# How Does the Offense-Defense Balance Scale?

Ben Garfinkel, Allan Dafoe*

May 21, 2019

**Abstract**

The offense-defense balance is a central concept for understanding the international security implications of new technologies. We ask how this balance scales, meaning how it changes as investments into a conflict increase. To do so we offer a general formalization of the offense-defense balance in terms of contest success functions. Simple models of ground invasions and cyberattacks that exploit undiscovered software vulnerabilities suggest that, in both cases, growth in investments will favor offense when investment levels are sufficiently low and favor defense when they are sufficiently high. We refer to this phenomenon as *OD-scaling*. A common mechanism is this: First, initial investments increase the attacker's ability to exploit points on an attack surface where the defender has provided relatively less coverage (*gap exploitation*). Second, beyond a certain investment level, the defender begins to saturate the attack surface, reducing any opportunities the attacker derives from differences in the two actors' patterns of coverage (*defensive saturation*). Such scaling effects might be especially important when considering the security implications of emerging technologies. For instance, progress in artificial intelligence and robotics may scale up both the number of weapons platforms that actors deploy and the number of software vulnerabilities that they can discover.[1][2]

*This work was supported by the European Research Council. We would also like to thank Al Brown, Ben Buchanan, Jeff Ding, Jon Lindsay, David Manheim, Anish Mohammed, Mike Montague, and Stefan Schubert for valuable comments.

[1]Keywords: Offense-defense theory, emerging technologies, strategic stability.

[2]Wordcount: 9998

# 1 Introduction

Consider two neighboring countries which, through steady investments, grow their armies to twice their initial sizes. How should we expect this development to affect each country's capacity to invade the other? Similarly, if investments into cybersecurity and into cyberattacks both double, should we expect successful attacks to become more or less feasible? What about investments into missile and anti-missile systems, into terror and counter-terror groups, or into biological weapons and biosecurity?

What unites these questions is a concern with how the balance between offense and defense *scales*, meaning how it changes as investment levels grow. Such questions have not been well-studied in the field of international security. However, they have theoretical and practical importance. Asking them, we will show, reveals ambiguities in classic formulations of the concept of an offense-defense balance. Answering them may also clarify the future prospects for security in a number of domains. The vertical proliferation of emerging weapons systems, such as cheap drones for use in swarms, might favor either offense or defense depending on the nature of the associated scaling effects.

This article is divided into four sections. In the first section, we consider the concept of an offense-defense balance and clarify what it means for the balance to scale. Our discussion draws upon the concept of a *contest success function* (*CSF*), which can be used to relate the expected outcome of an attack to investments made by the attacker and the defender. It is then natural to interpret the offense-defense balance associated with the attack as the relative efficacy of the two actors' investments. We consider a pair of formal metrics that describe this relative efficacy. We note that both metrics are themselves functions of the current investment level, meaning that they may change as investments grow.

In the second section, we consider two special cases of conflict: ground invasions and cyberattacks that exploit undiscovered software vulnerabilities. We present models suggesting that, in both of these cases, scaling up investments is likely to benefit the attacker when investment levels are sufficiently low and benefit the defender when investment levels are sufficiently high. We refer to this pattern as *OD-scaling*, for **O**ffensive-then-**D**efensive scaling.

In the third section, we explain this commonality. The general mechanism is this: First, initial investments increase the attacker's ability to exploit points on an attack surface where the defender has provided relatively less coverage (*gap exploitation*). Second, beyond a certain investment level, the defender begins to saturate the attack surface, reducing any opportunities the attacker derives from differences in the two actors' patterns of coverage (*defensive saturation*). We describe a set of abstract conditions that are sufficient for OD-scaling to emerge and use them to analyze additional cases. We argue that the conditions

are unlikely to be satisfied in the context of missile defense but may be satisfied in the context of defense against drone swarms.

In the final section, we argue that scaling effects are one of the core pathways by which technological progress can shift the relative efficacy of offense and defense. They may also play an increasingly large role in producing future shifts. We note that progress in artificial intelligence and in robotics will, in effect, scale up the number of weapons platforms that actors deploy, as well as the number of software vulnerabilities that they can discover. The role of this progress in reducing or exacerbating threats to security will depend to a significant extent on the associated scaling effects.[3]

# 2 Understanding the offense-defense balance

## 2.1 The offense-defense balance

Informally, the *offense-defense balance* refers to the relative ease of carrying out and defending against attacks.[4] Nearly all accounts, including both classic and contemporary ones, agree on this much. Accounts then diverge on a number of points, particularly on the concept's domain of applicability.[5] The earliest authors to discuss the concept used it to specifically describe the ease of capturing an opponent's territory through the defeat of their army. However, it has become increasingly common to apply the concept to other objectives and forms of conflict. Most notably, the cybersecurity and cyberwarfare literature includes frequent discussions of the "cyber offense-defense balance."[6]

[3]There are of course many other pathways, beyond shifting the offense-defense balance, by which emerging technologies might alter threats. Other pathways also discussed in this special issue include shifting offense-defense distinguishability (discussed by Volpe), shifting the likelihood of conflict escalation (discussed by Talmadge), and shifting the economic viability of conquest (discussed by Gartzke). Todd Sechser, Neil Narang, and Caitlin Talmadge, 'Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War'; Tristan Volpe, 'Dual-Use Distinguishability: How 3D-Printing Shapes the Security Dilemma for Nuclear Programs'; Caitlin Talmadge, 'Emerging Technologies and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today'; Erik Gartzke, 'Blood and Robots: How Remotely Piloted Vehicles and Related Technologies Affect the Politics of Violence'. For a discussion of the role of arms control in managing such threats, also in this issue, see Heather Williams, 'Asymmetric Arms Control and Strategic Stability: Scenarios for Limiting Hypersonic Glide Vehicles'.

[4]Robert Jervis, 'Cooperation under the Security Dilemma', *World Politics* 30/2 (1978).

[5]Charles L. Glaser and Chaim Kaufmann, 'What is the Offense-Defense Balance and How Can We Measure It?', *International Security* 22/4 (1998).

[6]Rebecca Slayton, 'What is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment', *International Security* 41/3 (2017); Lucas Kello, 'The Meaning of the Cyber Revolution: Perils to Theory and State-Craft', *International Security* 38/2 (2013); Erik Gartzke and Jon R. Lindsay, 'Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace', *Security Studies* 24/2 (2015); Jacquelyn Schneider, 'The Capability/Vulnerability Paradox and Military Revolutions: Implications for Computing, Cyber, and the Onset of

We will adopt this more flexible conception, under which it is sometimes more appropriate to speak of "an" offense-defense balance than "the" offense-defense balance. An offense-defense balance is defined relative to a particular objective, form of conflict, and set of actors. Most discussions of how "the" offense-defense balance has evolved over time, for instance, can be interpreted as discussions of the offense-defense balance specifically for European states attempting to capture territory.[7]

A number of scholars have argued that offense-defense balances play a significant role in determining the risk of conflict breaking out ("crisis stability") and the incentives to arms race ("peacetime stability"). Jervis argues that when the offense-defense balance in a system tilts toward offense, this exacerbates the security dilemma.[8] The easier offense is relative to defense, the more actors should feel threatened by one another and the more inclined they should be toward the use of preventative attacks as a means of reducing threats. Van Evera lists several other reasons why an offensively-tilted balance may produce instability.[9]

On the other hand, other authors have also suggested that offense-dominance within particular domains might actually support stability.[10] We will not weigh in on this debate, other than to note that on a number of different views offense-defense balances can play significant roles in shaping the frequency and character of conflict.

## 2.2 Formalism

### 2.2.1 Contest success functions

In his foundational article on the subject, Jervis suggests that there are two ways to describe the relative ease of offense and defense.[11] First, there is the relative efficacy of investments used to conduct attacks and investments used to

War'.

