, but deadly infection does drive the evolution of high-cost defense (Morran et al 2011). This suggests that the most important part of a risk profile for deciding defense investment is the rightmost tail: how severe are the worst-case attacks, and how probable? Unfortunately, both measures are notoriously difficult to estimate in cybersecurity, and the calculus can change over time.

Open Questions: What other information besides attack history can be used to construct accurate risk profiles? How should the uncertain right tail of a risk profile be conservatively estimated to best balance costs with prevention of worst-case scenarios?

Changes in the Context of Defense. If a defense system is carefully designed for a specific context, then unanticipated changes to that context may cause catastrophe. Thus, vigilance in monitoring not only defensive performance but also potential changes to the context of defense is essential. Contextual changes can be externally driven. For example, shifting political alliances among nations may change the origin, and thus the resources and techniques available, for cyberattacks. Contextual changes can also be driven by the defense system itself, in the form of unintended consequences. Antibodies generated during infection with one strain of Dengue or Zika virus actually prevent *de novo* generation of antibodies during an infection with a second Dengue strain, eliminating one of the key layers of defense and typically leading to more severe disease (Katzelnick et al 2020). In another example, imagine that avoidance of malaria-carrying mosquitos via bed nets were not only to succeed, but also to impose strong selective pressure for the *Plasmodium falciparum* protozoan to survive in other biting insects, perhaps with much wider geographic ranges. The context of defense would have changed drastically - a new third actor is involved, the spatial scale of attack has changed, and many new populations of people are at risk.

Unfortunately, changes to the context of defense appear difficult or impossible to predict, especially if they are not direct feedbacks of defensive action itself. While both biological immunity and cybersecurity are CASs with many components that span spatial and temporal scales, any predictive model of their behavior must nonetheless specify relevant components, their interactions, other aspects of their context, in advance. Unknown external forces that alter these assumptions cannot be fully accounted for in model predictions.

However, some approaches may aid in the evaluation and updating of defensive systems. A purely data-driven approach is to observe the time and trajectory taken by the defensive system to return to a stable, protected state after each attack. This may reveal critical slowing down: a phenomenon in which slower and slower returns to equilibrium predict that the system dynamics are gradually approaching a tipping point, where outcomes will suddenly become drastically different (Nazarimehr et al 2020). For example, chronic inflammation during old age markedly slows the rate at which cellular debris can be cleared from tissues after infection or injury, increasing the risk of tissue degeneracy and ultimate mortality (Sanada et al 2018). More generally, the cause of a gradual shift in system dynamics - perhaps external forces changing the context of defense - can remain entirely unknown, and yet an impending catastrophe can be predicted. In cybersecurity, early detection of critical slowing down can spur periods of greater investment in explicitly researching the context of defense, to adaptively modulate efforts.

Theoretical approaches are also available, in the form of sensitivity analyses. By identifying which (combinations of) parameters exert the strongest influences on model behavior and prediction uncertainty, sensitivity analyses can highlight which components and interactions in a defensive system are likely to be least robust against external forcing from contextual changes. Knowing such vulnerabilities, even without knowing which specific contextual changes to anticipate, could suggest further safeguards to prevent sudden failures in defense.

Open questions: What design features of defensive systems make them least susceptible not only to unpredictable attacks, but even to unpredictable changes in context?

Conclusion

Across evolved biological immunity and engineered cybersecurity, we find meaningful parallels in how the defensive contexts are framed, strategies chosen, and performance evaluated. Especially as technological advances allow these two defensive systems to resemble one another more closely, we believe that carefully drawn analogies between these systems can reveal general principles of defensive design to protect against unpredictable attacks. Lists of proposed principles already exist in some fields (e.g., Bergstrom & Antia 2006, Segel & Cohen 2001), but their generality across systems has not been examined in depth, either theoretically or practically. We hope the open questions articulated above will spark collaborative study, whether by sharing data and analytical techniques or constructing theoretical models. Finally, as general defensive design principles emerge, we hope to see them vetted and successfully deployed in other realms, such as national defense against domestic and international terrorism, and public health defense against zooneses and epidemics.

