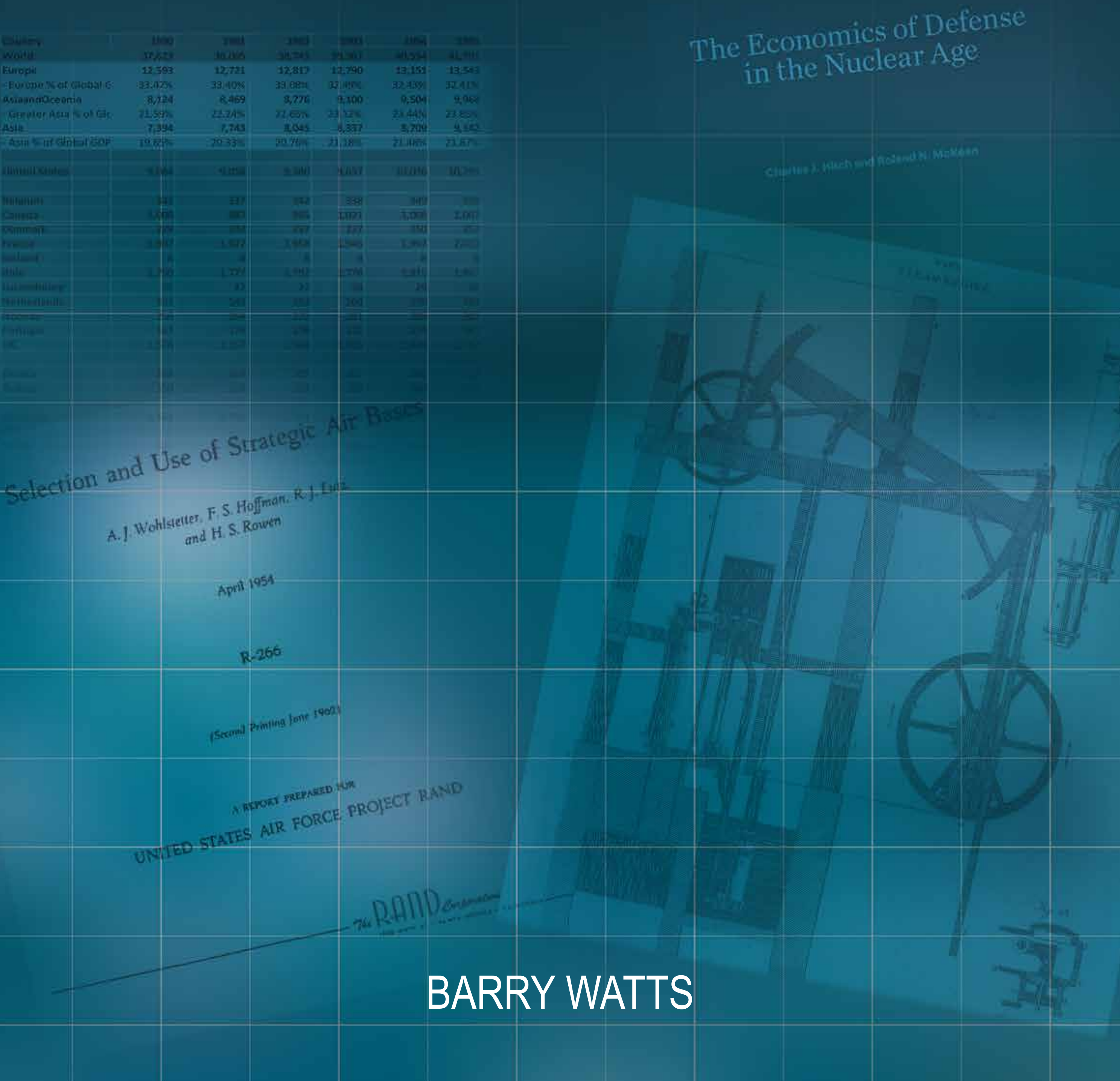
CSBA

Center for Strategic and Budgetary Assessments

ANALYTIC CRITERIA FOR JUDGMENTS, DECISIONS, AND ASSESSMENTS



BARRY WATTS

ANALYTIC CRITERIA FOR JUDGMENTS, DECISIONS, AND ASSESSMENTS

BARRY WATTS

CSBA

Center for Strategic and Budgetary Assessments

2017

ABOUT THE CENTER FOR STRATEGIC AND BUDGETARY ASSESSMENTS (CSBA)

The Center for Strategic and Budgetary Assessments is an independent, nonpartisan policy research institute established to promote innovative thinking and debate about national security strategy and investment options. CSBA's analysis focuses on key questions related to existing and emerging threats to U.S. national security, and its goal is to enable policymakers to make informed decisions on matters of strategy, security policy, and resource allocation.

ABOUT THE AUTHOR

Barry Watts was a senior fellow at the Center for Strategic & Budgetary Assessments (CSBA) from 2002 to 2014. There he wrote on such topics as the second nuclear age, strategic competence, realistic combat training, precision strike, the defense industrial base, diagnostic net assessment, and the military use of space. From 2001–2002 Mr. Watts directed the Office of Program Analysis and Evaluation in the Office of the Secretary of Defense. From 1986 to 2001 he was a member the Northrop Grumman Analysis Center, serving as its director from 1997 to 2001. During his years with Northrop, he published an analysis of Clausewitzian friction and future war and led the Gulf War Air Power Survey's work on the effects and effectiveness of air power in the 1991 Persian Gulf War. From 1965 to 1986 Mr. Watts was in the U.S. Air Force. During his Air Force career, he flew a combat tour in Vietnam in F-4s, taught logic and philosophy at the U.S. Air Force Academy, served two tours in the Office of Net Assessment under Andrew Marshall, and headed the Red Team in the Air Staff's Project Checkmate. His most recent publication, co-authored with Andrew Krepinevich, is *The Last Warrior: Andrew Marshall and the Shaping of Modern American Defense Strategy* (Basic Books, 2015). Mr. Watts is currently retired and living in North Carolina with his wife, Hope.

ACKNOWLEDGMENTS

The author is grateful to Thomas Mahnken, president of CSBA, as well as to Andrew Marshall and Andrew May from the Office of Net Assessment. Thanks go to the CSBA staff for editing and production support. The author is especially grateful for Kamilla Gunzinger's persistence in staying with this project for over more than a year. The opinions and analysis in this study are those of the author; any shortcomings are solely his responsibility. CSBA receives funding from a broad and diverse group of funders, including private foundations, government agencies, and corporations. A complete list of these organizations can be found on our website at www.csbaonline.org/about/contributors.

Cover: Report covers, GDP spreadsheets, and a steam engine schematic. U.S. Government and Public Domain sources.

Contents

FIGURES

TABLES

Foreword

The individual who provided the impetus for this investigation of analytic criteria for judgments, decisions, and assessments was Andrew W. Marshall, the Pentagon's Director of Net Assessment from 1973 to 2015. In the early 1990s, a number of us who had served on Marshall's small staff began encouraging him to document the intellectual history of the net assessment enterprise during the last two decades of the U.S.–Soviet Cold War. One of the subjects that came up time and again in the ensuing discussions of diagnostic net assessment was what a small group of RAND economists had termed "the criterion problem." This issue emerged from RAND's early efforts to employ systems analysis to advise the U.S. Air Force on the preferred instrumentalities and techniques of intercontinental air warfare in the nuclear age. Economists such as Marshall and Charles J. Hitch, who headed RAND's economics division from 1948 to 1961, realized that there were logical and other problems with the decision criteria used in systems analyses such as the massive, multi-year *Strategic Bombing Systems Study* completed in 1950 under Edwin W. Paxson.

It was not until I undertook the present study in 2014 that I began to realize just how ubiquitous and fundamental the selection of appropriate criterion problem was to judgments about historical outcomes, choices among strategic alternatives, and the conduct of useful net assessments. A related insight emerged from Hitch and James R. Schlesinger's comments on the criterion problem. Because there is no by-the-numbers formula or method for selecting appropriate analytic criteria, such choices inevitably contain a subjective element. Among other things, this insight reveals that judgments about historical outcomes are always open to revision. Similar certainties apply to the analytic criteria underlying strategic judgments and net assessments.

Barry Watts

October 2016

CHAPTER 1

Conjectures and Evidence

What of a constructive nature can be said about the selection of criteria? Clearly, there is no all-purpose criterion, for the appropriate test depends upon what alternatives are open to the decision-maker, upon what aspects of the situation must be taken as given, and even upon what measurements are feasible.

Charles Hitch and Roland McLean, 1960[1]

[F]or most higher-order problems adequate measures of merit have yet to be devised. Even if some improvement may be hoped for, the irreducibly subjective element in such measures will remain substantial.

James Schlesinger, 1967[2]

The Analytic Criteria Problem Stated

Choosing analytic criteria for making strategic choices or judging historical outcomes is a recurring, if not universal, problem. It recurs because no general method for choosing *appropriate* criteria is known to exist despite the early hopes for methodological innovations ranging from macroeconomic models to operations research, systems analysis, and game theory. In the absence of a general solution to the problem of selecting appropriate criteria, every analysis of potential strategies, force structure alternatives, organizational options, or historical outcomes involves contextual specifics that make each case unique. Among other things, the appropriateness of the criteria chosen in any instance depends on many things: the goals sought; the resources required and their opportunity costs; the completeness, accuracy, and clarity of the accessible data and information; the political and bureaucratic context

1 Charles J. Hitch and Roland N. McLean, *The Economics of Defense in the Nuclear Age* (Cambridge, MA: Harvard University Press, 1960), pp. 174–175, also published as Charles J. Hitch and Roland N. McLean, *The Economics of Defense in the Nuclear Age*, R-346 (Santa Monica, CA: The RAND Corporation, 1960).

2 James R. Schlesinger, *Systems Analysis and the Political Process*, P-3464 (Santa Monica, CA: The RAND Corporation, June 1967), p. 4.

surrounding the analysis; the uncertainties and incommensurables inherent in the situation; and, in the case of strategic choices, the time available for choosing appropriate criteria, conducting the analysis, and reaching judgments based on the chosen criteria, whether explicit or implicit.

Initial Evidence for the Problem's Insolubility

The supposition that the general problem of choosing analytic criteria has no universal solution should be initially approached as, at best, a conjecture—a hypothesis presumed to be true but has not yet been proved. Indeed, the claim that there is no universal solution to choosing appropriate criteria may not be provable in anything approaching a mathematical sense. Nevertheless, the hypothesis can be supported by examining a range of specific cases, which is the approach taken in this report.

The following questions highlight three long-studied cases from the 1940s in which the choice of metrics clearly influenced the answers commentators and scholars have reached about their outcomes and efficacy.

- What criteria drove the ebb and flow of the protracted struggle between German U-boats and Allied convoys in the Battle of the Atlantic during 1939–1945?
- How effective was the Anglo–American Combined Bomber Offensive (CBO) in achieving the progressive dislocation and destruction of the German economy as well as undermining the will of the German people to continue fighting?
- In the late 1940s and early 1950s, what were the U.S. Air Force's most cost-effective choices for providing active air defenses and modernizing Strategic Air Command's nuclear bomber force?

The first case explores causal criteria that explain the outcome of what Winston Churchill proclaimed in 1940 to be the "Battle of the Atlantic." The second assesses the effectiveness of Anglo–American strategic bombing doctrine in Europe during World War II with emphasis on how the choice of criteria for judging the campaign continues to influence conclusions about the CBO's effectiveness. The third deals with an early postwar application of systems analysis by Project RAND to advise the U.S. Air Force on the preferred instrumentalities and techniques for conducting intercontinental air warfare. Together these three cases are intended to provide initial steps toward building a strong empirical case for the claim that appropriate criteria vary widely from one instance to the next and are determined by the specifics of the situation being analyzed. Later chapters will explore other cases in an effort to place the insolubility of the analytic criteria problem beyond a reasonable doubt. These later cases include the causes of the West's rise to global economic and military dominance after 1500, the limits of widely used leading economic indicators such as gross domestic product (GDP), and the central role of criteria in diagnostic net assessments.

The selection of cases in what follows is not random. This fact raises the possibility of methodological bias: selecting only those cases that support the general insolubility of choosing appropriate decision criteria while ignoring those that do not. This problem is called sampling on the dependent variable or selection bias and emerges from the difficulties of identifying "causal mechanisms in a probabilistic world."[3] But insofar as strategic choices and judgments *always* depend on underlying decision criteria, all that need be shown is that at least some cases involve very different criteria. That some—but demonstrably not all—cases involve more or less the same decision criteria does not refute the absence of a universal solution to the problem of choosing appropriate ones. In the instances reprised in this chapter, it may be tempting in terms of aggregate outputs to suppose that costs and benefits are common to all three. But at the level of causal inputs, as we shall see, very different factors drove outcomes as well as judgments about efficacy in each of them. As Schlesinger observed in 1967: "For higher-order problems the measures of effectiveness must be based on some broad strategic criteria, in which political or psychological assessments unavoidably play a role, whether this be admitted or not."[4]

The U-boat Campaign in the North Atlantic

The protracted German effort to knock Britain out of the war by strangling Allied supply lines across the Atlantic was the longest continuous campaign of World War II. Clay Blair divides this contest between German submarines (*Unterseeboote*) and Allied merchant shipping into three distinct periods: September 1939 to December 1941, when the U-boats focused almost solely on the merchant shipping of the British Empire; December 1941 to August 1942, during which the U-boats mostly operated independently against shipping in American waters; and September 1942 to May 1945, when the U-boats returned to the North and Mid-Atlantic convoy routes between North America and the British Isles.[5] Most accounts of the U-boat war identify September 1942 to May 1943 as the decisive phase in which the Allied advances in anti-submarine warfare (ASW) capabilities and code-breaking defeated the German submarine threat in the Atlantic.

In all three periods, the ebb and flow of the struggle between Allied shipping and the U-boats have generally been portrayed using aggregate quantitative metrics: predominately the

3 Mie Augier, James G. March, and Andrew W. Marshall, "The Flaring of Intellectual Outliers: An Organizational Interpretation of the Generation of Novelty in the RAND Corporation," *Organization Science*, April 17, 2015, p. 4. For a recent examination of selection bias, see Bernard Forgues, "Sampling on the Dependent Variable Is Not Always Bad: Quantitative Case-Control Designs for Strategic Organization Research," *Strategic Organization* 10, no. 3, 2012, pp. 269–275.