[7] This more flexible conception can also accommodate both dyadic and systemic offense-defense balances. A *dyadic balance* is defined relative to a concrete pair of actors (such as England and France). A *systemic balance* is defined relative to a system of actors (such as 19th century Europe) and describes the relative ease of offense and defense for something like a typical or idealized pair of actors in the system.

[8] Jervis, 'Cooperation under the Security Dilemma'.

[9] Stephen Van Evera, 'Offense, Defense, and the Causes of War', *International Security* 22/4 (1998).

[10] For an argument that offense-dominance on land might increase the variance of military outcomes, and thus cause risk-averse leaders to refrain from initiating war, see James D. Fearon, 'The Offense-Defense Balance and War Since 1648', draft paper (1997). For an argument that offense-dominance in cyberspace can support defense in physical domains, see Erik Gartzke, 'The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth', *International Security* 38/2 (2013): 67.

[11] Jervis, 'Cooperation under the Security Dilemma'.

defend against attacks. Second, there is the advantage that one gains by being the actor who moves first in a conflict.

Most subsequent analyses of the concept have focused on only the first of these two components. Glaser and Kaufman argue that the oft-ignored second component, "first-strike advantage," ought to be separated out as an analytically distinct concept.[12] We accept this suggestion and place our focus on the efficacy of investments.

The offense-defense balance is then a partial characterization of the relationship between offensive investments, defensive investments, and attack outcomes. Before attempting to give the concept a more precise definition, we find it useful to first describe what a more complete characterization would entail.

Let $A$ be a measure of a potential attacker's investments. Let $D$ be a measure of a potential defender's investments. Let $S$ be a measure of attack success. Then we can consider a function $F(A, D)$ that describes, for any pair of investments, the degree of success that the attacker can expect to achieve.

$$\mathbb{E}[S] \ = F(A, D)$$

This function is a generalization of two-player *contest success functions* (*CSFs*) that economists sometimes use to study competition and conflict. We will therefore adopt the name.[13]

There will typically be more than one plausible measure that can be used to describe the outcome of an attack. For instance, for an attempt to seize territory from another state, one might use the area of territory captured, the economic value of the territory captured, or simply a binary variable that represents whether or not the attack was a "success."

Broadly, there are also three varieties of measures one can use to describe investments into a conflict. First, one can use a measure of the *resources* invested into the conflict. Second, one can use a measure of the *capabilities* invested into the conflict. Third, one can use a measure of the *utility* lost through investments into the conflict.

Consider the case of a ground invasion. The two actors' investments could be measured in dollars spent, in Armored Division Equivalents (ADEs), or in

[12] Glaser and Kaufmann, 'What is the Offense-Defense Balance and How Can We Measure It?'.

[13] Although they have not played a significant role in the offense-defense balance literature, CSFs have been applied to the study of military conflict. The standard formalism for CSFs differs from the one we use primarily by assuming a binary success metric. Michelle R. Garfinkel and Stergios Skaperdas, 'Economics of Conflict: An Overview', *Handbook of Defense Economics* 2 (2007); James D. Fearon, 'Cooperation, Conflict, and the Costs of Anarchy', *International Organization* 72/3 (2018).

the utility cost. As there are trade-offs associated with each, no measure is universally more useful than the others. For instance, utility lost is often the quantity most relevant to strategic analysis, but is hard to measure. By contrast, dollars spent is easier to measure, but may require additional assumptions to be theoretically relevant.

### 2.2.2 Metrics of the offense-defense balance

The most common metric for the relative efficacy of offensive and defensive investments is the minimum investment ratio that would allow the attacker to achieve a particular degree of success. One example of this approach is provided by Lynn-Jones:

> *[The offense-defense balance] is the offense/defense investment ratio required for the offensive state to achieve victory. If, for example, a state must invest $3 million in military capabilities in order to mount a successful offensive against a state that invests $1 million in its military capabilities and adopts a defensive strategy, the offense-defense balance is 3:1.*[14]

Another example is provided by Glaser and Kaufman:

> *We prefer to define the offense-defense balance as the ratio of the cost of the forces the attacker requires to take territory to the cost of the forces the defender has deployed.*[15]

To extend these definitions to cover additional forms of conflict, such as cyberattacks, one can simply replace their conquest-specific language. Roughly, this is the approach taken by many authors who discuss the "cyber offense-defense balance."[16] One can also extend them to cover non-monetary measures of investment.

However, even with such modifications, these definitions are not sufficiently general. They contain strong implicit assumptions about the forms of the contest success functions that underlie conflict.

First, Lynn-Jones' definition implies that there is some absolute investment ratio that is necessary and sufficient for victory. If $3 million in offensive invest-

[14] Sean M. Lynn-Jones, 'Offense-Defense Theory and its Critics', *Security Studies* 4/4 (1995).

[15] Charles L. Glaser and Chaim Kaufmann, 'What is the Offense-Defense Balance and How Can We Measure It?'.

[16] Patrick. J. Malone, 'Offense-Defense Balance in Cyberspace: A Proposed Model', Master's thesis, Naval Postgraduate School (2012); Keir Lieber, 'The Offense-Defense Balance and Cyber Warfare', *Cyber Analogies* (2014); Ilai Saltzman, 'Cyber Posturing and the Offense-Defense Balance', *Contemporary Security Policy* 34/1 (2013).

ments is just enough to offset $1 million in defensive investments, then, given that the ratio is the same, it follows that $6 million must be just enough to offset $2 million. Glaser and Kaufman's definition manages to avoid this assumption, by fixing the ratio to a particular level of investment made by the defender. Although they do not highlight this point, their definition transforms the offense-defense balance from a constant to a function of what the defender invests.

Both definitions also contain a second implicit assumption, that success is a binary variable and that investment levels produce success deterministically. While this may often be a useful simplifying assumption, there are also many cases where it is inappropriate. An attempt to disrupt another country's infrastructure with cyberattacks, for instance, might be most naturally described in the language of uncertainty and degrees of success.

To remove this limitation, we offer the following metric as a generalization of these authors' conceptions of the offense-defense balance. For a given contest success function $F$ and threshold of expected success $S^*$, we define the *ratio-for-success* $R_{S^*}$ as a function of the defender's investments.

$$R_{S^*}(D) \equiv \frac{D}{\min\{A : F(A, D) = S^*\}}$$

This represents the ratio of the defender's investment to the minimum offensive investment that would allow the attacker to secure some expected level of success. A larger value corresponds to an easier attack. Beyond a reversal of the numerator and denominator, the Glaser and Kaufman definition represents a special case of the ratio-for-success metric, for a binary CSF characterizing the success of a ground invasion and a success threshold set at certain success ($S^* = 1$).[17]

Ratio-for-success is not the only metric one can use to assess the relative efficacy of offensive and defensive investments. For instance, Van Evera notes that, in the context of conquest, another useful metric might be "the probability that a determined aggressor could conquer and subjugate a target state with comparable resources."[18]

We introduce *success-from-a-ratio* as a generalization of this additional metric, to cover other varieties of attack, non-binary success functions, and various degrees of comparability. For a given contest success function $F$ and investment ratio $R^*$, success-from-a-ratio $F_{R^*}$ describes how well the attacker can be expected to do when the defender's investment is $R^*$ times its own.

[17] We have chosen to define the metric in terms of $\frac{D}{A}$ rather than $\frac{A}{D}$ in order to accommodate the intuition that an increase in the offense-defense balance corresponds to a tilt toward offense.

[18] Van Evera, 'Offense, Defense, and the Causes of War'.

$$F_{R^*}(D) \equiv F\left(\frac{1}{R^*} * D, D\right)$$

We regard ratio-for-success and success-from-a-ratio as complementary metrics, describing closely related aspects of the offense-defense balance. Ratio-for-success answers the question: "What ratio of investments would allow the attacker to expect a certain level of success?" Success-from-a-ratio answers the question: "How successful could the attacker expect to be if there was a certain ratio of investments?"