Acknowledgements

The authors would like to thank Arizona State University's College of Liberal Arts and Sciences for providing the funding for the workshops that led to this paper as well as Princeton University for hosting one of the workshops. Benjamin Edwards contributed to Table 1. The authors would also like to acknowledge U.S. Army Research Office Grant No. W911NF-18-1-0325; National Science Foundation 2115075, 2211750; Defense Advanced Research Projects Agency N6600120C4020; and U.S. Air Force Research Laboratory AFRL FA8750-19-1-0501.

References

G. Altan-Bonnet, R. Mukherjee. Cytokine-mediated communication: a quantitative appraisal of immune complexity. *Nat Rev Immunol*. 19(4):205-217. (2019).

S.K. Atkinson, L.R. Sadofsky, A.H. Morice. How does rhinovirus cause the common cold cough? *BMJ Open Respir Res.* 3(1):e000118. (2016).

E.G. Barrantes, D.H. Ackley, S. Forrest, T.S. Palmer, D. Stefanovic, D.D. Zovi. Randomized instruction set emulation to disrupt binary code injection attacks. *Proceedings of the 10th ACM conference on Computer and communications security.* 281-289. (2003).

H. Berghel. Oh, what a tangled web: russian hacking, fake news, and the 2016 US presidential election. *IEEE Computer*. 50(9):87-91. (2017).

H. Berghel. Equifax and the latest round of identity theft roulette. *IEEE Computer*. 50(12):72-6. (2017).

C.T. Bergstrom, R. Antia. How do adaptive immune systems control pathogens while avoiding autoimmunity? *Trends in Ecology and Evolution*. 21(1):22-28. 2006.

S. Bhatkar, D. DuVarney, R. Sekar, Address obfuscation: an efficient approach to combat a broad range of memory error exploits. in USENIX Security Symposium (2003)

S. Bhatt, D.J. Weiss, E. Cameron, D. Bisanzio, B. Mappin, U. Dalrymple, et al. The effect of malaria control on *Plasmodium falciparum* in Africa between 2000 and 2015. *Nature.* 526, 207–211 (2015).

C.M. Booth. Vomiting Larry: a simulated vomiting system for assessing environmental contamination from projectile vomiting related to norovirus infection. *J Infect Prev.* 15(5):176-80. (2014).

F.M. Burnet, A modification of Jerne's theory of antibody production using the concept of clonal selection. *Austr. Jour. Sci.* 20, 67–69 (1957).

E. Chastain, R. Antia, C. T. Bergstrom, Defensive complexity and the phylogenetic conservation of immune control. arXiv preprint arXiv:1211.2878 (2012)

S. Chaturvedi, J. Klein, N. Vardi, C. Bolovan-Fritts, M. Wolf, K. Du, et al. A molecular mechanism for probabilistic bet hedging and its role in viral latency. *Proc Natl Acad Sci USA*. 117(29):17240-8. (2020).

B.Cox, D.Evans, A.Filipi, J.Rowanhill, W.Hu, J.Davidson, J.Knight, A. Nguyen-Tuong, and J. Hiser, N-variant systems: A secretless framework for security through diversity, in *Proceedings of the 15th Conference on USENIX Security Symposium* Vol. 15, ser. USENIX- SS'06. Berkeley, CA, USA: USENIX Association, 2006.

J.R. Crandall, D. Zinn, M. Byrd, E.T. Barr, R. East. ConceptDoppler: a weather tracker for internet censorship. *ACM Conference on Computer and Communication Security* 7:352-65 (2007).

C.E. Cressler, A.L. Graham, T. Day, Evolution of hosts paying manifold costs of defence. *Proc. R. Soc. B* 282, 20150065 (2015).

M.S. Datoo, H.M. Natama, A. Some, D. Bellamy, O. Traore, T. Rouamba, et al. Efficacy and immunogenicity of R21/Matrix-M vaccine against clinical malaria after 2 years' follow-up in children in Burkina Faso: a phase 1/2b randomised controlled trial. *Lancet Infect Dis.* 22(12):1728-36. (2022).

S.F. Elena, R.E. Lenski, Evolution experiments with microorganisms: The dynamics and genetic bases of adaptation. *Nat. Rev. Genet.* 4, 457–469 (2003).

S. Ellis, E.J. Lin, D. Tartar, Immunology of Wound Healing. *Current Dermatology Reports*. 7:350-358. 2018.

M.A. ElTanbouly, R.J. Noelle. Rethinking peripheral T cell tolerance: checkpoints across a T cell's journey. *Nat Rev Immunol*. 21:257-67. (2021).