4 James R. Schlesinger, *On Relating Non-Technical Elements to Systems Studies*, P-3545 (Santa Monica, CA: The RAND Corporation, February 1967), p. 4.

5 Clair Blair, *Hitler's U-boat War: The Hunters 1939–1942* (New York: Random House, 1996), pp. 418–427, 691–700. For a more fine-grained division of the U-boat war into seven periods, see Charles M. Sternhell and Alan M. Thorndike, *Antisubmarine Warfare in World War II*, Operations Analysis Group (OEG) Report No. 51 (Washington, DC: Navy Department, 1946), p. 11.

tonnages of Allied shipping sunk and U-boats lost from one month to the next or over the course of the campaign. From the Allies' perspective, the most critical metric consisted of monthly shipping losses. As P. M. S. Blackett wrote after the war, by November 1942, the Allied perception was that, unless rapidly checked, the German U-boats "might make the invasion of Europe in 1943 or even 1944 impossible through lack of shipping" from North America.[6] In that month losses to the U-boats in all theaters peaked at nearly 730,000 Gross Registered Tons (GRTs) and total losses—including those to German warships, armed merchants, aircraft, and mines—came to over 800,000 GRTs.[7] In exchange, the cost to the *Kriegsmarine* in November 1942 was a mere 13 U-boats, of which only five were lost attacking convoys in the North Atlantic.[8]

These aggregate attrition measures not only fail to tell the whole story, but they also offer little insight into why the Allies eventually prevailed. Given the sheer magnitude of the Allies' shipping losses (Table 1), many observers—both then and later—have agreed with Blackett's judgment that in late 1942 or early 1943 the Germans came close to cutting the vital lifeline between North America and the British Isles.[9] But neither facts nor recent analyses support this judgment despite the accelerating demands for men and materiel of Allied forces in both Europe and the Pacific during 1943 and 1944.

Table 1 shows the numbers and tonnage of Allied ocean-going merchant vessels (of 1,600 gross tons or more) lost to enemy action as well as non-combat "marine" casualties in all theaters from September 1939 through December 1944.[10] Some 25 percent of the merchants sunk, and 12 percent of Allied merchant tonnage lost, was due to marine casualties rather than enemy action. Over the course of the war, the Germans attacked Allied shipping with warships, armed merchants, aircraft, and mines in addition to U-boats. The U-boats, however, posed the gravest threat to Allied sea lines of communications across the Atlantic. As Table 1 shows, 1942 was the year of the heaviest Allied shipping losses—due in no small part

6 P. M. S. Blackett, *Studies of War: Nuclear and Conventional* (New York: Hill and Wang, 1962), p. 227; and Ralph Bennett, *Behind the Battle: Intelligence in the War with Germany, 1939–45* (London: Sinclair–Stevenson Ltd, 1994), p. 185.

7 Clair Blair, *Hitler's U-boat War: The Hunted 1942–1945* (New York: Random House, 1998), pp. 120, 736–737; and "Losses during the Battle of the Atlantic," *Wikipedia*, last modified March 15, 2017, available at http://en.wikipedia.org/wiki/Losses_during_the_Battle_of_the_Atlantic. Gross Registered Tons are a measure of the permanently enclosed internal volume of a ship; 100 cubic feet equals 1 GRT.

8 Axel Niestlé, *German U-Boat Losses During World War II: Details of Destruction* (Barnsley, England: Frontline Books, 2014), p. 199.

9 Blair argued in 1998 that, contrary to accepted wisdom, the "U-boats never even came close at any time to cutting the vital North Atlantic lifeline to the British Isles." Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 707.

10 Martin K. Metcalf, *History of Convoy and Routing 1939–1945*, Declassified FX-37 (Washington, DC: Navy Department, May 1945), pp. 1–4, Appendix I, available at https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/h/history-convoy-routing-1945.html. Table 1 ignores seven ships (including escorts) sunk in Atlantic convoys and three more lost in East Coast convoys during January–May 1945, the last five months of the European war.

to Admiral Ernest King's refusal during the first half of the year to institute a convoy system in the Caribbean and along the Atlantic coast of the United States.[11]

TABLE 1: ALLIED MERCHANT SHIPPING LOSSES 1939–1944[12]

	Losses to Enemy Action		Marine Casualties		Totals	
	Merchant Vessels	Gross Tons	Merchant Vessels	Gross Tons	Merchant Vessels	Gross Tons
Sep-Dec 1939	216	746,712	107	188,716	323	935,428
1940	982	3,877,394	363	672,286	1,345	4,549,680
1941	1,114	4,141,915	305	551,510	1,419	4,693,425
1942	1,562	7,713,119	290	597,936	1,852	8,311,055
1943	588	3,209,915	257	508,390	845	3,718,305
1944	193	1,036,904	233	400,689	426	1,437,593
	4,655	**20,725,959**	**1,555**	**2,919,527**	**6,210**	**23,645,486**

The eventual loss of more than 6,000 merchants comprising over 23,000,000 gross tons is a staggering total, especially when one considers the lives lost. But these numbers must be put in context. When the United States entered World War II in December 1941, the British Commonwealth and the United States controlled "about 41,000,000 gross tons of ocean-going merchant vessels," of which some 8,500,000 tons were American.[13] By December 1944, however, the Allies' ocean-going merchant capacity had grown to about 57,000,000 gross tons despite the loss of 2,343 merchant vessels to enemy action totaling 11,959,938 gross tons during the years 1942–1944 (in addition to the 2,312 merchants totaling 8,766,021 tons lost to enemy action during the years 1939–1941).[14] Both before and after the Japanese attack on Pearl Harbor the Allies were eventually able to more than offset shipping losses to enemy action with new construction and the acquisition of foreign shipping. "In 1942, American, British, and Canadian shipyards produced about 7.1 million gross tons, or about a million more gross

11 Phillips Payson O'Brien, *How the War Was Won: Air-Sea Power and Allied Victory in World War II* (Cambridge, UK: Cambridge University Press, 2015), pp. 234–249.

12 Metcalf, *History of Convoy and Routing*, p. 3. Gross tonnage is a unit-less index based on a ship's overall internal volume.

13 Ibid., p. 2.

14 Ibid., p. 3. Blair estimates that as of December 1941, the U.S. merchant fleet consisted of about 1,400 or 1,500 ships totaling 8.5 million gross tons, including about 390 tankers of about 2.8 million gross tons. He puts British controlled merchant capacity in December 1941 at some 20 million gross tons. Blair, *Hitler's U-boat War: The Hunters 1939–1942*, pp. 98, 446. FX-37, however, puts British Commonwealth merchant capacity at the time of Pearl Harbor figure to have been 32.5 million gross tons. *History of Convoy and Routing*, p. 2.

tons than were lost to the U-boats (6.1 million)" that year.[15] From May through September 1942 American shipyards alone built 2.6 million GRTs, enough to nearly offset the 2.8 million GRTs sunk by U-boats worldwide. And in the critical opening months of 1943, the prodigious output of U.S. shipyards was more than double the amount of Allied shipping losses (Table 2).[16]

TABLE 2: ALLIED SHIP LOSSES VERSUS U.S. CONSTRUCTION[17]

Month	Allied ships lost (all theaters)	Gross tons	Ships built in U.S. yards	Gross tons from U.S. yards
January-43	37	203,128	106	647,000
February-43	63	359,328	132	792,000
March-43	108	627,337	149	1,005,000
April-43	56	327,943	159	1,076,000
Totals	**264**	**1,517,736**	**546**	**3,520,000**

No comparable construction effort emerged on the German side to produce a fleet of ocean-going U-boats large enough to sever Allied sea lines of communications across the Atlantic. In March 1941 Admiral Karl Dönitz, the commander of U-boats (*Befehlshaber der Unterseeboote* or *B.d.U.*), "still controlled only 27 combat ready oceangoing boats," the number with which he had begun the war 19 months earlier.[18] Not until the fall of 1940 did Adolph Hitler accord "Special Stage" priority to U-boat and torpedo construction comparable to that previously accorded to the *Wehrmacht*, *Luftwaffe,* and the high seas fleet.[19] Hence, contrary to Allied wartime perceptions and many postwar assessments, Clair Blair concluded in the late 1990s that Dönitz never had anywhere near the number of U-boats he needed to win the Battle of the Atlantic.[20] The British historian Andrew Roberts agrees: however vicious

15 Blair, *Hitler's U-boat War: The Hunters 1939–1942*, p. 697.

16 Ibid., p. 697. In 1942 and 1943 American yards produced 2,709 merchant ships. Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 9.

17 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 168.

18 Blair, *Hitler's U-boat War: The Hunters 1939–1942*, p. 262.

19 Ibid., pp. 190–192, 218–220.

20 For a succinct summary of Blair's assessment of the inadequacy of the German U-boat campaign through December 1941, see Blair, *Hitler's U-boat War: The Hunters 1939–1942*, pp. 418–427. For later summaries of his position, see Blair, *Hitler's U-boat War: The Hunted 1942–1945*, pp. 164–169, 706–711. Before the war, Dönitz calculated that he needed 300 U-boats to win the Battle of the Atlantic. Blair's view is that Dönitz probably needed twice that number. The peak number of U-boats the *Kriegsmarine* possessed in 1943 was less than 440 oceangoing submarines. But less than 170 of these were on "active service" on all battlefronts. After the war, Dönitz commented that Germany should have started the war with 1,000 U-boats. Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 276; Sternhell and Thorndike, *Antisubmarine Warfare in World War II*, p. 83; Blair, *Hitler's U-boat War: The Hunters 1939–1942*, p. 240; Bennett, *Behind the Battle*, p. 197; and Admiral Karl Dönitz, *The Conduct of the War at Sea* (Washington, DC: Navy Department, January 15, 1946), p. 23.

and bitter the Battle of the Atlantic was, “Britain’s survival was never really in doubt, even though for most people on both sides it certainly did not look that way at the time.”[21]

Germany’s failure to field a large enough U-boat fleet points to the other important attrition metric widely used to assess the Allies campaign against German U-boats: the numbers of *Kriegsmarine* submarines lost over the course of the war. Through December 1944 German submarines sank almost 2,700 Allied merchants totaling over 14 million gross tons—69 percent of the total of more than 23 million gross tons lost to enemy action during this period.[22] Like inflated German wartime estimates of Allied merchant tonnage sunk, Allied wartime estimates of U-boats sunk were subject to the usual fog and frictions of war. A 1946 Navy Department study estimated that the Germans lost 705 U-boats.[23] Postwar research based on U.S. Navy, British Admiralty, and *B.d.U.* records have produced more accurate totals for the number of U-boats lost during the war. In what appears to be the most recent and authoritative estimate, Axel Niestlé concluded in 2014 that a total of 757 U-boats were lost to Allied action or other causes, both at sea and in port; of these 648 were lost at sea during frontline service.[24] He adds that the loss rate among frontline U-boat crews for the entire war exceeded 60 percent.

The total number of U-boats lost during World War II, like the aggregate tonnages of Allied merchant vessels sunk, is impressive. But these aggregate attrition figures reveal very little about how and why the Allies defeated the U-boat threat in the spring of 1943. The decisive period in the Battle of the Atlantic began in September 1942 when Dönitz decided to redeploy most of the operational U-boats to what he judged to be the decisive front in the Atlantic battle: the Atlantic convoy routes between North America and the British Isles.[25] The battle there between Allied convoys and the U-boats culminated in May 1943 when unsustainable losses forced Dönitz to withdraw his submarines from the North Atlantic.[26] True, the U-boats returned to the North Atlantic later in the year, and Allied losses to German submarines continued through May 1945. But after May 1943 the U-boats no longer posed a serious threat to Britain’s transatlantic lifeline.

21 Andrew Roberts, *The Storm of War: A New History of the Second World War* (London: Allen Lane, 2009), p. 374. More recently O’Brien reached the same conclusion as Roberts about the Battle of the Atlantic. Phillips Payson O’Brien, *How the War Was Won*, pp. 242, 262.

22 Metcalf, *History of Convoy and Routing*, p. 3. Clay Blair puts Allied losses to the U-boats at some 3,000 ships of all types totaling 15 million gross tons. Blair, *Hitler’s U-boat War: The Hunted 1942–1945*, pp. xi–xii.

23 Sternhell and Thorndike, *Antisubmarine Warfare in World War II*, p. 83.

24 Niestlé, *German U-Boat Losses During World War II*, p. 4. There remains considerable uncertainty in most data on the Battle of the Atlantic. In the case of U-boats, Niestlé noted in 2014 that the number “now recorded as lost to unknown cause has markedly increased.” Ibid., p. 3.

25 A. Timothy Warnock, *Air Power versus U-boats: Confronting Hitler’s Submarine Menace in the European Theater* (Washington, DC: Air Force History and Museums Program, 1999), p. 7.

26 OP-20-G, *Battle of the Atlantic*, Vol. I, *Allied Communications Intelligence, December 1942–May 1945*, SRH-009 (Washington, DC: National Security Agency, undated), available at http://www.ibiblio.org/hyperwar/ETO/Ultra/SRH-009/.

TABLE 3: RESULTS, NORTH ATLANTIC ROUTES, JANUARY–MAY 1943[27]

	Allied North Atlantic convoys (by arrival date)	Merchants in North Atlantic convoys	Merchants sunk (Metcalf)	U-boat patrols (by date sailed)	U-boats sunk (by date lost) (Blair)	U-boat patrols with no sinkings
January-43	14	461	18	58	3	33
February-43	12	529	20	69	14	37
March-43	15	687	58	49	11	29
April-43	16	658	16	90	12	69
May-43	21	922	19	58	35	46
Totals	**78**	**3,257**	**131**	**324**	**75**	**214**

Cumulative attrition figures aside, why did the Allies win the Battle of the Atlantic in the spring of 1943? To begin to answer this question, the focus needs to shift to more fine-grained metrics, particularly to the relative loss rates in early 1943 between Allied merchants and the U-boats. Table 3 shows that the U-boats had considerable success in March 1943 against Allied shipping across the North Atlantic, but by May U-boat losses to Allied anti-submarine warfare (ASW) forces became unsustainable.