## 2.3 Scaling

When we ask how a metric of the offense-defense balance *scales*, we are asking how it changes as investment levels increase. Specifically, given that both of our proposed metrics take the defender's investment $D$ as their independent variable, we are asking how the metric changes as $D$ increases.

If either metric increases with $D$, favoring offense, then we say that metric undergoes *offensive scaling*. If the metric decreases, favoring defense, then we say that it undergoes *defensive scaling*. If it undergoes offensive scaling for sufficiently low investment levels and defensive scaling for sufficiently high investment levels, then we say that, overall, it exhibits *OD-scaling*.

Both metrics will typically scale in similar ways, since they capture intertwined properties. It should thus often be sufficient to analyze scaling phenomena using only one of the two metrics.

Despite often being neglected in theoretical analyses, scaling can in principle have a dramatic impact on the offense-defense balance. Take the idealized case of a defender whose preparations for a war include building a long wall along its border. Given that the attacker has the ability to focus its attacks on any gaps in the wall, a defensive investment that is sufficient to cover its entire border will be much more than twice as effective and difficult to offset as an investment that only suffices to cover half its border. The defender, therefore, can decrease the local offense-defense balance simply by spending more.

As a concrete analogue, we can consider Germany's invasion of France in the Second World War. The successful German war plan took critical advantage of the fact that the Maginot Line, the string of defensive fortifications that France built along a portion of its Eastern border, was not extended to cover either the France-Belgium or Belgium-Germany border. Historical accounts suggest cost was a significant factor in the decision.[19] It is therefore plausible that a conflict

[19] Robert Jackson, *The Fall of France: May-June 1940* (Oxford: Oxford University Press 2003), 25-27.

with double the resources invested by both sides would have been notably more favorable to the defense.

# 3 Scaling effects for invasions and cyberattacks

In this section, we analyze the scaling effects associated with two varieties of attacks and two forms of investment. We consider, in particular, the following questions:

> **Invasions:** How does an attacker's ability to break through a defender's lines scale, as both actors devote more ground forces to holding or breaking through the lines?
>
> **Cyberattacks:** How does an attacker's ability to acquire "zero day" exploits for a defender's system scale, as both actors' increase their investments into vulnerability discovery?

We approach these questions by exploring idealized models, with our appendix also considering a number of possible relaxations to these models. In both cases, we find that there are strong reasons to expect OD-scaling. Scaling up investments first benefits the attacker and then benefits the defender.

## 3.1 Invasions

### 3.1.1 Background

In this case, an attacker attempts to move ground forces into a defender's territory. Once inside, the attacker may attempt to pursue a number of possible objectives, such as conquest or liberation. The defender, in turn, attempts to prevent the attacker from successfully penetrating too deeply or from achieving its objective.

Both actors have a range of possible strategies to choose from. Analysis is simplest, though, if we make the assumption that the defender adopts a forward defense. This means that the defender places a large portion of its forces close to its border, with the aim of preventing the attacker from penetrating its territory to any significant depth. When the defender adopts a forward defense, the attacker must break through some section of the defensive line in order to advance.

It is widely held that defenders hold a natural advantage in combat. This advantage has a number of sources, including the fact that one is naturally more

vulnerable when advancing than when remaining behind cover. Its practical significance is that the attacking force must typically achieve a much greater concentration at a single point to break through. A longstanding rule-of-thumb sets the necessary concentration ratio at 3-to-1.[20]

Although there are a number of ways an actor can make investments to improve its prospects for success, we will focus specifically on investments that increase the size of ground forces. We are interested in how scaling up the sizes of both ground forces will affect the attacker's ability to break through.

The first analyst to consider this question in significant depth was Liddell Hart, who concluded that greater force levels could significantly benefit the defender.[21] His argument was that there was some force-to-space ratio necessary to provide sufficiently unbroken coverage of a front. Below this force-to-space ratio, the attacker could easily break through by concentrating its forces along weak points or gaps in the defensive line. Other analysts, particularly Mearsheimer, later extended and popularized Hart's work on the topic, often with a focus on estimating the size of the defensive force necessary to repel a Soviet invasion of NATO countries.[22]

To our knowledge, Biddle et al. provide the most in-depth analysis of the role of force-to-space ratios in determining the outcomes of ground invasions.[23] They present a complex model, involving several dozen parameters, which predicts that increasing force levels above a certain point will diminish the attacker's ability to capture territory. On the other hand, increasing force levels will bolster the attacker's ability to capture territory when force levels are sufficiently low. In our terminology, the model predicts OD-scaling.

To isolate the core mechanism behind this prediction, we now present a significantly simpler model of a ground invasion.

### 3.1.2 An idealized model

We consider a contest success function in which the output is the expected size of the ground force that is able to penetrate the defensive line. The inputs are the total sizes of the attacking and defending ground forces. Possible units include simple headcounts or Armored Division Equivalents (ADEs), which adjust for

[20] John J. Mearsheimer, 'Assessing the Conventional Balance: The 3:1 Rule and its Critics', *International Security* 13/4 (1989).

[21] Basil L. Hart, 'The Ratio of Troops to Space', *Royal United Services Institution Journal* 105/618 (1960).

[22] John J. Mearsheimer, 'Numbers, Strategy, and the European Balance', *International Security* 12/4 (1988).

[23] Stephen D. Biddle et al., *Defense at Low Force Levels: The Effect of Force to Space Ratios on Conventional Combat Dynamics* (Alexandria, VA: Institute for Defense Analyses 1991).

the quality of each force.

$$S \equiv \text{Size of force that breaks through the defensive line}$$

$$A \equiv \text{Initial attacking force size}$$

$$D \equiv \text{Initial defensive force size}$$

Our choice to use the size of the force that penetrates the defensive line as a measure of success reflects the view that, across a wide range of scenarios, it is a reasonable proxy for the attacker's ability to achieve its objectives. Its ability to capture territory or to attack remaining defenders from the flanks and rear, for instance, will typically increase with $S$. In this case we treat the attacker as attempting to maximize the expected value of $S$ and the defender as trying to minimize it.

We accept the standard assumption, discussed above, that the attacker must establish a greater local force concentration to break through. We also accept the standard assumption that there is a limit to how densely a force can fruitfully concentrate at a particular point. Beyond a certain concentration, combatants and vehicles will face risks from collateral damage to one another, increase the expected effectiveness of enemy artillery, and limit each other's mobility.

Then, to produce an idealized model, we add four simplifying assumptions. First, we assume that breaking through at a given point requires complete attrition of the defender's forces at that point and that each side's losses grow linearly with the initial size of the other's force. Second, we assume that points on the front are homogeneous: they do not vary in limits on force concentration, defensive advantage, or other such factors. Third, we assume that reinforcements do not arrive during battles. Fourth, we assume that the defender cannot anticipate where the attacker will strike and that, given this fact, both sides distribute their forces optimally.[24]

We show in our appendix that a Nash equilibrium is for the defender to distribute its forces evenly across the points, while the attacker distributes its forces to achieve the maximum practical concentration at as many points as possible.

Let $N$ be the number of points along the front where the attacker might strike. Let $B$ be the force concentration ratio required to break through and $M$ the maximum practical force concentration. Then, as shown in our appendix, the contest success function is given by the following expression:

[24]This model resembles but is distinct from "Colonel Blotto" models, a well-studied class of game-theoretic models that also concern the optimal allocation of resources across battlefields. Brian Roberson, 'The Colonel Blotto Game', *Economic Theory* 29/1 (2006).