S. Forrest, A. S. Perelson, L. Allen, and R. Cherukuri. Self-nonself discrimination in a computer. In IEEE Symposium on Research in Security and Privacy, pages 202–212. *IEEE Computer Society Press*, 1994.

S. Forrest, S. A. Hofmeyr, A. Somayaji, and T. A. Longstaff. A sense of self for unix processes. In IEEE Symposium on Computer Security and Privacy, pages 120–128. *IEEE Computer Society Press*, 1996.

S.A. Frank. Immune response to parasitic attack: evolution of a pulsed character. J Theor Biol. 291(3):281-90. (2002).

S.A. Frank. Maladaptation and the Paradox of Robustness in Evolution. *PLoS One.* 2(10):e1021. (2007).

G.N.L. Galappaththy, S.D. Fernando, R.R. Abeyasinghe, Imported malaria: A possible threat to the elimination of malaria from Sri Lanka? *Trop. Med. Int. Health* 18(6), 761–768 (2013).

D. E. Geer, C. P. Pfleeger, B. Schneier, J. S. Quarterman, P. Metzger, R. Bace, and P. Gutmann. Cyberinsecurity: The Cost of Monopoly -- How the Dominance of Microsoft's Products Poses a Risk to Security. *Computer and Communications Industry Association.* (2003).

A.L. Graham, E.C. Schrom II, C.J.E. Metcalf. The evolution of powerful yet perilous immune systems. *Trends Immunol.* 43(2):117-31. (2022).

B.M. Greenwood, The microepidemiology of malaria and its importance to malaria control. *Trans. R. Soc. Trop. Med. Hyg.* 83 (Suppl), 25–29 (1989).

S.J. Gould, N. Eldredge, Punctuated equilibria: The tempo and mode of evolution reconsidered. *Paleobiology* 3(2), 115–151(1977).

X. He, E.H.Y. Lau, P. Wu, X. Deng, J. Wang, X. Hao et al. Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature Medicine*. 26: 672–675. (2020).

C. Herley. So Long, and No Thanks for the Externalities: The Rational Rejection of Security Advice by Users. *Proceedings of the 2009 Workshop on New Security Paradigms Workshop*, pp. 133--144, ACM. ISBN 978-1-60558-845-2.

S.A. Hofmeyr, S. Forrest, Architecture for an artificial immune system. *Evol. Comput.* 8(4), 443–473 (2000).

M.J. Hogan, N. Pardi. mRNA vaccines in the covid-19 pandemic and beyond. *Annual Review of Medicine*. 73:17-39. (2022).

J. Hurtado, H. Lobel. A. Soto. Overcoming catastrophic forgetting using sparse coding and meta learning. *IEEE Access*. 9:88279-90. (2021).

A.C. Jackson. Diabolical effects of rabies encephalitis. *J Neurovirol.* 22(1):8-13. (2016).

J.A. Jackson, I.M. Friberg, S. Little, J.E. Bradley. Review series on helminths, immune modulation and the hygiene hypothesis: Immunity against helminths and immunological phenomena in modern human populations: Coevolutionary legacies? *Immunology* 126(1), 18–27 (2009).

F. Jacob, I.M. Friberg, S. Little, J.E. Bradley, Evolution and tinkering. *Science* 196(4295), 1161–1166 (1977).

L.C. Katzelnick, C. Narvaez, S. Arguello, B.L. Mercado, D. Collado, O. Ampie, et al. Zika virus infection enhances future risk of severe dengue disease. *Science.* 369(6507):1123-8. (2020).

J. O. Kephart, G. B. Sorkin, W. C. Arnold, D. M. Chess, G. J. Tesauro, and S. R. White. Biologically inspired defenses against computer viruses. *International Joint Conference on Artificial Intelligence*, 1995.

G. King, J. Pan, M.E. Roberts. How censorship in China allows government criticism but silences collective expression. *Am Polit Sci Review.* 107(2):326-43. (2013).

A.J. Kucharski, W.J. Edmunds. Case fatality rate for Ebola virus disease in west Africa. *Lancet*. 384(9950):1260. (2014).
S.A. Levin, Ecosystems and the biosphere as Complex Adaptive Systems, *Ecosystems* 1, 431–436 (1998).

S.A. Levin. Fragile Dominion: Complexity and the Commons. *Perseus*. Reading, MA.(1999).