The first point to be made about Table 3 is that the 131 Allied merchant vessels sunk by U-boats in the North Atlantic during the first five months of 1943 constituted only 4 percent of the 3,257 merchants that sailed between North America and the British Isles; 96 percent of the merchants got through. Second, while losing more than 30 merchants from four east-bound convoys (Halifax 228 and 229 and Slow Convoys 121 and 122) to the U-boats during the first 20 days of March undoubtedly rattled the British Admiralty, the other 11 North Atlantic convoys, like most of the merchants they contained, got through unscathed.[28] This reflected the effectiveness of using evasive routing to avoid the U-boats, a strategy that had been the British Admiralty's "first line of defence for Britain's ocean-supply-line" since early in the war.[29] As Admiral King emphasized at the March 1943 Atlantic convoy conference, convoying was the *only* way of protecting the slower Allied merchant ships; hunter-killer operations were secondary to that strategic principle.[30] The fundamental operational challenge for both sides

27 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, Appendix 2, pp. 737–744; and Metcalf, *History of Convoy and Routing*, Appendix G, pp. G1–G5. Blair, based on Jürgen Rohwer's *Critical Convoy Battles of World War II: Crisis in the North Atlantic, March 1943* and *Axis Submarine Successes, 1939–45*, maintains that merchant losses during January–May 1943 totaled 118 ships vice the 131 in Metcalf. The losses are for the eastbound Halifax (HX) and Slow Convoy (SC) and westbound Outbound North (ON) and Outbound North Slow (ONS) convoys.

28 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 167.

29 Bennett, *Behind the Battle*, p. 173. Merchant ships that could sustain speeds of 13 knots or greater enjoyed considerable protection from the U-boats. O'Brien, *How the War Was Won*, p. 254. Ships like the *Queen Mary* that could travel up to 30 knots were virtually invulnerable.

30 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 241.

had been to find the other in the vastness of the Atlantic, and doing so had proven difficult. Third, many of the U-boats that aborted their patrols did so because of the damage inflicted by increasingly capable and numerous Allied ASW forces.[31] Finally, the percentage of U-boat patrols that failed to sink any Allied ships grew from half in January and February to three-quarters in April and May.

Table 4 uses the data in Table 3 to compute trends in the percentages of merchants and U-boats sunk during the first five months of 1943, the percentages of U-boats that achieved no sinkings, and the monthly merchant/U-boat exchange ratios. The most revealing of these trends is the collapse of U-boat productivity in May 1943 as measured by the merchant/U-boat exchange ratio. In that month the Germans lost almost two submarines for every Allied merchant sunk in the North Atlantic. As Karl Dönitz later wrote:

> The U-boat losses, which previously had been 13 percent of all the boats at sea, rose rapidly to 30 to 50 percent. These losses were suffered not only in convoy attacks, but everywhere at sea. There was no part of the Atlantic where the boats were safe from being located day and night by aircraft. All the U-boat entrance and exit channels in the Bay of Biscay were, in particular, most carefully watched. Losses here were especially high.[32]

TABLE 4: TRENDS, NORTH ATLANTIC ROUTES, FEBRUARY–MAY 1943[33]

	Percentage of convoy merchants sunk	Percentage of U-boats sunk	U-boat patrols with no sinkings	Merchant/U-boat exchange ratio
February-43	3.8%	20.3%	53.6%	**1.43**
March-43	8.4%	22.4%	59.2%	**5.27**
April-43	2.4%	13.3%	76.7%	**1.33**
May-43	2.1%	60.3%	79.3%	**0.54**

31 For data on U-boat aborts during January–June 1943, see Appendix 12 in Blair, *Hitler's U-boat War: The Hunted 1942–1945*, pp. 801–803.

32 Dönitz, *The Conduct of the War at Sea*, p. 23.

33 January 1943 has been dropped because the atrocious weather in the North Atlantic during that month severely limited the ability of the U-boats to locate Allied convoys. For similar estimates of the merchant/U-boat ratio, see C. E. Behrens, Operations Evaluation Group (OEG), *Effects on U-boat Performance of Intelligence Obtained from Decryption of Allied Communications*, OEG Study 533 (Arlington, VA: Center for Naval Analyses, April 28, 1954), p. 7. Behrens' 1954 report was informed by access to the *B.d.U.* war diary. His total of U-boats lost in the North Atlantic during February–May 1943 is 60 (vice the 73 in Table 3). Of Behrens' total, 30 were sunk by land-based aircraft, 27 by warships, and three by aircraft from merchant "jeep" escort carriers.

As a result, Dönitz was forced to suspend his campaign against Allied shipping in the North Atlantic on May 26, 1943.[34]

Dönitz's near-term response was to install heavier anti-aircraft armament on the older-type U-boats in hopes that these "Flak U-boats" could fight off attacks by Allied aircraft, thereby enabling other U-boats to attack on the surface. If they could, then the U-boats would use newly deployed acoustic homing torpedoes to break up the Allied escorts before turning to their main task of sinking Allied merchant vessels. But when this approach was initially tried in September 1943, the U-boats that got past the escorts were foiled because the convoy concealed itself behind a smoke screen.[35] The U-boats' situation in the North Atlantic became even worse in October 1943: the exchange rate was one merchant vessel for seven U-boats.[36] Thus Dönitz's Flak-U-boats failed, forcing him to conclude that the era of surface warfare for the U-boats "had come to an end," implying that the German Navy would have to bide its time until next-generation U-boats designed to reach high speeds underwater and to achieve longer ranges became available.[37]

The data in Tables 3 and 4 provide quantitative metrics that confirm the U-boats' defeat by May 1943. They do not, however, reveal the underlying causes of the Allied victory. These underlying causal factors can be grouped into two broad categories: (1) the contest between British and German cryptologists and (2) the quantitative growth and technological improvements in Allied ASW capabilities that began to affect the U-boat war in the early months of 1943.[38]

Early in the war, Dönitz decided to exercise direct centralized tactical control of the U-boats from his headquarters ashore. Individual U-boats generally received their operational orders and heading points by radio after putting to sea; thereafter they were required to send *B.d.U.* passage and position reports, provide detailed tactical situation reports whenever Allied convoys were detected, report fuel on hand with every transmission, and could not deviate from orders or start for home without receiving specific permission from *B.d.U.*[39] This book-keeping approach generated a wealth of high-frequency radio traffic between *B.d.U.* and the U-boats. Monitored and recorded by a network of British and American intercept sites,

34 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, pp. 339–340. In an attempt to mask the withdrawal of the U-boats from the North Atlantic front, Dönitz directed 13 submarines to remain there as long as possible and broadcast dummy messages. The Allies learned of the deception from naval Enigma.

35 Dönitz, *The Conduct of the War at Sea*, pp. 23–24. The Type XXI U-boat is considered the first true submarine.

36 Behrens, OEG, *Effects on U-boat Performance of Intelligence Obtained from Decryption of Allied Communications*, p. 8.

37 Dönitz, *The Conduct of the War at Sea*, p. 24; and OP-20-G, *Battle of the Atlantic*, Vol. I, p. 54.

38 A code is a mapping between some meaningful units (words, sentences, phrases) into something else, usually a shorter group of symbols. For example, "accountant" has been used as a code for "Come at once. Do not delay." A cipher is a mechanical operation or algorithm performed on individual or small groups of letters—for example the Caesar cipher that replaces each plain text letter with a letter some fixed number of positions up or down the alphabet. Codes deal with meaning, whereas ciphers deal with syntax.

39 OP-20-G, *Battle of the Atlantic*, Vol. I, pp. 21–23; and Dönitz, *The Conduct of the War at Sea*, p. 19.

the Germans' unrestricted radio traffic constituted a weakness that, if the Allies managed to exploit it, could strip the U-boats of the stealth that was their fundamental advantage.

Of course, messages to and from the U-boats were encrypted—initially using three-rotor Enigma machines with plugboards.[40] In May 1941 the British Government Code and Cipher School (GC&CS) at Bletchley Park broke the three-rotor Enigma, and by June it was able to decipher U-boat radio traffic concurrently.[41] But in February 1942 the German Navy introduced four-rotor Enigma machines, and Bletchley Park did not break the four-rotor naval Enigma until December. From January to May 1943 there were 23 days that could not be read at all.[42]

During this same period, the German Navy's observation service, *Beobachtungs-Dienst* (*B-Dienst*), was reading the encrypted messages the Allies used to communicate with the convoys via their Broadcast to Allied Merchant Ships (BAMS).[43] At the time these communications were encrypted with the British Admiralty's Naval Cipher No. 3. *B-Dienst* broke Naval Cipher No. 3 in September 1942, and the British Admiralty, reluctant to believe it had been broken, did not switch to Naval Cipher No. 5 until June 1943, after the Battle of the Atlantic had turned decisively against the Germans.[44]

The upshot of both sides reading the other's communications was a "gigantic game of naval chess on the North Atlantic run" between the opposing command, control, and communications systems (C^3).[45] Using naval Enigma decrypts and high-frequency direction finding that located U-boats and their intercept lines, the Allies rerouted convoys to avoid the German submarines. But timely *B-Dienst* intercepts of rerouting instructions to the convoys enabled the U-boats to respond. This cat-and-mouse game between Bletchley Park and *B-Dienst* cryptographers produced some epic convoy battles. The costliest exchange occurred during the ten days from March 10–19, 1943 in which Bletchley Park temporarily lost naval Enigma.[46] The protracted battle involved nearly 40 U-boats and two eastbound convoys: Slow Convoy 122 and Halifax 229. When all was said and done, SC 122 lost nine out of 50 merchants

40 For details on the history and workings of the German Navy's Enigma machines, see Andrew Hodges, *Alan Turing: The Enigma* (Princeton and Oxford: Princeton University Press, 2014), pp. 202–222. The plugboard increased the number of states of the Enigma machine to 1,305,093,289,500 ways of connecting the seven pairs of letters on the plugboard for each of the 6 x 17,576 rotor states. Ibid., p. 214.

41 Hodges, *Alan Turing: The Enigma*, p. 253.

42 OP-20-G, *Battle of the Atlantic*, Vol. I, p. 20; and Hodges, *Alan Turing: The Enigma*, pp. 280–283.

43 David Kahn, *The Codebreakers: The Comprehensive History of Secret Communication from Ancient Times to the Internet* (New York: Scriber, 1996), pp. 465–466.

44 Bennett, *Behind the Battle*, p. 177; and OP-20-G, *Battle of the Atlantic*, Vol. III, *German Naval Communication Intelligence*, SRH-024 (Washington, DC: National Security Agency, undated), pp. 35–45, available at https://www.history.navy.mil/research/library/online-reading-room/title-list-alphabetically/b/battle-atlantic-volume-3-german-naval-communication-iIntelligence.html.

45 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 191. Eventually Bletchley Park was able not just to read messages but capture the entire U-boat C3 system. Hodges, *Alan Turing: The Enigma*, p. 241.

46 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 257.

(18 percent), and HX 229 lost 13 out of 40 merchants (32.5 percent); the cost to the *Kriegsmarine* was only two U-boats (U-384 and U-665).[47]

Despite this signal success, the tide was turning rapidly and decisively against the U-boats. In June 1943 Naval Cipher No. 5 replaced Naval Cipher No. 3, largely blinding *B-Dienst* to evasive routing instructions sent to Allied convoys for the rest of the war.[48] Starting in September 1943 the Allies were able to read naval Enigma continuously and completely, generally within less than 24 hours.[49] Thus, by the fall of 1943, Allied cryptographers had gained a major advantage over German cryptographers that persisted to the war's end.[50]

Even before then, the growing numbers and lethality of Allied ASW forces were making U-boat operations increasingly perilous and unproductive. As Clay Blair summarized this turn of events:

> By the end of April 1943, the Germans confronted a naval rout in the Atlantic. Allied centimetric-wavelength radar, land- and ship-based Huff Duff [high-frequency direction finding], and superb intelligence evaluation and operational-research teams in London and Washington had finally defeated the U-boat force. No existing production U-boats could enter combat in the North Atlantic with even the slightest confidence. Attacks on Allied convoys had become near-suicidal endeavors and would only become more dangerous.[51]

As early as March, Allied escort groups began growing larger. In 1942 the average number of escorts assigned to North and Mid-Atlantic convoys was 6.1; in 1943 it grew to 8.3 escorts per convoy.[52] By April escort aircraft carriers and Merchant Aircraft Carriers were starting to appear in well-trained, aggressive escort groups.[53] Very long-range aircraft, notably the B-24, were beginning to close the Mid-Atlantic gap, and around 60 of these bombers were assigned to frontline ASW squadrons in the North Atlantic.[54] Whereas the kill rate of Allied

---

47 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, pp. 264–266. See also Hodges, *Alan Turing: The Enigma*, pp. 325–329.

48 OP-20-G, *Battle of the Atlantic*, Vol. III, p. 45.

49 OP-20-G, *Battle of the Atlantic*, Vol. I, p. 20. Enigma encrypted messages were intercepted by a network of British and American radio-intercept stations. High frequency direction finding (Huff Duff) also plotted U-boat locations based on their transmissions. For details on the U.S. intercept net, see OP-20-G, *Battle of the Atlantic*, Vol. I, pp. xv, xxi–xxiii. Enigma messages were then decrypted by electromechanical computers (or Bombes), which translators and intelligence officers then rendered into militarily comprehensible English. In the case of GC&CS, Hut 8 at Bletchley Park did the decryption, and Hut 4 translated and interpreted the German messages. Bennett, *Behind the Battle*, p. xxiii.

50 By the end of 1943, decrypting and interpreting naval Enigma traffic was taken over by the U.S. Navy. Hodges, *Alan Turing: The Enigma*, p. 329.

51 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 311.

52 Metcalf, *History of Convoy and Routing*, p. 31.

53 Sternhell and Thorndike, *Antisubmarine Warfare in World War II*, p. 36; and Warnock, A*ir Power versus U-boats*, pp. 7, 9–10. The Mid-Atlantic gap was a 500-mile wide area between 25ºW and 45ºW longitude in which the U-boats had been able to operate with little fear of land-based Allied ASW aircraft prior to early 1943.