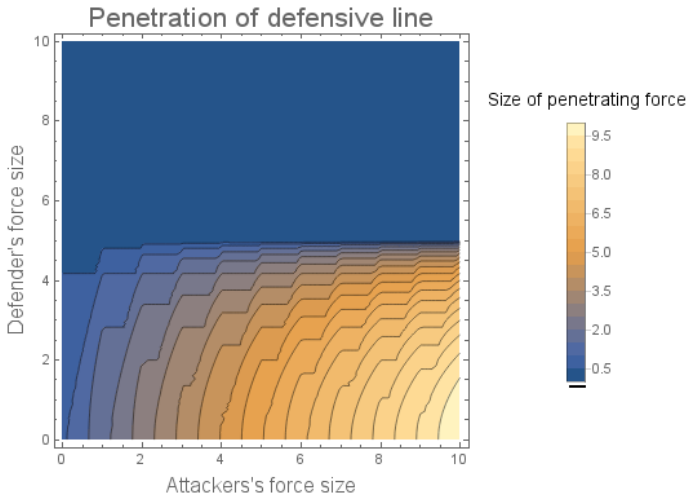


Figure 1: The CSF for our idealized model. The chosen parameters correspond to 10 possible points of attack, a 2:1 force ratio requirement for achieving breakthrough, and force size units normalized to the maximum concentration level at each point.

$$\mathbb{E}[S] = \sum_{i=1}^{N} max\Big\{0, max\big\{0, min\{M, A - M * (i-1)\}\big\} - B * min\big\{M, \frac{D}{N}\big\}\Big\}$$

Figure 1 presents a graphical depiction of the function for a representative set of parameters. Figures 2 and 3 present the relevant scaling effects more clearly, by depicting success-from-a-ratio for an equal investment ratio and ratio-for-success for a threshold of one unit of penetration. Clearly, OD-scaling is observed.

We can explain the basic mechanism responsible for OD-scaling in this case by focusing on success-from-a-ratio and considering what happens as investments increase. Initially, the attacker can use surprise to achieve much higher degrees of force concentration at a small number of points. These differences in concentration allow the attacker to break through the defender's lines. For some time, higher force levels will also enable larger breakthroughs. However, given that space is finite, the defender will eventually begin to *saturate* the front by achieving close to the maximum concentration at a large number of points. This will place increasingly strict limits on how different the two actors' patterns of concentration along the front can be. Since defense is also *locally superior* – in the sense that the defender has a natural edge in battles where concentrations are equal – the attacker's capacity to break through must decline too.

In our appendix, we consider each of the assumptions used in this simple model

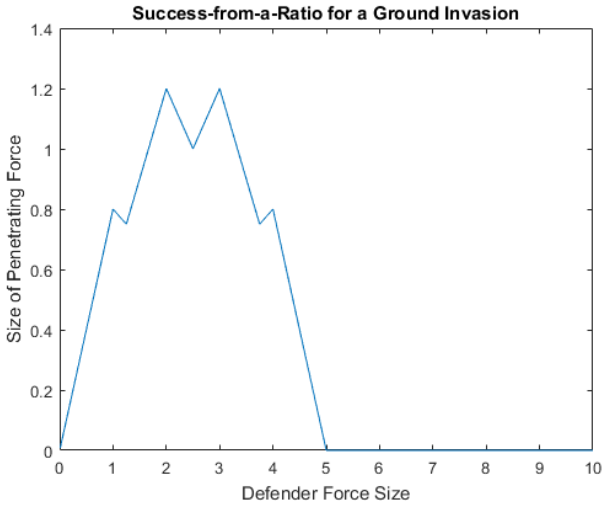


Figure 2: Success-from-a-ratio for our idealized model, for an equal investment ratio. Parameters same as above.

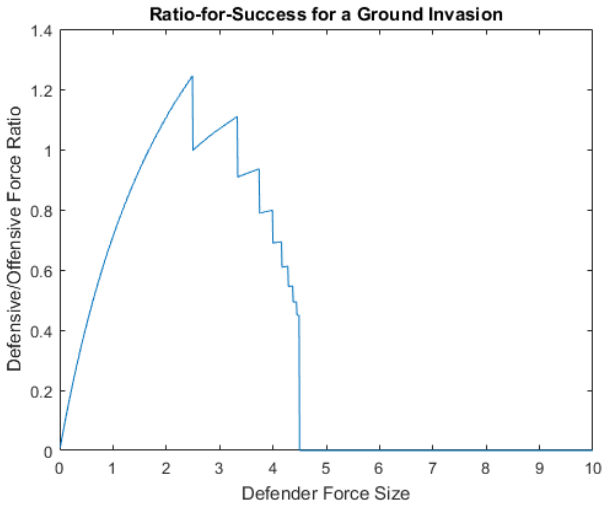


Figure 3: Ratio-for-success for our idealized model, for a threshold of one unit. Parameters same as above.

in turn. We find that although they are highly idealized, it appears to be the case that reasonable adjustments to the assumptions, for instance to allow for reinforcement dynamics and a more realistic attrition law, do not challenge the basic mechanism described here. The more complex Biddle et al. model also arrives at the same conclusion.

### 3.1.3 Case study: Normandy landings

The Allied invasion of France in the Second World War provides a useful illustration of the dynamics discussed above. In this case, the Allied forces entered France from the sea rather than over land. However, this distinction does not imply any significant change to the abstract model.

First, it was the case that the Normandy landings depended heavily on the Allies' ability to exploit superior force concentrations at the points where they landed.[25] They were able to achieve the necessary concentrations by creating uncertainty about their chosen landing points, including by running a protracted disinformation campaign, which forced the German military to spread its forces out between a larger number of plausible landing points. However, even with the benefit of a lopsided local force ratio, the inherent difficulty of amphibious landings caused the Allies to struggle and suffer several times greater casualties. In addition, despite being spread between five beaches and arriving in multiple waves, the Allies also appeared to be approaching the maximum useful levels of concentration – suffering from congestion issues, particularly, on Sword Beach.

A counterfactual version of the Normandy landings with much higher force levels on both sides could have allowed the German forces to more comfortably defend each of the landing sites they considered plausible. On the other hand, a counterfactual version with much lower force levels could have prevented the Allies from landing a large enough force to achieve their objectives on the mainland.

## 3.2 Cyberattacks

### 3.2.1 Background

A cyberattack is an attempt to exploit another actor's computer systems. A cyberattack may be associated with a number of possible objectives, including stealing confidential information, disrupting the availability of a service, or damaging connected physical objects and infrastructure.

[25] Olivier Wieviorka, *Normandy: The Landings to the Liberation of Paris* (Cambridge, MA: Harvard University Press 2008).

One particularly noteworthy class of cyberattacks involves the exploitation of *zero days*, or software vulnerabilities that are unknown to developers and users of the software. A wide range of political actors, including national governments and criminal organizations, are believed to stockpile knowledge of zero days for potential use in future offensive operations. For instance, the American/Israeli Stuxnet attack on the Iranian Natanz nuclear facility exploited four distinct zero days for the Windows operating system.[26]

We can regard any party attempting to acquire zero days for a particular system as an attacker and the set of parties with an interest in patching vulnerabilities, collectively, as the defender. An attacker acquires a zero day, then, when it discovers a vulnerability that the defender has not.

Since both attackers and defenders benefit from discovering vulnerabilities, we can ask about the effect of scaling up their discovery capabilities. In fact, this question is far from hypothetical. We should want to know, for instance, whether ongoing projects to develop more effective vulnerability discovery tools are more likely to empower attackers or defenders.

Only a small literature has examined scaling effects. One relevant but inconclusive empirical debate considers whether publishing software source code – which is a useful resource for anyone hoping to discover vulnerabilities – tends to overall benefit attackers or defenders more.[27] A related debate considers the returns to defensive vulnerability discovery. Rescorla claims that any vulnerability discovered by a defender is unlikely to have been rediscovered by an attacker.[28] (The most sophisticated analysis of vulnerability rediscovery to date, although also not conclusive, supports this claim.)[29] From this, he argues that defensive investments have only a minimal effect on an attacker's ability to acquire zero days; scaling up investments would strongly favor the attacker.

We now present a simple model that, while consistent with Rescorla's analysis for sufficiently low investment levels, ultimately predicts OD-scaling.

### 3.2.2 An idealized model

We consider a contest success function in which the output is the expected number of zero days that the attacker acquires over a given window of time. The inputs are quantities of effort devoted to vulnerability discovery over this time.