J. Lukes, J.M. Archibald, P.J. Keeling, W.F. Doolittle, M.W. Gray. How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life.* 63(7):528-37. (2011).

A. Mayer, T. Mora, O. Rivoire, A.M. Walczak. Diversity of immune strategies explained by adaptation to pathogen statistics. *Proc Natl Acad Sci USA.* 113(31):8630-5. (2016).

A. Mayer, V. Balasubramanian, A.M. Walczak, Mora T. How a well-adapting immune system remembers. *Proc Natl Acad Sci USA.* 116(18):8815-23. (2019).

R. Medzhitov, D.S. Schneider, M.P. Soares, Disease tolerance as a defense strategy. *Science* 335(6071), 936–941 (2012).

C.J.E. Metcalf, A.T. Tate, A.L. Graham, Demographically framing tradeoffs between sensitivity and specificity illuminates selection on immunity. *Nat. Ecol. Evol.* 1, 1766–72 (2017).

L.T. Morran, O.G. Schmidt, I.A. Gelarden, R.C. Parrish II, C.M. Lively. Running with the Red Queen: host-parasite coevolution selects for biparental sex. *Science*. 333(6039):216-8. (2011).

F. Nazarimehr, S. Jafari, M. Perc, J.C. Sprott. Critical slowing down indicators. *EPL*. 132(1):18001 (2020).

S. Nish, R. Medzhitov. Host defense pathways: role of redundancy and compensation in infectious disease phenotypes. *Immunity*. 34(5):629-36. (2011).

A. Nourmohammad, J. Otwinowski, J.B. Plotkin. Host-pathogen coevolution and the emergence of broadly neutralizing antibodies in chronic infections. *PLoS Genet*. 12(7):e1006171. (2016).

H. Okhravi, A. Comella, E. Robinson, and J. Haines, Creating a cyber moving target for critical infrastructure applications using platform diversity. *International Journal of Critical Infrastructure Protection.* 5(1) pp. 30–39 (2012).

E.M. Pasini, A.V. van der Wel, N. Heijmans, O. Klop, A-M. Zeeman, H. Oostermeijer, et al. Sterile protection against relapsing malaria with a single-shot vaccine. *NPJ Vaccines.* 7(1):126. (2022).

M.V. Periago, J.M. Bethony. Hookworm virulence factors: making the most of the host. *Microbes Infect.* 14(15):1451-64. (2012).

A. Peters, K. Delhey, S. Nakagawa, A. Aulsebrook, S. Verhulst, Immunosenescence in wild animals: Meta-analysis and outlook. *Ecol. Lett. 22*(10), 1709–1722 (2019).

J.H. Pinzón, L. Dornberger, J. Beach-Letendre, E. Weil, L.D. Mydlarz, The link between immunity and life history traits in scleractinian corals. *PeerJ* 2(4), e628 (2014).

P. Romero, F., Arnold, Exploring protein fitness landscapes by directed evolution. *Nat. Rev. Mol. Cell Biol.* 10, 866–876 (2009).

J. Rossaint, A. Zarbock, Pathogenesis of multiple organ failure in sepsis. *Crit. Rev. Immunol.* 35, 277–91 (2015).

B.A. Roy, J.W. Kirchner, Evolutionary dynamics of pathogen resistance and tolerance. *Evol.* 54(1), 51–63 (2000).

RTS,S Clinical Trials Partnership. Efficacy and safety of the RTS,S/AS01 malaria vaccine during 18 months after vaccination: a phase 3 randomized, controlled trial in children and young infants at 11 African sites. *PLoS Med*. 11(7):e10011685. (2014).

A.F. Salvador, K.A. de Lima, J. Kipnis. Neuromodulation by the immune system: a focus on cytokines. *Nature Reviews Immunology*. 21: 526-41. 2021.

F. Sanada, Y. Taniyama, J. Muratsu, R. Otsu, H. Shimizu, H. Rakugi, R. Morishita. Source of chronic inflammation in aging. *Front Cardiovasc Med*. 5:12. (2018).

I.H. Sarker. Deep cybersecurity: a comprehensive overview from neural network and deep learning perspective. *SN Comput Sci*. 2:154. (2021).