54 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 800. Counting B-24 units enroute, equipping, or working up, over 200 B-24s were assigned to North Atlantic ASW in March 1943. A B-24 could dwell over a convoy for three hours at a range of 1,000 miles from its base.

aircraft against U-boats during World War I had been "essentially zero," the addition of Leigh Lights for pursuing U-boats at night, airborne ASV (Air-to-Surface Vessel) Mark III 10-centimeter radars using the cavity magnetron, and the American air-delivered "Fido" acoustic homing torpedo during World War II made aircraft the most effective U-boat killers.[55] Because German radar warning receivers could not detect ASV Mark III radars, they enabled the aircraft to surprise surfaced U-boats.[56] By the war's end, Allied aircraft were credited with sinking over 320 U-boats. And, circling back to intelligence on the locations of patrolling U-boats, the Allies not only developed high-frequency direction finding (DF) that was more accurate than the Germans thought possible, but also, no less importantly, they improved "the processing of DF information." Miniaturization likewise permitted more and more escort vessels to be equipped with Huff Duff.[57]

At the same time, there were a number of weaknesses on the German side. The Type VII and IX U-boats were not true submarines. Their underwater speeds—7–8 knots—were several knots slower than the 9–10 knots of the faster Allied merchant ships. Even a 1- or 2-knot advantage over the U-boats could provide a large boost in protection, as British operations research discovered.[58] In addition, the submerged endurance and cruising ranges of Type VIIs and IXs were limited, their lookout capabilities poor, and their vulnerability to Allied anti-submarine weaponry substantial.[59] Furthermore, Dönitz uncritically accepted consistently inflated estimates of the tonnages his U-boat captains reported sinking and was reassured in late 1941 and again in early 1943 that naval Enigma remained secure.[60]

What, then, were the principal criteria that led to the U-boats' defeat in the Battle of the Atlantic during the spring of 1943? The monthly loss data (Allied merchant tonnage and U-boats sunk) are comparable to the score of a football game from one point in the contest to the next: they reflect which team was ahead at any stage but give no insight into why one team had the lead at any point—or ultimately won. In the football analogy, the winning team's offensive line might have prevented its quarterback from being hurried or sacked; the victor might have had a particularly effective defense; turnovers and special teams could have provided the margin of victory; or, either team could have made critical mistakes or they

55 Blair, *Hitler's U-boat War: The Hunters 1939–1942*, p. 14.

56 Blair, *Hitler's U-boat War: The Hunted 1942–1945*, p. 30.

57 Blair, *Hitler's U-boat War: The Hunters 1939–1942*, pp. 270, 426; and *Hitler's U-boat War: The Hunted 1942–1945*, pp. 791–792.

58 O'Brien, *How the War Was Won*, p. 255. O'Brien credits P. M. S. Blackett and his operations researchers with the discovery that the determining factor in merchant ship losses for any convoy was not the relative number of escort vessels to merchant ships but the number of escort vessels to submarines attacking. Ibid., 256.

59 OP-20-G, *Battle of the Atlantic*, Vol. I, p. 18.

60 For some of Rohwer's examples of inflated German tonnage claims, see Blair, *Hitler's U-boat War: The Hunters 1939–1942*, pp. 199–200. For the dismissal in 1941 of concerns about whether the Allies were reading the three-rotor naval Enigma traffic, see Ibid., pp. 386–387. For the dismissal of similar concerns in early 1943, see *Hitler's U-boat War: The Hunted 1942–1945*, p. 256; and Bennett, *Behind the Battle*, p. 177. During World War II Bennett was a leading producer of Ultra intelligence at Bletchley Park.

simply got lucky. It is the interplay of such factors that drives the score throughout the contest and leads to its eventual outcome. In the case of the Battle of the Atlantic, there is a long list of causal factors whose interactions affected the ebb and flow of the campaign: the decrypts of German naval Enigma traffic being but one among many.[61] Reflecting what had emerged as generally accepted view once the secret of Ultra intelligence was revealed in the 1970s, Ralph Bennett argued that the balance of informed opinion "favors Ultra as the decisive element which created a new situation" that eventually tipped the contest in favor of the Allies.[62] But he also observed that when events have multiple, shifting causes, "It is unwise to offer comparative estimates of the responsibilities of each."[63] Furthermore, he was unquestionably right to point out that even the best intelligence is only valuable if accompanied by the means to exploit it, and by April 1943 the Allies were acquiring the requisite ASW capabilities to do so.[64]

Several broader points about effectiveness criteria emerge from these observations. First, although the obvious global metrics—monthly losses of Allied merchant vessels, monthly U-boat losses, and the monthly average of U-boats at sea—reflect the campaign's ebb and flow, they do not shed much light on why the Battle of the Atlantic unfolded as it did. Second, the drivers—the real causes underlying the global data—involved the complex interplay of various technical and tactical factors: the changes and innovations that the two sides either made over time (or failed to make). Third, given the complexity of the interactions between so many different factors over such a protracted campaign, the conclusion that the cryptologic competition was the decisive element is, at best, a matter of judgment rather than a provable proposition in any rigorous sense.[65] The causal complexity of the Battle of the Atlantic is simply too great. Again, the aggregate metrics traditionally used to track the campaign's ebb and flow were driven by more fundamental factors. One obvious example would be the changes in the merchant/U-boat exchange ratio. Another criterion, which may have been fundamental, is the change over time in the frequency with which Allied convoys were able to avoid being located by the U-boats. And in both cases, the most important aspect of these metrics was their rates of change.

61 "Ultra" was the designation adopted by British intelligence in June 1941 for wartime signals intelligence obtained by breaking high-level encrypted enemy radio and teleprinter communications at GC&CS.

62 Bennett, *Behind the Battle*, p.182. Those Bennett cited as representing the balance of informed opinion included Patrick Beesly, Jürgen Rohwer, and Harry Hinsley. During the war, Hinsley had been the chief intelligence officer of the Naval Section at GC&CS, and he was later the chief author of the official British intelligence history of World War II.

63 Bennett, *Behind the Battle*, p. 182.

64 Bennett, *Behind the Battle*, pp. 200–201.

65 O'Brien argues that Ultra, while important, was a "subsidiary player" in the outcome of the Battle of the Atlantic. O'Brien, *How the War Was Won*, p. 242. We shall return to the issue of monocausal explanations of major historical changes such as the industrial revolution in Chapter 3.

The Strategic Bombing of Germany

In April 1943 the British and American air forces agreed on a coordinated plan for the strategic bombing of Germany. In May the Combined Chiefs of Staff approved this plan at the Trident conference under the codename POINTBLANK.[66] The stated aim of the Anglo–American Combined Bomber Offensive was the "progressive destruction and dislocation of the German military, industrial and economic system and the undermining of the morale of the German people to a point where their capacity for armed resistance is fatally weakened."[67] The CBO plan to accomplish this aim envisioned attacking six target systems containing 76 precision targets during the daytime while Royal Air Force (RAF) Bomber Command (BC) continued area attacks on industrial centers at night.[68] In addition, the plan identified German fighter strength "as an intermediate objective second to none in priority."[69]

To give a sense of the weight of the British effort during the year or so before the CBO, the night of May 30–31, 1942, RAF BC dispatched 1,046 bombers against the city of Cologne (of which, 940 attacked).[70] And in July 1943 RAF BC succeeded in burning Hamburg to the ground, killing at least 40,000 people over the course of nine nights. On the first night, July 24–25, Bomber Command's Air Marshall Arthur "Bomber" Harris sent 792 bombers against the city; on the final night, Bomber Command attacked with 740.[71]

The CBO envisioned the U.S. Eighth Air Force in England needing similarly large forces to attack the 76 precision targets identified in the plan. American planners estimated that by June 1943 the Eighth Air Force would need 944 heavy and 200 medium bombers plus extensive fighter support to accomplish the first phase of the American daylight bombing program; the CBO plan projected that by December 1943, the Eighth Air Force would need 1,746 heavy and 600 medium bombers.[72] Thus, by early 1943 both the British and the Americans were investing heavily in strategic bombing.

66 Robert C. Ehrhart, "The European Theater of Operations, 1943–1945," in John F. Kreis, ed., *Piercing the Fog: Intelligence and Army Air Forces Operations in World War II* (Washington, DC: Air Force History and Museums Program, 1996), p. 194.

67 U.S. Eighth Air Force, "The Combined Bomber Offensive from the U.K.," April 12, 1943, U.S. National Archives, RG 218, CCS 381, Box 594, 2, p. 1. Emphasis in the original.

68 The target systems were: (1) submarine construction yards, (2) the German aircraft industry, (3) ball bearings, (4) oil, (5) synthetic rubber and tires, and (6) military transport vehicles. Eighth Air Force, "The Combined Bomber Offensive from the U.K.," pp. 1–2.

69 Eighth Air Force, "The Combined Bomber Offensive from the U.K.," p. 3.

70 Williamson Murray and Allan R. Millett, *A War to Be Won: Fighting the Second World War* (Cambridge and London: Belknap Press, 2000), pp. 307–308; and Richard G. Davis, *Bombing the European Axis Powers: A Historical Digest of the Combined Bomber Offensive, 1939–1945* (Maxwell Air Force Base, AL: Air University Press, April 2006), p. 65.

71 O'Brien, *How the War Was Won*, p. 279. The firestorm occurred on the second night of Operation Gomorrah. "The inferno reached temperatures as high as 1,000°F, while superheated air rushed through the city at speeds close to 300 mph." Murray and Millett, *A War to Be Won*, p. 310.

72 Eighth Air Force, "The Combined Bomber Offensive from the U.K.," p. 5.

Even before the United States entered the war, President Franklin Roosevelt had decided that "Complete dominance in the air was the key step to achieving victory," and that control of the air could be gained by sheer numbers of aircraft.[73] In January 1942 he, therefore, set ambitious aircraft production targets for 1942 and 1943 (60,000 and 125,000 planes, respectively).[74] After it became clear in August that production was going to fall considerably short of these goals, Roosevelt and Harry Hopkins made it clear to General George Marshall and Donald Nelson, chairman of the War Production Board, that the aircraft program would be given highest priority and every effort would be made to produce 107,000 planes in 1943 (of which 82,000 would be combat types).[75] Phillips O'Brien argues that this was Roosevelt's "most important and crucial military intervention" of the war.[76]

Given Roosevelt's views about air power's importance in mid-twentieth-century warfare, it is not surprising that in late 1944 he suggested to Henry Stimson, the secretary of war, that there was a need for "a comprehensive survey of the effects of the strategic bombing offensive against Germany."[77] Stimson, in turn, had General H. H. "Hap" Arnold, commander of the U.S. Army Air Forces (USAAF), persuade the 67-year-old Franklin D'Olier, then president of the Prudential Insurance Company of America, to chair the effort to evaluate the effects of strategic bombing in Europe.[78] On October 21, 1944, D'Olier met with General Arnold and Robert Lovett, Stimson's assistant secretary of war for air, and agreed to take on the challenging task of running the bombing survey.[79] A series of implementing directives established the United States Strategic Bombing Survey (USSBS) on November 3, 1944.[80] The organization grew rapidly. By August 1945 the survey employed over 300 civilians and some 1,400 military personnel.[81] The USSBS ultimately produced 204 reports on the war in Europe and another 105 on the Pacific theater of operations.

73 O'Brien, *How the War Was Won*, p. 46. Roosevelt's views on the importance of air power were reinforced in May 1940 by the reporting of William Bullitt, the U.S. ambassador in Paris, during the fall of France. Ibid., p. 128.

74 These aircraft production goals were not pulled out of the air. They resulted from interaction between the president, Army chief of staff General George Marshall, and chief of the Air Corps Lieutenant General H. H. "Hap" Arnold. David MacIsaac, *Strategic Bombing in World War II: The Story of the United States Strategic Bombing Survey* (New York and London: Garland, 1976), pp. 10–12. This interaction started in the summer of 1941 with an air staff planning exercise, AWPD-1 (Air War Plans Divisiion-1) that resulted from Roosevelt asking the secretaries of the Army and Navy to estimate the "overall production requirements required to defeat our potential enemies." Ibid., p. 12.

75 O'Brien, *How the War Was Won*, pp. 51–52, 156. The United States produced 299,230 aircraft during World War II, of which 31,890 were heavy bombers (B-17s and B-24s), 3,764 were very heavy bombers (B-29s), and 100,554 were fighters. Office of Statistical Control, *United States Air Force Statistical Digest World War II* (Washington, DC: Comptroller of the Air Force, December 1945), p. 112.

76 O'Brien, *How the War Was Won*, pp. 49, 243.

77 MacIsaac, *Strategic Bombing in World War II*, pp. x, 3.

78 Ibid., pp. 1–3, 21.

79 Ibid., pp. 51–52.

80 Ibid., p. 56.

81 Ibid., p. 68. In Europe, four members of the USSBS were killed and four others wounded. Ibid., p. 94.

The USSBS's assessment of the effects of strategic bombing on the German war economy was largely the work of Overall Economics Effects Division headed by the economist John Kenneth Galbraith and his assistant director Burton H. Klein. Galbraith's division reached three main conclusions using the output of German munitions plants as the primary analytic criterion for judging the CBO's effectiveness.[82] First, prior to mid-1943 Anglo–American strategic bombing "had no appreciable effect either on German munitions production or on the national output in general."[83] During the first three years of the conflict, the Germans did not mobilize their economy for a long war, and the weight of British and American strategic bombing was insufficient to offset the slack in the German economy. By 1943 the value of German armaments output was about one-third of total industrial production and approximately 15 percent of Germany's gross national product (GNP).[84] Second, as the CBO intensified in 1944, the loss from direct damage "was less important than the sacrifice in output caused by dispersal and other defensive measures taken."[85] Bomber Command had shifted to area bombing of German cities at night in 1940, while the Americans later opted for precision bombing during the daytime—an approach that proved unsustainable without fighters that could escort the bombers all the way to their targets and back. Not until early 1944 did the Eighth Air Force receive enough P-51 long-range fighters to provide the fighter escorts needed to sustain daylight attacks against industrial targets deep in Germany.[86] Third, while the index of total German munitions output continued to grow through mid-1944, the loss of armaments output that could be credited to bombing rose from about 5 percent in the last half of 1943 to somewhat above 15 percent in the last half of 1944.[87] Nor did strategic bombing erode support for the war effort among the German populace until the final months of 1944 and early 1945 due largely to "the terrorist control of the population by the Nazis and, in part, to the cultural patterns of the German people."[88]

---

82 From the 1930s through World War II, American airmen understood the word "strategic" as meaning bombing *independent* of land or sea campaigns. Subsequently, with the advent of atomic and thermonuclear weapons, the word came to mean *nuclear*.

83 United States Strategic Bombing Survey (USSBS), *The Effects of Strategic Bombing on the German War Economy* (Washington, DC: U.S. GPO, 1945), p. 11.