[26] Thomas M. Chen and Saeed Abu-Nimeh, 'Lessons from Stuxnet', *Computer* 44/4 (2011).
[27] Guido Schryen, 'Is Open Source Security a Myth?', *Communications of the ACM* 54/5 (2011).
[28] Eric Rescorla, 'Is Finding Security Holes a Good Idea?', *IEEE Security & Privacy* 3/1 (2005).
[29] Ablon and Bogart, *Zero Days, Thousands of Nights.*

$$S \equiv \text{Zero days acquired by attacker}$$
$$A \equiv \text{Attacker effort}$$
$$D \equiv \text{Defender effort}$$

For our idealized model, we make three simplifying assumptions. First, we assume that there is no correlation between the vulnerabilities the attacker discovers and the vulnerabilities the defender discovers. Second, we assume that no new vulnerabilities are introduced over the period of time under consideration. Third, we assume that, until all vulnerabilities are discovered, the number of vulnerabilities each actor discovers grows linearly with effort.

Now, let $N$ represent total number of vulnerabilities in the defenders' systems. In addition, let the units of effort be normalized so that one unit of effort corresponds to one expected discovery. Then, as shown in our appendix, it follows that the contest success function takes the form:

$$\mathbb{E}[S] = \min\{N, A\} \cdot \frac{N - \min\{N, D\}}{N}$$

Figure 4 presents a graphical depiction of the function. One can see that increasing investments tends to benefit the attacker when investment levels are low and tends to benefit the defender when they are sufficiently high. Figures 5 and 6 show this phenomenon more clearly, depicting the scaling of success-from-a-ratio, for an even investment ratio, and ratio-for-success, for a threshold of one zero day. For both metrics, we observe OD-scaling.

Why, intuitively, does the model predict OD-scaling? The underlying mechanism is essentially the same as in the ground invasion case. We again focus on success-from-a-ratio and consider what happens as $D$ increases.

Initially, due to randomness, the two actors are unlikely to discover precisely the same vulnerabilities. When their investments begin to increase from nothing, the attacker will therefore tend to find some vulnerabilities that the defender does not. However, given that the number of vulnerabilities is finite, the defender will eventually begin to achieve *saturation* by finding very large portions of them. This will place increasingly strict limits on how different the sets of vulnerabilities discovered by the two actors can be. Since defense is also *locally superior* – in the sense that any vulnerability the attacker and defender both discover does not become a zero day – the expected number of zero days must eventually begin to shrink too.

In our appendix, we consider each of the assumptions used in this simple model in turn. We again find that although they are highly idealized, it appears to be the case that reasonable adjustments to the assumptions, for instance to allow

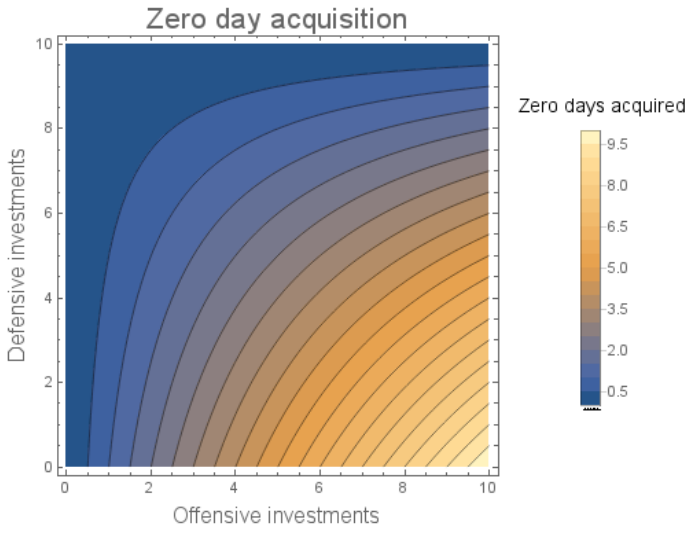


Figure 4: The CSF for our idealized model, for a case with 10 vulnerabilities.

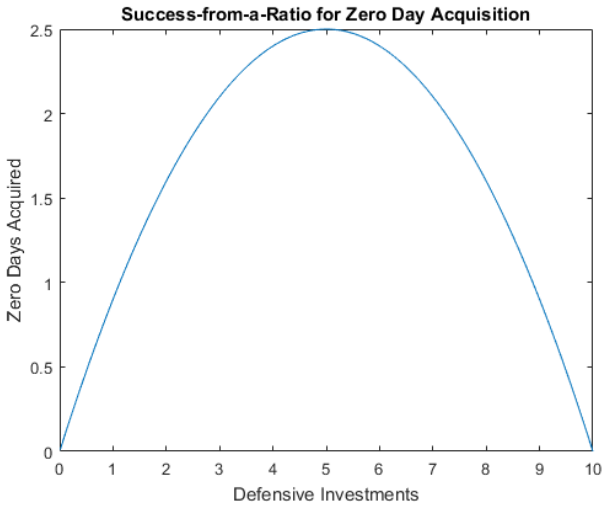


Figure 5: Success-from-a-ratio for our idealized model, for a case with 10 vulnerabilities and an even investment ratio.

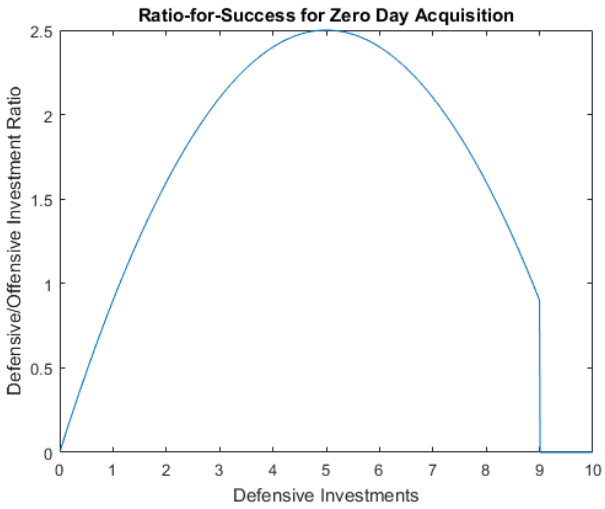


Figure 6: Ratio-for-success for our idealized model, for a case with 10 vulnerabilities and a threshold of one expected zero day.

for correlations between discoveries and diminishing returns on effort, do not challenge the basic mechanism described here.

### 3.2.3 Case study: Automated vulnerability discovery

The above analysis can also help us to understand the significance of ongoing work on automated vulnerability discovery.

Currently, manual testing and inspection still consistently allow actors to discover vulnerabilities that automated tools cannot.[30] However, there has recently been some noteworthy progress in adapting techniques from the field of artificial intelligence to the problem of vulnerability discovery.[31]

One obvious question is whether we should expect this progress, if it continues, to favor offense or defense. To answer this question, let us consider again our simplified model. If we take investments to be in units of vulnerability discovery capability, then we can interpret improvements in the efficiency of discovery tools

[30] Andrew Austin and Laurie Williams, 'One Technique is Not Enough: A Comparison of Vulnerability Discovery Techniques', *2011 International Symposium on Empirical Software Engineering and Measurement* (2011).

[31] Michael Kan, 'Mayhem Supercomputer Takes on Humans at Def Con', *Computer World*, 6 Aug. 2016; Teresa N. Brooks, 'Survey of Automated Vulnerability Detection and Exploit Generation Techniques in Cyber Reasoning Systems', *arXiv preprint arXiv:1702.06162* (2017); Konstantin Böttinger, Patrice Godefroid, and Rishabh Singh, 'Deep Reinforcement Fuzzing', *arXiv preprint arXiv:1801.04589* (2018).

as increases in the two actors' investments. Our model suggests that, all else being equal, these increases will tend to benefit attackers up until some point where they become large enough to benefit defenders.