P. Schmid-Hempel. Parasite immune evasion: A momentous molecular war. *Trends in Ecol. Evol.* 23(6), 318–326 (2008).

P. Schmid-Hempel, *Evolutionary Parasitology* (Oxford University Press, 2012), p. 516.

E.C. Schrom II, S.A. Levin, A.L. Graham. Quorum sensing via dynamic cytokine signaling comprehensively explains divergent patterns of effector choice among helper T cells. *PloS Comput Biol.* 16(7):e1008051. (2020).

L.A. Segel, I.R. Cohen, Eds. *Design Principles for the Immune System and Other Distributed Autonomous Systems*. Oxford University Press. 2001.

M.T. Sofonea, L. Aldakak, L.F.V.V. Boullosa, S. Alizon. Can Ebola virus evolve to be less virulent in humans? *J Evol Biol.* 31(3):382-392. (2018).

A. Somayaji, S. Forrest. Automated response using system-call delays. *Proceedings of the 9th USENIX Security Symposium.* (2000).

M.C. Urban, R. Bürger, D.I. Bolnick, Asymmetric selection and the evolution of extraordinary defences. *Nat. Commun.* 4, 2085 (2013).

E.R. Westra, S. van Houte, S. Oyesiku-Blakemore, B. Makin, J.M. Broniewski, A. Best et al., Parasite exposure drives selective evolution of constitutive versus inducible defense. *Curr Biol.*, 25(8):1043–9 (2015).

S. Wong, K. Park, A. Gola, A.P. Baptista, C.H. Miller, D. Deep, et al. A local regulatory T cell feedback circuit maintains immune homeostasis by pruning self-activated T cells. *Cell*. 184(15):3981-97. (2021).

J.C. Wooley, H.S. Lin, Eds, *Catalyzing Inquiry at the Interface of Computing and Biology* (National Research Council, National Academies Press, 2005).

World Health Organization. World malaria report 2022. License: CC BY-NC-SA 3.0 IGO.

Box 1: Engineered vs. Evolved Systems

The most glaring difference between biological defense systems and cyber systems is how they have arisen: One system was produced by a natural evolutionary process and the other by human ingenuity. We argue that the division between these two processes is ambiguous, that modern engineering processes have more in common with evolutionary processes than is commonly believed, and that inadvertent evolutionary dynamics are particularly relevant in computer security.

At first glance, the goal-directed nature of engineering, with designs produced by intelligent beings, is quite different from biological evolution, where natural selection responds to undirected random variations and drift. For example, Jacob (1977) argues that evolution through natural selection is akin to tinkering and fundamentally different from the work of the master craftsman: "The engineer works according to a preconceived plan in that he foresees the product of his efforts," "The objects produced by the engineer approach the level of perfection made possible by the technology of the time." But no one would argue that today's computer systems approach perfection, nor that our software infrastructure, which is so vulnerable to attack, was produced according to a preconceived plan, even if, as humans, we can indeed foresee some futures.

In practice, engineering and evolution share many features, and it is often challenging to distinguish between the two. Many of today's engineered systems were produced at least in part by natural evolutionary processes. An obvious example is Arnold's Nobel Prize winning work using directed mutation in chemistry to optimize protein function (Romero and Arnold 2009). Similarly, in computing, tinkering is the norm, and clean slate design is unusual. That is, we rarely get to go back in time and redesign systems from scratch. Why? Many systems are required to maintain backward compatibility, both for communication and networking and also for user experience; it is more expensive and error-prone to redesign from scratch than to reuse existing components. This is similar to evolutionary processes, which can only "work" (evolve) with components and processes already in place, the very arguments that underlie Jacob's thesis. Despite these constraints, evolutionary processes sometimes create large shifts that can be seen on the macro scale in punctuated equilibrium (Gould and Eldredge 1972) and on the micro scale in microbes that evolve the ability to digest new carbon sources (Elena and Lenski 2003)—more akin to the large-scale shifts we might associate with foresight and design, but that require neither.

We hypothesize that simple inspection of an artifact cannot always reveal the process that produced it and that at best we can make a probabilistic guess, which prompts us to ask: What are the distinct properties of engineered and evolved systems that are reflected in the designs they produce? One can even imagine a kind of Turing test that asks how one could distinguish between a product of an

evolutionary process versus an engineered process? What are the hallmarks of each? Suppose, for example, that you were presented with an immune system, a cryptography system, and a modern enterprise software system with all of its defenses, would you be able to distinguish whether each was evolved or engineered?