84 Ibid., p. 139. German GNP was the value in reichsmarks of the goods and services produced within Germany during a given year or quarter. For fuller definition, see footnote 64 in Chapter 3.

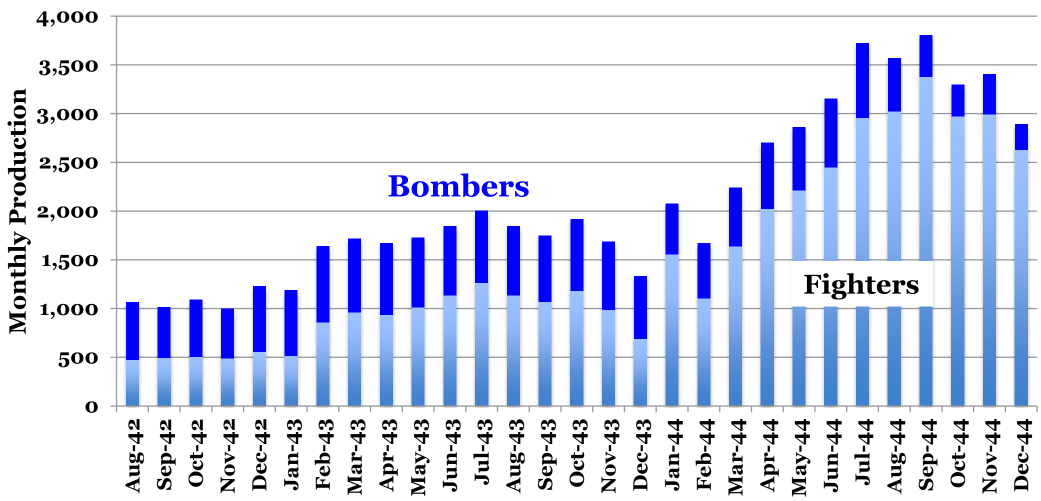
85 Ibid., p. 11. The Eighth Air Force dropped 687,052 tons of ordnance during World War II, and Bomber Command dropped 1,011,510 tons. Davis, *Bombing the European Axis Powers*, p. 568.

86 Ehrhart, "The European Theater of Operations, 1943–1945" in Kreis, *Piercing the Fog: Intelligence and Army Air Forces Operations in World War II*, pp. 198–199.

87 USSBS, *The Effects of Strategic Bombing on the German War Economy*, p. 13.

88 USSBS, *The Effects of Strategic Bombing on German Morale*, Vol. I (Washington, DC: U.S. GPO, 1947), p. 7. See also Bennett, *Behind the Battle*, pp. 134–136, 145, 147–150, 159.

FIGURE 1: GERMAN FIGHTER AND BOMBER PRODUCTION, AUGUST 1942–DECEMBER 1944[89]



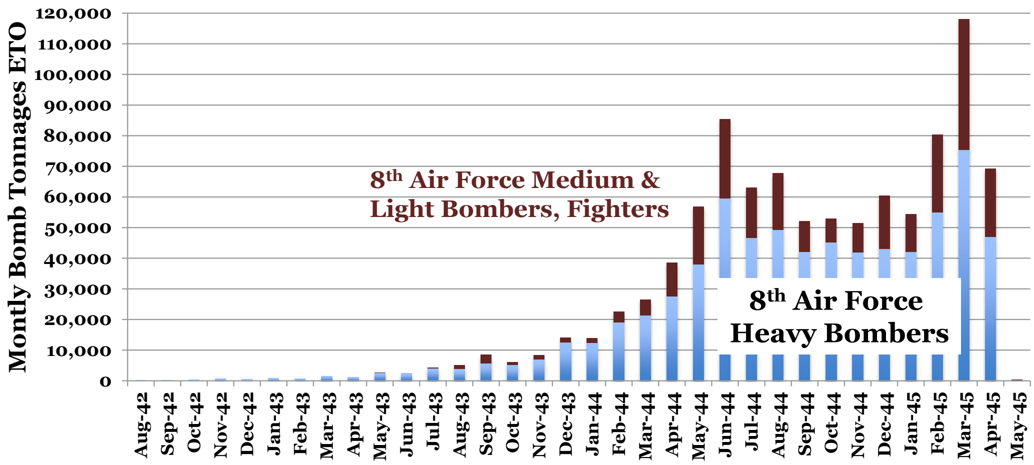
The extent to which Allied strategic bombing failed to bring about the advertised effects on German war production until late 1944 was not apparent until after the war in Europe had ended. In contrast to the role decrypts of Enigma traffic played in the Battle of the Atlantic, Ultra provided little insight into the effects of strategic bombing. Ample landline communications existed in France and Germany for conducting Luftwaffe air operations and managing Germany's wartime economy without the heavy use of vulnerable radio communications. "The result was that no worthwhile assessment of the effects of bombing could be secured" during the war either by post-strike aerial photography or communications intelligence.[90] Only after the USSBS began acquiring German records did it become possible to assess the direct effects of the strategic bombing campaign, particularly on German aircraft production.

Again, the CBO plan identified German fighter strength in Western Europe as an intermediate objective second to none. This was especially true for Eighth Air Force's doctrinal commitment to daylight precision bombing. But as Figure 1 makes clear, German fighter production continued to rise until finally peaking in September 1944. Yet, from late 1943 through the Normandy landings on June 6, 1944, the weight of the Eighth Air Force's daylight strategic bombing in the European Theater of Operations (ETO) grew steadily, as Figure 2 documents. Taken together Figures 1 and 2 can be interpreted as making a *prima facie* case that strategic bombing was a failure.

89 USSBS, *The Effects of Strategic Bombing on the German War Economy*, Table 102, p. 277. The U.S. Eighth Air Force flew its first heavy bomber mission against targets in occupied France on August 17, 1942.

90 Bennett, *Behind the Battle*, p. 136. "Factory managers' damage reports and estimates of repair-time and the return of something like normality went by telephone or teleprinter," which meant that Enigma decrypts provided little insight into the effects of Allied strategic bombing. Moreover, in 1943 the Allies had few agents in Germany that could provide human intelligence on the effects of POINTBLANK.

FIGURE 2: 8TH AIR FORCE BOMB TONNAGES, AUGUST 1942–MAY 1945[91]



The principal criterion for reaching this conclusion is the campaign's lack of a prompt, direct impact on German war production: the quantities of armaments that emerged from the factory doors over time. Galbraith later argued that a general disruption of the German war economy "could not be meaningful" if it did not affect the production of military end items such as Panzers (a term encompassing tanks, assault guns, and self-propelled guns); anti-aircraft artillery; ammunition; and, above all, single-engine fighters. He noted in 1981 that average monthly Panzer production grew steadily from 136 in 1940 (the first full year of the war) to 1,005 in 1943 (the year the bombing began in earnest) and reached 1,583 in 1944.[92] In the case of the intermediate objective second to none, output grew through July 1943, suffered some plateaus and downturns during late 1943 and early 1944, and then, with the exception of one downturn in August 1944, grew to a peak in September (see Figure 1).

At the same time, the cost in blood and treasure to the Allies was substantial. From August 1942 to May 1945 the U.S. Eighth Air Force operating out of England lost over 4,100 aircraft on operational missions and suffered perhaps 58,000 casualties, of which some 18,000 were killed, 6,500 wounded, and 33,500 missing in action.[93] Bomber Command paid an even greater price, losing over 7,000 aircraft on wartime operations and sustaining some 64,100 operational casualties, of which over 47,000 were killed.[94] Moreover, both organizations lost large numbers of aircraft and aircrews in non-operational accidents. Small wonder, then, that Galbraith, who was appalled at the wholesale destruction of German cities, concluded that the aircraft, manpower, and ordnance consumed by the CBO had cost the American and British

91 Office of Statistical Control, *United States Air Force Statistical Digest World War II*, Table 143, p. 243. Figure 2 does not contain bombing in the Mediterranean Theater of Operations.

92 John Kenneth Galbraith, *A Life in Our Times* (Boston: Houghton Mifflin, 1981), p. 205.

93 Davis, *Bombing the European Axis Powers*, pp. 568, 585.

94 Ibid., pp. 568, 583–584.

economies far more in output than they had cost Germany.[95] The Allied decision to pursue the strategic bombing of German's war industry had been "one of the greatest, perhaps the greatest, miscalculation of the war."[96]

This harsh judgment only follows, however, if one focuses narrowly on the output of German armaments factories under Allied bombing as the sole metric for assessing the CBO. It ignores the broader, more indirect effects strategic bombing had on Germany's overall war effort and suggests that it would be wise to explore some alternative metrics, starting with Germany's allocation of air defense assets among the main operational theaters. In mid-1943 German leaders concluded that "by far the greatest threat to German power" was the Combined Bomber Offensive.[97] This decision led to a major redeployment of Luftwaffe aircraft by the end of 1943. In December 1942, the Luftwaffe had around 1,500 aircraft deployed both in Russia and in the Reich, with another 900 aircraft in the Mediterranean and North Africa. A year later 54 percent of all German aircraft and almost 70 percent of Luftwaffe fighters were in Germany or on the Western Front.[98] Indeed, as early as September 1943 almost all German fighters had been withdrawn from the Mediterranean. In Russia, Luftflotte 6, which was tasked with protecting Army Group Centre, had only 40 single-engine fighters in working order.[99] This withdrawal of Luftwaffe fighters back to the Reich and the Western Front stripped German ground forces in Russia and the Mediterranean of fighter cover and direct air support—developments that proved especially catastrophic on the Russian front during the Red Army's Belorussian offensive (Operation Bagration, June 23–August 29, 1944). Against the Army Group Centre's handful of operational single-engine fighters, the Soviet Air Force deployed some 5,400 aircraft and over 2,500 fighters to support the three fronts that mounted Operation Bagration and averaged some 4,500 sorties a day through July 4, 1944.[100] Without adequate fighter cover, German ground forces were either pinned down and immobile or subjected to Soviet air and ground attack when they attempted to retreat. Army Group Centre's 25 divisions were largely encircled and destroyed within 12 days.[101] By the end of August, this offensive had destroyed the Germans' Army Group Centre and carried the Red Army some 500 kilometers westward to the borders of East Prussia.[102]

95 Galbraith, *A Life in Our Times*, pp. 299, 226.

96 Ibid., p. 206.

97 O'Brien, *How the War Was Won*, pp. 290. Shortly after Hamburg was burned to the ground, Field Marshal Erhard Milch, the Luftwaffe's chief of production, announced that Hitler had put top priority on fighter production, which was to increase to 2,000 a month by mid-1944. Murray and Millett, *A War to Be Won*, p. 315.

98 O'Brien, *How the War Was Won*, pp. 290–291; and Murray and Millett, *A War to Be Won*, p. 316.

99 O'Brien, *How the War Was Won*, pp. 360–361.

100 Von Hardesty, *Red Phoenix: The Rise of Soviet Air Power, 1941–1945* (Washington, DC: Smithsonian Institution Press, 1982), p. 195; and O'Brien, *How the War Was Won*, p. 363.

101 Murray and Millett, *A War to Be Won*, p. 448; and Von Hardesty, *Red Phoenix*, p. 196.

102 David M. Glantz, "The Great Patriotic War and the Maturation of Soviet Operational Art: 1941–1945," draft, Soviet Army Studies Office, April 1987, p. 76, available at http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA194152.

FIGURE 3: FLAK TOWER NEAR THE BERLIN ZOO[103]

Two other priorities emerged from the decision in 1943 to focus on defending Germany against Allied bombing. One was increased investment in anti-aircraft defenses. As early as 1940 the Germans had begun building anti-aircraft batteries around Berlin in response to British bombing, and by mid-1943 there were 89 flak batteries defending the city against Allied air attack.[104] Production of the famous 88-mm anti-aircraft/anti-tank cannon began accelerating in 1942, and during 1944 the Germans produced more than five times as many as they had in 1940.[105] As Allied bombing intensified in 1943 with growing contributions of the Eighth Air Force in the daytime, the Germans began building concrete emplacements for their flak guns. Figure 3 shows the massive flak tower built near the Berlin Zoo. Such emplacements consumed large quantities of both concrete and manpower. By November 1943 German flak forces contained 13,500 heavy guns, 37,500 light guns, and over 1,360,000 personnel throughout the German-held territory, including Luftwaffe field and SS (*Schutzstaffel*) divisions as well as naval flak.[106] How much of Germany's war effort was

103 "German Flak Tower from Above," Time-Life photo, available at https://militaryberlin.wordpress.com/2011/11/18/berlin-flak-tower-from-above/. It took the British Army three attempts to demolish this tower after the war.

104 Murray and Millett, *A War to Be Won*, p. 332.

105 USSBS, *The Effects of Strategic Bombing on the German War Economy*, Table 114, p. 285.

106 O'Brien, *How the War Was Won*, pp. 305–306.

eventually devoted to air defense? Galbraith's deputy in the USSBS's Overall Economics Effects Division, Burton Kline, later offered this assessment:

> From 1942 to the first half of 1944 expenditures on air defense armaments—defensive fighter planes and their armament, antiaircraft weapons, and ammunition—nearly tripled, and at the time of the [Normandy] invasion amounted to about one third of Germany's entire munitions output. Indeed, in mid-1944 production of air defense armaments was at a higher level than was munitions output as a whole at the time Germany went to war with Russia. It can be seen, therefore, that where the preinvasion attacks really paid off was not nearly so much in the damage they did, but rather in the effect they had on causing the Germans to put a very significant part of their total war effort into air defense.[107]

The CBO, then, diverted substantial German resources and manpower into defending the Reich against Allied strategic bombing.

The second priority that emerged in 1943 from the growing weight of Anglo–American bombing was building the V-1 flying bomb and the A-4 (V-2) rocket as a means of revenge—bombarding England to terrorize the British. Adolf Hitler's hope was that these new weapons would provide the retaliatory reprisal he believed the German people demanded and that he promised to deliver.[108] Toward this end, the V-2 program, which had begun in 1934, became the "single highest-priority weapons program for the Nazi state" in the summer of 1943.[109]

The USSBS concluded that the Germans produced 30,000–32,000 V-1s and approximately 6,000 V-2s.[110] The first V-1s were launched at London from France on the night of June 12–13, 1944. Through March 1945 an estimated 8,200 were fired against England and another 7,800 at targets on the Continent (chiefly Antwerp).[111] V-2 attacks on England began the first week of September 1944. By the war's end, an estimated 1,115 V-2s had reached England, and another 1,675 had been fired against Continental targets (again mostly Antwerp).[112] The accuracy of the V-2s was so poor that they could only be used against targets the size of a major city such as London.