However, all else might not be equal. One especially relevant possibility to note is that the vulnerabilities discovered by highly optimized tools might be more correlated than the vulnerabilities that are discovered in a more ad-hoc manner by humans today. Our robustness analysis, included in the appendix, shows that an increase in the correlation between discoveries would favor the defender.

In all, the short-term impact of increased automation is ambiguous. There appears to be at least one factor favoring the attacker and at least one other factor plausibly favoring the defender. A more thorough analysis would attempt to weigh these factors against each other, perhaps on the basis of data about discovery correlations between automated tools. The long-run impact will also depend on just how effective these tools can become. If they ever become efficient enough to allow defenders to discover significant majorities of the vulnerabilities in their systems, outpacing the rate at which new vulnerabilities are introduced, then the picture would seem to decisively favor defense.[32]

# 4 When should we expect OD-scaling?

## 4.1 A general model of conflict

We now present a general model that can be applied to many different forms of conflict, with the goal of explaining the conditions under which OD-scaling is likely to occur. For simplicity, we focus specifically on the success-from-a-ratio metric.

### 4.1.1 The model

In this model, there is some *attack surface* that can be thought to consist of discrete *attack vectors*. By making investments the attacker and defender can increase their *coverage* of these vectors, each up to some *maximum coverage level*. The manner in which the actors' investments produce patterns of coverage is determined by their respective *investment functions*. Each attack vector is furthermore associated with a *local contest success function*, which maps the two actors' coverage levels to offensive success at that point. Finally, there is

[32] For a recent comment on the plausibility of this long-run scenario, see Bruce Schneier, 'Artificial Intelligence and the Attack/Defense Balance', *IEEE Security & Privacy* 2 (2018).

an *aggregation function* that maps the values of all of the local contest success functions to the overall success of the attack.

### 4.1.2 The invasion and cyberattack models as special cases

The idealized models in sections 3.1.2 and 3.2.2 can be understood as special cases of this more general model.

For our ground invasion model, the attack surface is the set of points that make up the defensive line. Local coverage levels $A_i$ and $D_i$ represents how large a force each actor has concentrated at each point $i$. For all points, and both actors, the maximum coverage level is some constant $N$. The defender's investment function converts investments into equal levels of coverage across the points. The attacker's investment function converts investments into maximum coverage of as many points as possible. The local CSF $F_i(A_i, D_i)$ indicates how large of an attacking force breaks through at a given point. It is based on a linear attrition law, with $B$ being the local coverage ratio necessary to break through.

$$F_i(A_i, D_i) = \begin{cases} A_i - B * D_i, & A_i > B * D_i \\ 0, & otherwise \end{cases}$$

The aggregation function adds up the values of each of the local CSFs to find $S$, the total size of the attacking force that breaks through. This then implies the global contest success presented in Section 3.1.2.

For our vulnerability discovery model, the attack surface is the set of software vulnerabilities in the defender's system. Coverage is a binary variable that represents whether or not each actor has discovered the given vulnerability. For all points, and for both actors, the maximum coverage level is 1. The actors' investment functions convert investments into coverage of random and uncorrelated sets of vulnerabilities. The local CSF indicates whether or not a given vulnerability becomes a zero day. For a particular point $i$ and local coverage levels $A_i$ and $D_i$, it is given by:

$$F_i(A_i, D_i) = \begin{cases} 1, & A_i > D_i \\ 0, & otherwise \end{cases}$$

The aggregation function adds up the values of the local CSFs to find $S$, the total number of zero days the attacker acquires. This then implies the global contest success function presented in section 3.2.2.[33]

[33] One limitation of the general model is that it represents conflict as consisting of only a

### 4.1.3 A common mechanism for OD-scaling

Both of the special cases just described have a number of common properties. These properties are jointly sufficient (although not necessary) to produce OD-scaling.

- **Local defense superiority:** When the defender provides maximum coverage of any given attack vector, the output of the associated local CSF is approximately zero.
- **Defensive saturation:** Above some level of investment, the defender provides maximum coverage of all attack vectors.
- **Gap exploitation:** For some lower level of investment, the output of at least one local CSF is likely to be non-negligible. This is the result of the attacker's coverage of the associated attack vector substantially exceeding the defender's.
- **Aggregative success:** The output of the global CSF is approximately zero if the outputs of all local CSFs are approximately zero. It is also non-negligible if the output of at least one local CSF is non-negligible.

Together local defense superiority, defensive saturation, and aggregative success imply that the output of the global CSF will be approximately zero when investments are sufficiently high. Before saturation occurs, however, gap exploitation and aggregative success imply that the output of the global CSF will be non-negligible.

Success-from-a-ratio begins at zero. As the investment level grows, it rises to some non-negligible value. Ultimately, it must decrease toward zero again. OD-scaling occurs.

## 4.2 Additional special cases

Briefly, we now consider two additional cases: missile attacks and drone swarm attacks. We use the general model just described as a tool for assessing what scaling behavior these two cases are likely to exhibit. In particular, we ask whether the four properties sufficient for OD-scaling hold.

single round, in which all investments are applied. It is therefore unable to represent changes in coverage levels over time. For example, it cannot represent the arrival of reinforcements or multiple waves of attackers in the ground invasion case or represent multiple salvos in the missile defense case discussed below. We hypothesize that extending the relevant models to account for multiple rounds will not typically change the basic scaling phenomena predicted in each case, but we have not yet explored this hypothesis in depth.

Other cases that it may be worth considering, in future work, include cyberattacks based on social engineering (such as phishing attacks), blockades, aerial bombings, attacks with pathogens, and various forms of terrorism

### 4.2.1 Missile attacks

In this case, the attacker attempts to destroy a target by striking it with missiles. In turn, the defender attempts to protect the target by destroying the attacker's missiles. To make the case more concrete, we will take the target to be a warship and the defender to be relying exclusively on surface-to-air missiles.

We can conceptualize the attack vectors as non-overlapping trajectories by which the attacker's missiles can hit the ship. Furthermore, we can conceptualize the attacker and the defender as covering the same vector if a projectile launched by the defender follows a trajectory that leads it to intercept one of the attacker's missiles. Investments are measured in units of missiles.

The local CSFs indicate how many missiles associated with a given attack vector hit the ship, over a fixed period of time. They exhibit local defense superiority, since a missile that is intercepted will not reach its target. In addition, consistent with aggregative success, the global CSF indicates the total number of missiles that hit the ship.

In this case, gap exploitation occurs due to the inaccuracy of defensive systems. In particular, each surface-to-air missile has some non-zero chance of following a trajectory that does not lead it to intercept the intended offensive missile. This means that, as total investments scale up, the expected number of offensive missiles that fail to be intercepted will initially increase.

This somewhat abstract argument is consistent with models that have been developed in the missile defense literature. For instance, Washburn and Kress present a simple single-round model in which the defender distributes its missiles evenly among the offensive missiles, each defensive missiles has some probability $P$ of hitting the intended offensive missile (and no probability of hitting any other), and each offensive missile has some probability $Q$ of hitting the target if it is not destroyed first.[34] Let $A$ and $D$ be the number of missiles launched by the two actors. Then, making the mathematically convenient assumption that $\frac{D}{A}$ is an integer, we show in our appendix that the expected number of missiles that hit the target is given by:

$$E[S] = Q * A * (1 - P)^{\frac{D}{A}}$$

This expression implies offensive scaling. The associated function for success-

[34] Alan R. Washburn and Moshe Kress, *Combat Modeling* (New York: Springer 2009), 65-67.

from-a-ratio, $F_{R^*}(D) = Q * D * (1 - P)^{R^*}$, increases linearly with $D$.

The key remaining question, then, is whether the property of defensive saturation holds. This property would seem to imply that, for sufficiently high investment levels, the defender could create a virtual wall of missiles around the ship. In this case, the model just described would cease to be appropriate, as all trajectories that an offensive missile could take would lead to interception.

However, the plausibility of defensive saturation is questionable. Even if the defender possessed an essentially unlimited supply of surface-to-air missiles, it seems that the need to prevent these missiles from interfering with one another and the need to avoid fratricide would be key limiting factors.[35]

In short, in contrast with the primary two cases considered above, it seems likely that missile defense presents a case of purely offensive scaling with regard to the number of missiles used by both sides.

#### 4.2.2 Drone swarm attacks

In this case, a large number of drones launched by an attacker attempt to navigate close enough to a target to damage it with short-range munitions. The defender attempts to destroy the drones before they can get in range of the target. For instance, Allen and Chan consider a scenario in which an attacker directs millions of cheap kamikaze drones at an aircraft carrier battle group.[36]

This variety of attack is speculative. The only military drone swarms known to be in development, such as the US Department of Defense's Perdix systems, are not designed to carry munitions.[37] However, some analysts predict that, as the technology progresses, the possibility of using large swarms to destroy targets will become irresistible.

Individually, drones offer a number of advantages over human combatants or manned vehicles.[38] Since they do not need to carry or protect a human passenger, they can be especially long-ranged, fast, small, cheap to produce, and expendable. To the extent that they exhibit autonomy, they can also process information to make decisions much more quickly. Acting together, though,

[35] Note that our analysis here considers what would happen if the number of missiles used by both actors increased enormously. Once we also take into account practical limits on the number of missiles that could plausibly be stored and launched by an individual ship, it becomes more doubtful that defensive saturation could ultimately be observed in practice.

[36] Greg Allen and Taniel Chan, *Artificial Intelligence and National Security* (Cambridge, MA: Belfer Center for Science and International Affairs 2017).

[37] Paul Scharre, *Army of None* (New York: Norton 2018).

[38] Paul Scharre, *Robotics on the Battlefield: The Coming Swarm* (Washington DC: Center for a New American Security 2014).

their advantages can multiply. For orders-of-magnitude less than the cost of, for instance, a modern fighter jet or warship, a well-coordinated swarm of drones may be able to overwhelm the defenses of a target, dividing its attention and forcing it to expend its ammunition. Even if individual drones are not especially sophisticated or deadly, a large swarm may still have a very high probability of destroying its target.

Although a number of possible defenses against swarming attacks have been proposed, we focus on the case where the counter-swarm is the primary line of defense and ask what the impact of growing swarm sizes would likely be.[39]

At first glance, this case is structurally quite similar to missile defense, with individual defensive drones attempting to destroy individual offensive drones before they can reach the target. However, the potential ability of drones within a swarm to coordinate their movements closely with one another, hover within a fixed region of space, and defend against multiple attacking drones sequentially suggests that defensive saturation is more plausible than it is in the missile defense case.

Ultimately, it is plausible but far from certain that swarm defense will eventually present another case of OD-scaling with regard to the number of drones used by both sides.

# 5 Three effects of technological progress

It is well-understood that technological progress can impact offense-defense balances. In fact, perhaps the primary motivation for developing the concept has been to understand the distinctions between different eras of military technology.

For instance, European powers' failure to predict the grueling attrition warfare that would characterize much of the First World War is often attributed to their failure to recognize that new technologies, such as machine guns and barbed wire, had shifted the European offense-defense balance for conquest significantly toward defense.[40]

We see three primary ways in which technological progress can influence offense-defense balances.

Pathway 1 is to introduce a *new form* of conflict with distinct offense-defense dynamics. For instance, the emergence of cyberattacks as a form of conflict has

[39] Paul Scharre, 'Counter-Swarm: A Guide to Defeating Robotic Swarms', *War on the Rocks*, 31 March 2015.

[40] Stephen Van Evera, 'The Cult of the Offensive and the Origins of the First World War', *International Security* 9/1 (1984).

made it possible to discuss the "cyber offense-defense balance."

Pathway 2 is to change the *character* of an existing form of conflict in a manner that implies an updated contest success function. A CSF that maps relative force sizes into the expected success of a ground invasion, for example, will be different if they are equipped with mid-nineteenth century weaponry than if they are equipped with early twentieth century weaponry.[41]

Pathway 3 is to change the *quantity* of investments applied to the associated CSF. Technological progress can support an increase in investments – whether measured in units of financial cost (such as dollars) or in units of capability (such as Armored Division Equivalents) – by generating wealth through economic growth or by increasing the power and cost-effectiveness of technologies used in conflict. We have seen, in turn, that rising investments can significantly influence the offense-defense balance.

Pathway 1 and Pathway 2 are widely recognized. The third pathway, however, appears to have received rather less attention.[42] Arguably, this relative neglect has been understandable. The concept of an offense-defense balance has been discussed primarily in the context of ground invasions, often with a focus on head counts as inputs. Technological progress does not support the "scaling up" of ground force levels nearly as directly as it supports the scaling up of, for example, the number of software vulnerabilities an actor discovers. In addition, the effects of changing investment levels may also be obscured by the fact that technological progress is continually changing the associated conflict domains and CSFs as well.

However, we argue that this third pathway will become increasingly difficult to ignore. Brundage et al. argue that *scalability* is one of the key distinguishing features of digital and artificially intelligent systems.[43] The marginal cost of increasing the effectiveness of a given piece of software or increasing the number of copies in operation will often be quite low; the marginal cost can also often be expected to decline exponentially with the cost of computing power over

[41] Concretely, holding force sizes fixed, the conventional wisdom holds that a conflict with mid-nineteenth century technology could be expected to produce a better outcome for the attacker than a conflict with early twentieth century technology. See, for instance, Van Evera, 'Offense, Defense, and the Causes of War'.

[42] Glaser and Kaufman briefly discuss the significance of force-to-space ratios in their classic article on the offense-defense balance. Elsewhere, without directly linking their result to the offense-defense balance literature, Johnson and MacKay suggest that defense may be more difficult in the context of conflicts between small-scale societies (such as hunter-gatherer groups) than in the context of conflicts between large-scale societies (such as countries). Lindsay also suggests that the relative costs of carrying out anonymous cyberattacks and attributing them may scale in a way that ultimately favors defenders. Dominic D.P. Johnson and Niall J. MacKay, 'Fight the Power: Lanchester's Laws of Combat in Human Evolution', *Evolution and Human Behavior* 36/2 (2015); Jon R. Lindsay, 'Tipping the Scales: The Attribution Problem and the Feasibility of Deterrence against Cyberattack', *Journal of Cybersecurity* 1/1 (2015).

[43] Miles Brundage, et al., 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation', *arXiv preprint arXiv:1802.07228* (2018).

time. The manufacture and operation of certain robots may also lend itself to unusually rapid scaling. For instance, anticipating a continual decline in manufacturing costs, Scharre suggests that it may be possible to deploy swarms of "billions" of drones in the future.[44] Overall, a trend toward more highly automated and digitally-oriented conflict suggests that scaling phenomena may become much more salient.

In Sections 3.2.3 and 4.2.2, we discussed two concrete instances of this trend. First, we argued that progress in developing and applying automated vulnerability tools may dramatically increase actors' abilities to discover software vulnerabilities. All else being equal, a plausible effect of this trend would be to tilt the relevant offense-defense balance further toward offense in the short-run, then toward defense in the long run.[45]

Second, as Scharre argues, progress in developing drone swarms for use in combat may dramatically increase the number of weapons platforms available to military actors. This raises the tactical possibility of using large swarms of individually expendable drones to overwhelm the defenses of targets, with the *quantity* of drones playing perhaps an even greater role than their *quality* does in determining their chance of success. Our analysis suggests that, for target defense, the net effect of a trend of growing swarm sizes is likely to initially benefit attackers. Whether further increases eventually benefit defenders will depend on whether it is feasible for defenders to essentially saturate the airspace around a target with defensive drones.[46]

# 6 Conclusion

In this paper we have shown how investment levels can play an important role in determining the offense-defense balance. For both ground invasions and attempts to acquire zero days, our models predict that scaling up investments will initially benefit offense and ultimately benefit defense. We argued that this pattern, which we call OD-scaling, is a useful baseline assumption across a range of cases. A foothold for understanding the case of a missile attack, for example, is asking whether or not the case displays certain features that are sufficient for OD-scaling.