The V-1 was relatively inexpensive to produce. Kurt Tank, who was involved in developing the V-1's pulsejet engine, estimated that the V-1 required only 300 man-hours to produce compared to 3,700 for the Fw-190 fighter.[113] The V-2 was another story. It was the most expensive weapon system the Nazis ever undertook and probably cost as much as the U.S. atomic

107 Burton H. Klein, *Germany's Economic Preparations for War* (Cambridge, MA: Harvard University Press, 1959), pp. 232–233. This book was based on Klein's Ph.D. dissertation at Harvard.

108 Murray and Millett, *A War to Be Won*, pp. 315–316.

109 O'Brien, *How the War Was Won*, p. 340.

110 USSBS, *V-Weapons (Crossbow) Campaign* (Washington, DC: U.S. GPO, 1947), p. 4. However, the director of the main V-2 assembly plant claimed that 7,000 V-2s were produced. O'Brien, *How the War Was Won*, p. 341.

111 USSBS, *V-Weapons (Crossbow) Campaign*, p. 15.

112 Ibid., p. 16.

113 O'Brien, *How the War Was Won*, pp. 341, 342.

bomb program cost the United States.[114] Albert Speer's estimate was that one V-2 took 22,200 to 25,900 man-hours to construct. The USSBS estimated that the V-1 and V-2 programs together cost the Germans the equivalent of 24,000 fighters.[115] But this total appears to be a substantial underestimate even ignoring the costs of development, launch-site construction, and moving V-1 and V-2 production facilities underground. Using V-1 and V-2 production quantities together with V-1, V-2, and Fw-190 man-hour production estimates, the V-weapons cost the equivalent of some 38,000 to 44,000 Fw-190s.[116] To put these totals in perspective, from April 1943, when the CBO plan was finalized, to December 1944, the Germans produced around 38,200 fighters. Plainly the opportunity cost of the V-1 and V-2 programs was immense. And like the redeployment of German fighters to defend the Reich against the CBO and the accompanying investment in flak, the V-1 and V-2 programs were a direct response to the Allied strategic bombing.

Even worse for the Germans, in terms of affecting the course and outcome of the war, the V-1 and V-2 programs accomplished little. Their aim was to break the British will to continue fighting, but they no more did this than Allied bombing broke German morale. Further, the direct damage they inflicted was minimal by the standards of the Second World War. The V-1 was responsible for 5,864 deaths in the United Kingdom (UK) with another 17,200 people seriously injured; 2,863 people were killed by V-2s in the UK and another 6,286 seriously injured.[117] In the case of the V-2, measured against its return on investment it "was undoubtedly the most cost-ineffective weapon of the war."[118]

Although the V-weapons diverted substantial German technological and industrial resources into ineffective investments, they also caused a considerable diversion of the Allies' air campaign into attacking targets associated with the V-1 and V-2 programs. On June 29, 1943, an expanded British War Cabinet met to discuss intelligence evidence that the Germans were developing long-range bombardment weapons.[119] Most of those present concluded that they were developing such weapons, and on the evening of August 17, Bomber Command dispatched 596 Lancasters (of which 571 reached the target) to attack the research facilities at Peenemünde.[120] This attack prompted the Germans to begin moving the V-2 production facilities underground with the attendant delays and cost increases. It was also the beginning of the Crossbow campaign that eventually absorbed nearly 69,000 attack sorties and dropped more than 120,000 tons of bombs—a total that is over 12 percent of the approximately 971,580

114 Ibid., p. 340. Albert Speer's guess was that one V-2 took 22,200 to 25,900 man-hours to construct. Ibid., p. 341.

115 USSBS, *V-Weapons (Crossbow) Campaign*, p. 35.

116 The calculation of the high estimate is as follows: 32,000 V-1s times 300 man-hours = 9,600,000 man-hours; 6,000 V-2s times 25,900 man-hours = 155,400,000 man-hours. Adding these totals together and dividing by the 3,700 man-hours required for a single Fw-190 yields 44,595 Fw-190 equivalents.

117 O'Brien, *How the War Was Won*, p. 348.

118 Murray and Millett, *A War to Be Won*, p. 333.

119 O'Brien, *How the War Was Won*, p. 342.

120 Davis, *Bombing the European Axis Powers*, pp. 157–158.

tons of bombs the U.S. Army Air Forces delivered in the ETO during 1942–1945.[121] In addition, approximately 4,000 reconnaissance sorties were flown to locate launch sites and other elements of the V-weapons program.

The diversion of German resources into active defenses against Anglo–American bombers and the unproductive investment in V-weapons are two additional criteria for assessing the CBO that go well beyond the quantities of weapons and munitions that came out German factory doors in 1943 and 1944. A further (and related) consideration is that German efforts to shield the Reich against strategic bombing ultimately failed. From a strategic perspective, the most important failure was the Luftwaffe's loss of daylight air superiority on the Western Front by the spring of 1944.

In early October 1943, the U.S. Eighth Air Force had mounted a series of attacks against cities and facilities in Germany. The target for the last of these attacks, on October 14, 1943, consisted of the ball bearing plants at Schweinfurt. The Eighth had first attacked Schweinfurt and the Bf-109 assembly plants at Regensburg on August 17, 1943, a mission that cost the Americans 60 bombers. The premise of both Schweinfurt missions was that American bombers operating in tight formations beyond the range of U.S. escort fighters had the defensive firepower to hold their own against German fighters. On the second Schweinfurt raid, the Eighth Air Force dispatched 320 B-17s and B-24s, of which 60 (18.8 percent) were lost, and another 145 bombers suffered battle damage (including seven beyond repair).[122] Although this fact was not discovered until after the war, German losses that day were light: 31 fighters destroyed, 12 written off, and 34 damaged.[123] "Given the heavy losses sustained in the second Schweinfurt raid, the Eighth threw in the towel and accepted, for the time being, that it could not launch unescorted bombers deep into Germany. For the rest of the year, it confined itself to raids in France, the Ruhr, and the German coast—all within the range of Allied fighter escort."[124]

By December 1943 the Eighth Air Force began receiving P-51B fighters powered by the Rolls-Royce Merlin 61 engine. With two 75-gallon drop tanks, the P-51B's combat radius was more than 650 miles, and with a pair of 108-gallon drop tanks, their potential radius increased to

121 USSBS, *V-Weapons (Crossbow) Campaign*, p. 17; Office of Statistical Control, *United States Air Force Statistical Digest World War II*, Table 143, p. 243.

122 Roger A. Freeman, with Alan Crouchman and Vic Maslen, *Mighty Eighth War Diary* (New York: Jane's, 1981), p.126. Of the 320 bombers dispatched, 29 were a diversionary mission, and 229 attacked the ball bearing plants.

123 Davis, *Bombing the European Axis Powers*, pp. 182–183. Eighth Air Force bomber crews claimed to have shot down 186 German fighters and damaged nearly 150 others during this mission. Freeman, *Mighty Eighth War Diary*, p.126. It was not until the second volume of *The Army Air Forces in World War II* was nearing completion that its authors became aware of how exaggerated the "kill" claims of American and British bomber crews had been. Wesley Frank Craven and James Lea Cate, eds., *The Army Air Forces in World War II*, Vol. 2 (Chicago: University of Chicago, 1949), pp. xii–xiii.

124 Davis, *Bombing the European Axis Powers*, p. 184. On the missions of October 4, 8, 9, 10, and 14, 1943, the Eighth Air Force lost 165 bombers and their crews. While this was not the American perception at the time, the official history concluded that, "The Eighth Air Force had for the time being lost air superiority over Germany." Arthur B. Ferguson, "Pointblank" in Craven and Cate, *The Army Air Forces in World War II*, Vol. 2, p. 705.

850 miles.[125] The P-51 enabled the Eighth Air Force to escort its bombers on daylight raids to and from targets throughout Germany. The other significant change in early 1944 was a revised doctrine for employing escort fighters. Up until January 1944, the Eighth's fighters had been tied closely to the bomber streams and had not been permitted to desert the bombers to pursue German fighters. But when Lieutenant General James H. Doolittle became commander of the Eighth Air Force on January 6, 1944, he promptly promulgated a doctrine of "ultimate pursuit" of German fighters. Escort fighters were permitted to pursue German fighters until they were destroyed, whether in the air or on the ground, and fighter pilots were encouraged to strafe German airfields, transportation, and other ground targets while returning to base.[126]

With the invasion of the Continent (Operation OVERLORD) planned for mid-1944, American air commanders set out to achieve daylight air superiority over the Luftwaffe before D-Day. From the outset, Allied planners had recognized that OVERLORD had little chance of success unless they could control the skies over the Normandy beaches.[127] As of January 1, 1944, General Carl Spaatz assumed command of U.S. Strategic Air Forces in Europe (USSTAF), which included the Eighth Air Force under Doolittle and the Fifteenth under Lieutenant General Ira Eaker.[128] Spaatz and Doolittle's approach to achieving adequate air control by D-Day was one of annihilation. Bomber targets would be chosen to force the Luftwaffe's day fighters to do battle, thereby taking advantage of the German daylight fighter force's greatest vulnerability—its inability to man its planes adequately with experienced pilots.[129]

This campaign of annihilation began in mid-January 1944 when the Eighth Air Force resumed attacking targets deep in Germany. The first targets were the massive I. G. Farben chemical complex and the aircraft plants at Oschersleben and Halberstadt.[130] Although bad weather and losses had kept the Eighth out of Germany until the end of January, February 20–25 saw the Eighth and Fifteenth Air Forces mount six major bombing missions (known collectively as the Big Week) against Germany's aircraft industry. Both sides suffered heavy losses. But whereas the Americans could afford the attrition, the Germans could not.[131]

In the daytime, German Bf-109s and Fw-190s were increasingly forced to fight their way through Allied escort fighters to get at the American bomber formations. At the same time, the

125 Major General W. E. Kepner, *Eighth Air Force Tactical Development, August 1942–May 1945* (England: Eighth Air Force and Army Air Forces Evaluation Board, July 9, 1945), p. 50. The potential radii assume the fighters were not forced to jettison their drop tanks before they were dry.

126 Ibid., pp. 50, 56. To incentivize the fighter pilots, the Eighth Air Force began giving kill credits for German fighters destroyed on the ground as well as in the air.

127 Richard P. Hallion, *D-Day 1944: Air Power Over the Normandy Beaches and Beyond* (Washington, DC: Air Force History and Museums Program, 1994), p. 1.

128 Arthur B. Ferguson and Albert F. Simpson, "Final Reorganization," in Craven and Cate, *The Army Air Forces in World War II*, Vol. 2, pp. 744–745, 749, 754.

129 Davis, *Bombing the European Axis Powers*, p. 268.

130 Ibid., pp. 269–270.

131 Ibid., p. 277.

doctrine of "ultimate pursuit" gave German fighters little respite even at their own airfields. As Adolph Galland, who commanded the German daylight fighter force in 1944, later wrote:

> Wherever our fighters appeared, the Americans hurled themselves at them. They went over to low-level attacks on our airfields. Nowhere were we safe from them, and we had to skulk on our own bases. During take-off, assembling, climb and approach to the bombers, when we were in contact with them, on our way back, during landing, and even after that the American fighters attacked with overwhelming superiority.[132]

These relentless attacks were especially devastating to the Luftwaffe's capacity to provide adequate training for replacement pilots. By the second half of 1943, British and American fighter pilots were generally receiving twice as many flying hours (over 300) as their German counterparts prior to their first combat missions.[133] Inadequate training was particularly problematic when new Luftwaffe pilots had to rely on their instruments to climb through low overcasts to reach the bomber streams at higher altitudes in clearer air.[134] In May 1944, when the Eighth Air Force began targeting Germany's synthetic oil industry, severe shortages in aviation gasoline resulted "in catastrophic curtailment of training programs and operations," and by July, "Hundreds of newly assembled fighters were grounded by the lack of fuel."[135] The wastage inflicted on the aircraft the Germans did manage to produce was also substantial. In mid-1944, for example, a Luftwaffe representative on the General Staff reported that all but 92 of 1,000 planes produced at one aircraft factory were destroyed on the ground after acceptance.[136] From 1942 to 1943 the total number of Luftwaffe aircraft lost or damaged during non-operational activities doubled to over 7,000.[137] And by late 1944, when German aircraft production peaked, at least one-fifth of German aircraft production was being lost in transit flights from production facilities to Luftwaffe bases.[138]

Another consequential result of the Big Week was the German decision to undertake an immediate, large-scale dispersal of their aircraft industry. The 29 major aircraft producers were divided into 85 airframe factories, and aero-engine production was scattered to 249 sites.[139] This dispersal program eventually made the aircraft industry relatively resistant to air attack—at least so long as Germany's rail transportation system remained intact. But it caused more production delays, increased indirect labor costs by 20 percent, reduced the quality of many of the completed aircraft, and robbed the German aircraft industry of economies of scale.

---

132 Adolph Galland, *The First and the Last: The German Fighter Force in World War II* (Mesa, AZ: Champlin Fighter Museum Press, 1986), p. 276. This is a reprint of the 1955 British edition.

133 O'Brien, *How the War Was Won*, p. 294. The RAF and USAAF required at least 300 hours before operations; by July 1944 new German pilots were being sent into combat with as little as 100 hours. Ibid., p. 336.

134 Ibid., pp. 294–295.

135 Davis, *Bombing the European Axis Powers*, p. 287.

136 USSBS, *The Effects of Strategic Bombing on the German War Economy*, p. 159.

137 O'Brien, *How the War Was Won*, Figure 40, p. 296.

138 USSBS, *The Effects of Strategic Bombing on the German War Economy*, p. 159.

139 Davis, *Bombing the European Axis Powers*, pp. 287–288; O'Brien, *How the War Was Won*, Map 6, p. xxvi.

In sum, by April 1944 the CBO had brought about the effective destruction of the Luftwaffe not by running the Germans out of fighters but by running them out of experienced pilots.