Finally, we have suggested that scaling effects can be expected to play a par-

[44] Paul Scharre, *Robotics on the Battlefield: The Coming Swarm.*

[45] However, as noted in the earlier section, automation may also change the CSF in a manner favorable to the defender, by increasing the correlation between the vulnerabilities the actors discover. This effect could conceivably offset the initial benefit attackers derive from scaling effects.

[46] Of course, it is not yet known how significant a role counter-swarms will play in defense against offensive drone swarms. If gun systems or area denial weapons play larger roles, for instance, then this result has less direct significance.

ticularly large role in determining the offense-defense balance associated with drone swarm warfare and automated vulnerability discovery. Scaling effects, therefore, are likely to become increasingly difficult to ignore and essential to understand.

We have shied away from drawing specific conclusions about the relationship between scaling effects and international security. The apparent implications will vary with one's views on unresolved controversies in offense-defense theory, particular the balance's relationship with crisis and peacetime stability. While some hold that shifts toward offense-dominance obviously favor conflict and arms racing, this position has been challenged on a number of grounds. It has even been suggested that shifts toward offense-dominance can increase stability in a number of cases.

Providing a full account of these controversies is beyond the scope of this paper. Nevertheless, further work to understand the security implications of scaling effects will require direct engagement with offense-defense theory. Given that offense-defense theory has also been developed primarily under the assumption that the offense-defense balance is a constant property of a system and technological era, rather than a variable that changes with the actors' investments, our work may also suggest the need for substantial theoretical revision.

# 7 Bibliography


Ablon, Lillian and Andy Bogart, *Zero Days, Thousands of Nights: The Life and Times of Zero-Day Vulnerabilities and Their Exploits* (Santa Monica, CA: RAND Corporation 2017).

Allen, Greg and Taniel Chan, *Artificial Intelligence and National Security* (Cambridge, MA: Belfer Center for Science and International Affairs 2017).

Austin, Andrew and Laurie Williams, 'One Technique is Not Enough: A Comparison of Vulnerability Discovery Techniques', *2011 International Symposium on Empirical Software Engineering and Measurement* (2011).

Biddle, Stephen D., et al., *Defense at Low Force Levels: The Effect of Force to Space Ratios on Conventional Combat Dynamics* (Alexandria, VA: Institute for Defense Analyses 1991).

Böttinger, Konstantin, Patrice Godefroid, and Rishabh Singh, 'Deep Reinforcement Fuzzing', *arXiv preprint arXiv:1801.04589* (2018).

Brooks, Teresa N., 'Survey of Automated Vulnerability Detection and Exploit Generation Techniques in Cyber Reasoning Systems', *arXiv preprint arXiv:1702.06162*

(2017).

Brundage, Miles, et al., 'The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation', *arXiv preprint arXiv:1802.07228* (2018).

Chen, Thomas M. and Saeed Abu-Nimeh, 'Lessons from Stuxnet', *Computer* 44/4 (2011).

Fearon, James D., 'The Offense-Defense Balance and War Since 1648', draft paper (1997).

Fearon, James D., 'Cooperation, Conflict, and the Costs of Anarchy', *International Organization* 72/3 (2018).

Gartzke, Erik, 'The Myth of Cyberwar: Bringing War in Cyberspace Back Down to Earth', *International Security* 38/2 (2013).

Gartzke, Erik and Jon R. Lindsay, 'Weaving Tangled Webs: Offense, Defense, and Deception in Cyberspace', *Security Studies* 24/2 (2015).

Gartzke, Erik, 'Blood and Robots: How Remotely Piloted Vehicles and Related Technologies Affect the Politics of Violence'.

Glaser, Charles L. and Chaim Kaufmann, 'What is the Offense-Defense Balance and How Can We Measure It?', *International security* 22/4 (1998).

Hart, Basil L., 'The Ratio of Troops to Space', *Royal United Services Institution Journal* 105/618 (1960).

Jackson, Robert, *The Fall of France: May-June 1940* (Oxford: Oxford University Press 2003).

Jervis, Robert, 'Cooperation under the Security Dilemma', *World Politics* 30/2 (1978).

Johnson, Dominic D.P. and Niall J. MacKay, 'Fight the Power: Lanchester's Laws of Combat in Human Evolution', *Evolution and Human Behavior* 36/2 (2015).

Kan, Michael, 'Mayhem Supercomputer Takes on Humans at Def Con', *Computer World*, 6 Aug. 2016.

Kello, Lucas, 'The Meaning of the Cyber Revolution: Perils to Theory and State-Craft', *International Security* 38/2 (2013).

Lieber, Keir, 'The Offense-Defense Balance and Cyber Warfare', *Cyber Analogies* (2014).

Lindsay, Jon R., 'Tipping the Scales: The Attribution Problem and the Feasibility of Deterrence against Cyberattack', *Journal of Cybersecurity* 1/1 (2015).

Lynn-Jones, Sean M., 'Offense-Defense Theory and its Critics', *Security Studies* 4/4 (1995).

Malone, Patrick. J., 'Offense-Defense Balance in Cyberspace: A Proposed Model', Master's thesis, Naval Postgraduate School (2012).

Mearsheimer, John J., 'Assessing the Conventional Balance: The 3:1 Rule and its Critics', *International Security* 13/4 (1989).

Mearsheimer, John J., 'Numbers, Strategy, and the European Balance', *International Security* 12/4 (1988).

Rescorla, Eric, 'Is Finding Security Holes a Good Idea?', *IEEE Security & Privacy* 3/1 (2005).

Roberson, Brian, 'The Colonel Blotto Game', *Economic Theory* 29/1 (2006).

Saltzman, Ilai, 'Cyber Posturing and the Offense-Defense Balance', *Contemporary Security Policy* 34/1 (2013).

Scharre, Paul, *Army of None* (New York: Norton 2018).

Scharre, Paul, 'Counter-Swarm: A Guide to Defeating Robotic Swarms', *War on the Rocks*, 31 March 2015.

Scharre, Paul, *Robotics on the Battlefield: The Coming Swarm* (Washington DC: Center for a New American Security 2014).

Schneider, Jacquelyn, 'The Capability/Vulnerability Paradox and Military Revolutions: Implications for Computing, Cyber, and the Onset of War'.

Schneier, Bruce, 'Artificial Intelligence and the Attack/Defense Balance', *IEEE Security & Privacy* 2 (2018).

Schryen, Guido, 'Is Open Source Security a Myth?', *Communications of the ACM* 54/5 (2011).

Sechser, Todd, Neil Narang, and Caitlin Talmadge, 'Emerging Technologies and Strategic Stability in Peacetime, Crisis, and War'

Michelle R. Garfinkel and Stergios Skaperdas, 'Economics of Conflict: An Overview', *Handbook of Defense Economics* 2 (2007).

Slayton, Rebecca, 'What is the Cyber Offense-Defense Balance? Conceptions, Causes, and Assessment', *International Security* 41/3 (2017).

Talmadge, Caitlin, 'Emerging Technologies and Intra-War Escalation Risks: Evidence from the Cold War, Implications for Today'.

Van Evera, Stephen, 'The Cult of the Offensive and the Origins of the First World War', *International Security* 9/1 (1984).

Van Evera, Stephen, 'Offense, Defense, and the Causes of War', *International Security* 22/4 (1998).

Volpe, Tristan, 'Dual-Use Distinguishability: How 3D-Printing Shapes the Security Dilemma for Nuclear Programs'.

Washburn, Alan R., and Moshe Kress, *Combat Modeling* (New York: Springer 2009).

Wieviorka, Olivier, *Normandy: The Landings to the Liberation of Paris* (Cambridge, MA: Harvard University Press 2008).

Williams, Heather, 'Asymmetric Arms Control and Strategic Stability: Scenarios for Limiting Hypersonic Glide Vehicles'.