> Between January and June 1944—the five months before D-Day—2,262 German fighter pilots died . . . In May alone, no less than 25 percent of Germany's total fighter pilot force (which averaged 2,283 at any one time during this period) perished. During Big Week, American air forces targeted the German aircraft industry for special treatment; while production continued, the fighter force took staggering losses. In March 1944, fully 56 percent of the available German fighters were lost, dipping to 43 percent in April (as the bomber effort switched to Germany's petroleum production), and rising again to just over 50 percent in May, on the eve of Normandy.[140]

So complete was Allied daylight air superiority by D-Day that on June 6, Allied air forces mounted over 14,700 sorties in support of the Normandy landings, whereas the Luftwaffe was barely able to generate 100 sorties over northern France.[141] Granted, the U.S. Eighth Air Force in England paid a heavy price for this victory, losing 1,012 heavy bombers to Luftwaffe fighters, another 447 heavies to flak, and 933 fighters to all causes during January–May 1944.[142] Nevertheless, Allied air superiority was a *sine qua non* for the success of the Normandy landings.[143] Without the attainment of daylight air superiority, it is difficult to envision how OVERLORD could have possibly succeeded. Control of the air precluded the Germans from quickly bringing Panzer units to Normandy to throw the Allies back into the sea as the Germans had planned. For example, the Panzer Lehr Division, which was reckoned to have four times the fighting strength of standard Panzer divisions prior to June 6, was ordered to concentrate near Caen on D-Day; the division not only suffered attrition from Allied air attacks during its deployment, but its advance was slowed to between 6 and 8 miles per hour, and much of the division was not in action until July 10.[144]

In light of this brief review of the CBO, the judgment of Galbraith and other critics that the strategic bombing of Germany was a failure appears to have overlooked much, including the collapse of the rail and inland waterways systems that transported the coal on which the entire

140 Hallion, *D-Day 1944: Air Power Over the Normandy Beaches and Beyond*, p. 2.

141 Ryan Crane, "Allied Air Forces Paved Way for D-Day," *Air Force News Service*, May 30, 2014, available at http://www.af.mil/News/ArticleDisplay/tabid/223/Article/485137/allied-air-forces-paved-way-for-d-day.aspx.

142 Office of Statistical Control, *Army Air Forces Statistical Digest: World War II,* Table 159, p. 255.

143 Another sine qua non for OVERLORD's success was the Allied deception plan. Operations FORTITUDE NORTH and FORTITUDE SOUTH misled the German high command as to OVERLORD's intended landing sites and tied down German forces that could very well have stopped the Allies on the Normandy beaches. Roberts, *The Storm of War*, p. 463. Ultra enabled the Allies to monitor the success of their deception plan in real time. Hodges, *Alan Turning: The Enigma*, p. 362.

144 O'Brien, *How the War Was Won*, pp. 316–317. Seven weeks later, Panzer Lehr was completely destroyed by aerial bombardment.

Germany economy depended for power.[145] Galbraith's assessment was based primarily on a single criterion: how long it took the Anglo–American bombing campaign to bring about declines in German war production—declines that only came very late in the war. But there are other relevant decision criteria. These include 1) the diversion of German resources into the wasteful V-weapons programs; 2) the redeployment of the Luftwaffe fighter forces to Germany and the Western Front, thereby stripping other fronts of air support; 3) the diversion of a third of Germany's war effort into air defenses against Allied strategic bombing; 4) the dispersal of Germany's aircraft industry; and 5) the CBO's attainment of air superiority prior to D-Day by running the Germans out of adequately trained pilots. Hence Phillips O'Brien created his tongue-in-cheek but appropriate description of the CBO as "the war-winning failure."[146]

Of course, O'Brien's argument about how and why the Allies won World War II goes well beyond suggesting a fresh look at the CBO. His point of departure was the observation that the United States, Britain, Japan, and even Germany all allocated at least two-thirds of their annual wartime production to air and sea weapons.[147] Land armaments, by comparison, "were only a small part of munitions output for Germany and Japan—and the USA and UK as well."[148] O'Brien, therefore, argues that the dominant, decisive campaign of World War II was the contest between the two sides' production and employment of air and sea forces in a "super battlefield" that covered thousands of miles in length and breadth.[149] Histories of the Second World War, he insists, have tended to focus on the conflict's great land battles—the second Battle of El Alamein, Midway, Stalingrad, Kursk, and the Battle of the Bulge. Due to the fact that, "Out of every five Germans killed in [land] combat—that is, on the battlefield rather than in aerial bombing or through other means—four died on the Eastern Front," Andrew Roberts has asserted that the war against Germany was won on the Eastern Front.[150] O'Brien disagrees:

> While the overwhelming consensus of historians is that Germany was defeated primarily through the interaction of the large land armies on the Eastern Front, this is true only if one believes that the number of soldiers deployed was the best indicator of national effort. On the other hand, if economic, technological, and overall domestic allocation of resources is a better measurement, the air-sea war between the Germans and the British and Americans was the defining campaign of the war—by a considerable measure. This contest was by far the more modern of the two, involving the most advanced, expensive machinery, in the largest quantities and controlled

145 The Allies' "transportation offensive severely weakened the German economy from October 1944 and induced its collapse in early January 1945." Alfred C. Mierzejewski, *The Collapse of the German War Economy, 1944–1945: Allied Air Power and the German National Railway* (Chapel Hill and London: University of North Carolina Press, 1988), pp. 184, 198. For a recent (2015) review of the positions various authors have taken on the efficacy (or inefficacy) of the CBO, see O'Brien, *How the War Was Won*, pp. 6–13.

146 O'Brien, *How the War Was Won*, p. 268.

147 Ibid., p. 3.

148 Ibid., p. 2.

149 Ibid., p. 5.

150 Roberts, *The Storm of War*, pp. 520, 603.

> by the best-trained warriors. It cost the most, destroyed the most and determined strategy for the last three years of the war. That the Eastern Front was responsible for the greater human suffering is undeniable. But this human suffering should not obscure the fact that it was a secondary theater in terms of production and technology, where force was actually used in a far more limited front-based manner which caused considerably fewer choices to be made and smaller amounts of equipment to be utilized and destroyed.[151]

For present purposes we only need insist upon a single point: that the selection and interpretation of appropriate analytic measures are central to the judgments one reaches about the how and why of historical outcomes in wartime. As we will see next, appropriate analytic measures are also central to decisions about future force structures.

Designing U.S. Strategic Forces in the Late 1940s

In December 1945 the U.S. Army Air Forces established Project RAND (an acronym for "Research ANd Development"). RAND was initially housed within the Douglas Aircraft Company and drew on its personnel. General "Hap" Arnold's motivation for creating Project RAND was to place the AAF's "postwar and next-war research and development programs . . . on a sound and continuing basis."[152] During World War II the close cooperation between the military, government agencies, universities, and industry achieved by Vannevar Bush's wartime Office of Scientific Research and Development (OSRD) had produced a number of weapons and systems "unknown at the outbreak of hostilities"; the postwar view was that several of these innovations had "decisively" affected the course and outcome of the war.[153] Among the more important were the centimeter-wavelength search radars used in the Battle of the Atlantic; proximity fuses; computerized fire direction systems employing microwave radars developed by the Massachusetts Institute of Technology's Radiation Laboratory; and, of course, the atomic bomb. General Arnold was convinced that "a nucleus of scientists working full time and continuously on the problems of the military" was not only needed but also one of the most important challenges facing the AAF in the immediate aftermath of World War II.[154] By establishing RAND, Arnold hoped to ensure that the AAF would stay on the forefront of scientific developments in weaponry and related military capabilities.

At the outset, RAND's official purpose was to recommend to the AAF, and the U.S. Air Force (USAF) starting in 1947, the "preferred instrumentalities" and techniques for conducting

151 O'Brien, *How the War Was Won*, pp. 484–485.

152 General H. H. Arnold, "Memorandum for Dr. von Kármán," November 7, 1944, in Theodore von Kármán, *Toward New Horizons: Science, the Key to Air Supremacy* (Washington, DC: Army Air Forces Scientific Advisory Group, 1945), p. iv. Project RAND was initially established within the Douglas Aircraft Company in 1946. Others who were directly involved in creating RAND included Edward Bowles, generals Lauris Norstad and Curtis LeMay, Donald Douglas, Arthur Raymond, and Franklin Collbohm.

153 Irvin Stewart, *Organizing Scientific Research for War: The Administrative History of the Office of Scientific Research and Development* (Boston: Little, Brown and Company, 1948), p. ix.

154 C. J. Hitch, amended by J. R. Goldstein, *RAND: Its History, Organization and Character*, B-200 (Santa Monica, CA: The RAND Corporation, July 20, 1960), p. 2.

intercontinental air warfare.[155] In the early years of the U.S.–Soviet Cold War, there were two areas of immediate concern to the fledgling Air Force: (1) active air defense of the continental United States against a Soviet atomic attack; and (2) offensive bombing systems to attack the Soviet homeland. In the late 1940s, RAND undertook major studies to provide its Air Force customer with the think tank's best scientific advice regarding both the defensive and offensive dimensions of intercontinental air warfare.

These two studies occurred during a period of great optimism about the potential of mathematical and scientific methods to provide solutions to the problems of modern warfare. Among those sharing this optimism was Warren Weaver. During the war, Weaver had been chief of the OSRD's Applied Mathematics Panel. His efforts there had "focused increasingly on constructing a general theory of military worth," and in January 1946 he published a classified essay that detailed his thoughts on "a mathematically-based science of warfare."[156] Seeing the close fit between Weaver's ambitious goal and RAND's objectives, in 1946 Frank Collbohm tried to recruit Weaver to serve as Project RAND's director.[157] Weaver, however, preferred to return to his pre-war position at the Rockefeller Foundation and recommended Collbohm instead hire John D. Williams,[158] who had led the Applied Mathematics Panel's Statistical Research Group and shared Weaver's view that mathematics could "provide the foundation for a science of warfare that incorporated contributions from a wide range of academic disciplines."[159] Collbohm took Weaver's recommendation and hired Williams to head Project RAND's "Evaluation of Military Worth Section."[160] Williams remained with the evaluation group until 1948 when it was divided into three departments—mathematics, economics, and social science—and Williams took over the new mathematics department.[161] Early 1948 was also the juncture at which discomfort between the think tank's mission as an objective advisor

155 A. A. Alchian, G. D. Bodehorn, S. Enke, C. J. Hitch, J. Hirshleifer, and A. W. Marshall, *What Is the Best System?* D-860 (Santa Monica, CA: The RAND Corporation, January 4, 1951), p. 1.

156 David R. Jardini, *Thinking Through the Cold War: RAND, National Security and Domestic Policy, 1945–1975* (Los Gatos, CA: Smashwords, 2013), pp. 23–24, available at https://www.smashwords.com/books/view/352765.

157 Ibid., p. 23.

158 According to Andrew May, Collbohm also tried to recruit the physicist Louis Ridenour to run Project RAND, but, like Weaver, Ridenour preferred to remain a RAND consultant rather than join the think tank's staff. Andrew D. May, "The RAND Corporation and the Dynamics of American Strategic Thought, 1946–1962," unpublished Ph.D. dissertation, Emory University, Department of History, Chapter 2, p. 2. Each chapter in the electronic files of May's dissertation starts with page 1.

159 Jardini, *Thinking Through the Cold War*, p. 24.

160 Weaver understood military worth as efficiency (if not optimization): producing "the largest margin of profit—the largest excess of return over cost" relative to the force employed and the results achieved. May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 9.

161 Jardini, *Thinking Through the Cold War*, p. 46. The other departments at this time were aircraft, missiles, nuclear physics, and electronics.

to the Air Force and the product-oriented culture of Douglas Aircraft led to the establishment of RAND as a freestanding, private, nonprofit corporation in the state of California.[162]

The reason for highlighting Collbohm and Williams in the context of Weaver's interest in military worth lay in their shared conviction that the problems confronting RAND researchers could not be satisfactorily addressed by "hard" disciplines such as aeronautical engineering and the physical sciences. In the early days, the majority of those Project RAND hired were engineers and mathematicians recruited from the aircraft industry, many of whom had experience with wartime operations research (OR).[163] But as Philip Morse and George Kimball noted in their classic account of OR, "measures of value are rather unfamiliar to the physical scientist."[164] Recognizing from the outset that advising the Air Force on the preferred instrumentalities and techniques for conducting intercontinental air warfare, Collbohm convened a conference of social scientists in New York City in September 1947 to "judge the nature and extent of the contribution which the human sciences" could make to RAND's problem-driven research.[165] The issue of measuring military worth came up repeatedly during the conference. As Weaver stated one point, "There is a considerable group in the home office staff of RAND who are concerned with attempts to see whether or not a useful concept of military worth can be constructed and given quantitative value."[166]

The conclusion Collbohm drew from the September 1947 conference was that social scientists could help RAND deal with issues such as military worth. In 1948 he established two new departments that formally enlisted the social sciences. Williams recruited Charles Hitch, whom he had known at Oxford, to run Project RAND's economics division; Hans Speier was hired to head the social sciences division.[167] Subsequently, Collbohm, Williams, and Hitch consistently encouraged interdisciplinary research and solutions to RAND's problems.[168] Indeed, when RAND's management decided to construct a new building in Santa Monica, Williams argued for a lattice-like set of enclosed patios to encourage chance meetings among

162 David Hounshell, "The Cold War, RAND, and the Generation of Knowledge, 1946–1962," *Historical Studies in the Physical and Biological Sciences* 27, no. 2, 1997, p. 242. With Project RAND's incorporation Collbohm became its president, a position he held until 1967.

163 From 1948 to 1962, RAND grew from 225 employees and an annual budget of $3.5 million (some $46 million in fiscal year 2015 constant dollars) to 1,100 employees and a budget of over $20 million (over $146 million in FY 2015 constant dollars). From 1951 to 1963 the proportion of researchers fell from 51 percent to 39 percent, while RAND's proportion of administrators grew from 49 to 61 percent. Augier, March, and Marshall, "The Flaring of Intellectual Outliers," p. 14.

164 Philip M. Morse and George E. Kimball, Operations Evaluation Group, Office of the CNO, *Methods of Operations Research*, OEG Report No. 54 (Washington, DC: Navy Department, 1946), p.8.

165 Project RAND, *Conference of Social Sciences: September 14 to 19, 1947—New York*, R-106 (Santa Monica, CA: The RAND Corporation, June 9, 1948), p. vii.

166 Ibid., p. 281.

167 C. J. Hitch, amended by J. R. Goldstein, *RAND: Its History, Organization and Character*, p. 8.

168 Augier, March, and Marshall, "The Flaring of Intellectual Outliers," p. 10.

researchers from different academic disciplines.[169] Thus, both Project RAND's initial air defense and bombing systems studies took place in the context of a growing commitment to interdisciplinary research by Collbohm, Williams, Hitch, and others.

Together these two studies reinforce the importance of choosing good decision criteria. The authors of the 1948 active air defense study made sensible choices of analytic criteria to structure their analysis—choices that illuminated the nature and complexity of the air defense problem. In the case of the strategic bombing systems study completed in 1950, the analysis went in the opposite direction. The desideratum of scientifically quantifying the military worth of alternative combinations of atomic bombs and bombers by maximizing the ratio of damage inflicted on Soviet targets to various measures of cost led to recommendations that the Air Force rightly rejected.

In December 1946 General Carl Spaatz, who had succeeded General Arnold as chief of the Army Air Forces (AAF), asked Theodore von Kármán, chair of the AAF's Scientific Advisory Group, to propose a detailed study of active (as opposed to passive) air defense and suggested that Project RAND be given the assignment.[170] The questions the Air Force posed to RAND went beyond the technical capabilities of the weapons that might be employed to counter an air attack on the continental United States or its overseas bases. The Air Force also asked the study to emphasize the economic feasibility of any recommended air defense program.[171]

James Lipp, an aeronautical engineer who had worked for Douglas Aircraft from 1935 until he joined Project RAND after the war, directed the study.[172] As Andrew May has observed, given the academic backgrounds of Lipp and others involved in the project—including the physicist Louis Ridenour—one might have expected a narrow, engineering examination of how active air defenses might be designed, organized, and deployed to defend the United States against a surprise attack by Soviet bombers armed with atomic weapons; instead, Lipp's group went in the opposite direction and produced "a broad-gauged study of how active defense would fit into America's postwar national security strategy."[173]

Based on preliminary papers by Lipp's group written during the first half of 1947, RAND issued a summary report, *Active Defense of the United States against Air Attack* (RA-15058),

---

169 J. D. Williams, "Comments on RAND Building Program," letter to RAND staff, December 26, 1950, available at http://www.rand.org/pubs/classics/building.html.

170 Shelters, dispersal, and camouflage typify passive air defenses. By the 1950s U.S. active air defenses included interceptor aircraft, surface-to-air missiles, radars, and the semi-automatic ground environment (SAGE) consisting of computers and networking equipment.

171 Project RAND, "Active Defense of the United States against Air Attack (A Preliminary Study of the Problem)," RAND Research Memorandum (RM) 28, February 5, 1948, p. 1.

172 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 2.

173 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 3. Lipp, for example, had been involved in analyzing the structural weight issues in the design of an earth-circling satellite for Project RAND's first report. He also headed the think tank's missile division.

on July 10, 1947.[174] In response to comments from both civilian and military readers, the original report was revised and re-issued as Research Memorandum 28, or RM-28, in February 1948. On the whole, RM-28 was not very optimistic about the prospects for active air defense, stating at one point that "fully effective interception" against a surprise, all-out Soviet atomic attack in the 1952–1960 timeframe would be "difficult or even impossible."[175]

This conclusion was not without precedent. Earlier, in 1946, Ridenour had written an article in *One World or None* arguing that there was no prospect of the United States defending against atomic air attack due to the sheer destructiveness of atomic weapons, the speed and trajectories of ballistic missiles like the V-2, and the likelihood that the next major war would begin with a "Pearl Harbor" kind of attack.[176] Since Ridenour, in his capacity as a RAND consultant, also contributed one of the preliminary papers done for the air defense study, it is hardly surprising that his bleak assessment carried over into RM-28.[177]

The decision criteria that structured RM-28 were implied in a series of assumptions that took up the first half of this short report (comprising only ten pages). First and foremost was the assumption that "a central purpose" of U.S. policy "must be to prevent major war" on the grounds that atomic weapons, if employed on a sufficient scale, could "render large parts of the globe literally uninhabitable."[178] Next, RM-28 assumed that Russia and its satellites constituted the only enemy with the war potential to develop the weaponry and forces to mount an atomic attack on the United States and that ideological conflict with the Union of Soviet Socialist Republics (USSR) was already underway.[179] True, Russia had not yet achieved a capability for atomic attack. Moscow's first atomic test (known as RDS-1 or First Lightning in Russia and as Joe-1 in the West) did not occur until August 1949.[180] Not until the mid-1950s did the Soviets begin fielding bombers with the range to reach the northeastern United States

---

174 Project RAND, RM-28, p. ii.

175 Ibid., p. 7.

176 Louis Ridenour, "There Is No Defense," in Dexter Masters and Katherine Way, eds., *One World or None: A Report to the Public on the Full Meaning of the Atomic Bomb* (New York: The New Press, 2007), pp. 90–106. Originally published by McGraw-Hill in 1946, Other contributors to this volume included Niels Bohr, Albert Einstein, General H. H. Arnold, J. Robert Oppenheimer, and Leo Szilard. The book advocated international control of atomic weapons. As Oppenheimer wrote, "The vastly increased powers of destruction that atomic weapons give us have brought with them a profound change in the balance between national and international interests. The common interest of all in the prevention of atomic warfare would seem immensely to overshadow any purely national interest, whether of welfare or of security." Ibid., p. 38.

177 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 3. Bernard Brodie had made the same point in 1946. Bernard Brodie, "Implications for Military Policy," in Frederick S. Dunn, Bernard Brodie, Arnold Wolfers, Percy E. Corbett, and William T. R. Fox, *The Absolute Weapon: Atomic Power and World Order* (New York: Harcourt, Brace & Company, 1946), p. 76.

178 Project RAND, RM-28, pp. 1–2.

179 Ibid., p. 3.

180 Thomas C. Reed and Danny B. Stillman, *The Nuclear Express: A Political History of the Bomb and Its Proliferation* (Minneapolis, MN: Zenith Press, 2009), pp. 29, 33–34. In mid-1948, the Central Intelligence Agency (CIA) "had estimated that the Soviet Union most probably would not have the bomb before mid-1953." David M. Kunsman and Douglas B. Lawson, *A Primer on U.S. Strategic Nuclear Policy* (Albuquerque, NM: Sandia National Laboratories, January 2001), p. 24.

from Soviet bases on two-way missions.[181] But in light of President Harry Truman's March 1947 declaration that, "The United States would provide political, military and economic assistance to all democratic nations under threat from external or internal authoritarian forces," the Cold War was unquestionably underway.[182] Finally, RM-28 assumed that American traditions and laws would not "permit the United States to strike a massive offensive blow without warning or explicit provocation."[183] In other words, RM-28 ruled out a preventative attack on the USSR before Moscow could field intercontinental nuclear forces. This argued that the Russians would almost certainly have the considerable advantage of striking a massive first blow.

The decision criteria implicit in these various assumptions appear to boil down to two. First, any advice about air defense against atomic air attack had to be couched within the evolving international security situation confronting the United States—along with likely U.S. political policies and military strategies for dealing with the evolving Cold War. The study's assumptions placed so much importance on taking these issues into account that RM-28 did not even attempt to propose instrumentalities and techniques for an economically feasible air defense system. Instead, the report concluded that nothing less than the structure of international relations must be considered in deciding whether a technically feasible and apparently desirable defensive system genuinely represented a safeguard required by, or even desirable to, the United States under the conditions then existing in the world.[184] Indeed, given the uncertainties about Soviet intentions and nuclear capabilities, RM-28 argued that funds lavished on a

181 The Tupolev Tu-4 bomber that the Soviets had reverse engineered from early versions of the B-29 only had a combat radius of ~820 nm (compared with 1,390–1,560 nm for the B-29), and on a round-trip profile, they could not reach any targets in the continental United States even from bases as far north as Anadyr in the upper Chukotsky Peninsula of eastern Siberia. Steven J. Zaloga, *The Kremlin's Nuclear Sword: The Rise and Fall of Russia's Strategic Nuclear Forces, 1945–2000* (Washington, DC: Smithsonian Institution Press, 2002), p. 15. Even launching from the Kola Peninsula on a one-way mission, the Tu-4 could not reach targets in the northeastern United States. Ibid., p. 16.

182 Office of the Historian, U.S. Department of State, "The Truman Doctrine, 1947," available at https://history.state.gov/milestones/1945-1952/truman-doctrine. In his February 1946 Long Telegram from the Moscow embassy, George Kennan assessed the USSR's rulers as "committed fanatically to the belief that with the U.S. there can be no permanent modus vivendi, that it is desirable and necessary that the internal harmony of our society be disrupted, our traditional way of life be destroyed, and the international authority of our state be broken, if Soviet power is to be secure." George Kennan telegram to George Marshall (the "Long Telegram"), February 22, 1946, p. 14, available at https://www.trumanlibrary.org/whistlestop/study_collections/coldwar/documents/pdf/6-6.pdf. The following month Winston Churchill famously observed that an "iron curtain" had descended across Central and Eastern Europe leaving Warsaw, Berlin, Prague, Vienna, Budapest, Belgrade, Bucharest, Sofia, and the populations around them in the "Soviet sphere." Winston Churchill, "The Sinews of Peace ('Iron Curtain Speech')," Westminster College, Fulton, Missouri, The Churchill Centre, March 5, 1946, available at http://www.winstonchurchill.org/resources/speeches/1946-1963-elder-statesman.

183 Project RAND, RM-28, p. 4.

184 Ibid., p. 8. Of course U.S. leaders did opt for investing heavily in continental air defenses, initially against manned bombers. By mid-1957 the Army Air Defense Command had deployed 58 Nike missile battalions, the Air Force's Air Defense Command had 69 all-weather interceptor squadrons, and the Distant Early Warning line of radars was semi-operational. Continental Air Defense Command, (CONAD), *Historical Summary July 1956–June 1957* (Headquarters CONAD, 1957), pp. 38, 46, 63, 69, available at http://www.northcom.mil/Portals/28/Documents/Supporting%20documents/%28U%29%201956-1957%20NORAD%20CONAD%20History.pdf.

"really first-class intelligence system" would buy more air defense for the continental United States than an equal quantity of money spent on any conceivable defense system.[185]

The other criterion underlying RM-28 was more technical. It focused on the fact first demonstrated at Hiroshima that a single atomic bomb could destroy a city. Given this level of sheer destructiveness, RM-28 drew the obvious conclusion that the offense would henceforth enjoy a tremendous advantage over the defense. If a single Soviet bomber getting through U.S. air defenses could destroy an American city, then the prospect for a fully effective defense against long-range bombers armed with atomic weapons appeared dim indeed. Even a defense that was 90 percent effective would entail catastrophic destruction for the defending nation.[186] And while RM-28 did not use the word *deterrence*, its authors suggested that America would have to place major reliance on its offensive atomic forces, which should be maintained in constant readiness and at full strength, to deter a major attack on the United States.[187]

Seven decades later, this author would not gainsay Andrew May's judgment that RM-28 was "a remarkable piece of analysis."[188] The report's explicit assumptions and implicit metrics framed the issue of active air defense in its broader political, economic, and military context. Without first examining that context, it made no sense to Lipp, Ridenour, and others involved in the study to pose any specific solution to the air defense problem. In this sense, they not only asked the right questions about the military's future role in active air defense but also selected appropriate measures for structuring the analysis, albeit implicitly.

The *Strategic Bombing Systems Study* (R-173), completed in March 1950, was less successful in its choice of decision criteria. Like the 1948 active air defense study, R-173 began in 1946. From the outset, R-173 had two objectives. For Weaver, Collbohm, Williams, and other RAND staff members who became involved in the analysis of bombing systems, the preeminent aim was methodological development: to exploit advances in applied mathematics and computation in order to "quantify and manipulate the thousands of variables associated with modern warfare," including those found in the social sciences.[189] The hope was to put "methodological meat on the conceptual skeleton" of the mathematically based science of warfare that Weaver had envisioned in his January 1946 classified essay "Comments on a General Theory

185 Project RAND, RM-28, p.8. As early as 1939, Russian physicists Yuliy B. Khariton and Yakov B. Zeldovich published three papers on the steps needed to achieve an atomic bomb, and in early 1941 they calculated the critical mass of a U-235 bomb would have needed to be about 22 pounds. Reed and Stillman, *The Nuclear Express*, pp. 26–27. During World War II they kept abreast of Allied progress toward an atomic bomb through spies such as Klaus Fuchs, whose efforts included providing the Russians with dimensional drawings of the U.S. implosion design dropped on Nagasaki (Fat Man). Ibid., pp. 29–30. And in August 1945 Josef Stalin put his security chief, Lavrenti Beria, in charge of a special committee with orders to build the bomb as quickly as possible.

186 Ridenour in Masters and Way, *One World or None*, p. 106.

187 Project RAND, RM-28, p. 7.

188 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 6. Jardini's less sympathetic assessment was that RM-28 was "mired in analytical complexity" just like the *Strategic Bombing Systems Study*. Jardini, *Thinking Through the Cold War*, p. 30.

189 Ibid., p. 24.

of Air Warfare."[190] By early 1947 this ambitious methodological development came to be called "systems analysis."[191]

The other goal of the early bombing systems study that led to R-173 was to provide the AAF with "immediately useful information on aircraft systems that would be available in the short run": the B-29, B-50, and B-36.[192] Under the direction of L. E. "Gene" Root, the study team initially concentrated on analyzing the interrelationships among such variables as the weight, payload, engine type, and wing design of existing bombers. As a result, the 1946 interim study was almost exclusively technical in nature. However, its authors made it clear that follow-on work would incorporate not only jet and rocket propulsion, pilotless aircraft, and long-range missiles, but also detail on the location and importance of Soviet targets and forces, as well as the vulnerabilities of U.S. bombers to enemy defenses.[193]

Incorporating all these additional complications meant, of course, that the subsequent bombing systems analysis would be of far greater complexity than the Interim Study. By mid-June 1947 Root's study group was calculating out to 1955 time-dependent probabilities for usable bases, reaching the target area despite Soviet defenses, target identification, successful weapon release and functioning, and bomber crew aiming error (dependent on training and the level of combat stress), as well as the probability of a hit for each mission.[194] The direction of such an ambitious analysis required an equally ambitious project leader. By mid-1947 RAND had recruited Edwin W. Paxson to take charge of the core analyses. Paxson had worked with both Warren Weaver and John Williams at the Applied Mathematics Panel during the war. There he had conducted operations research for the U.S. Eighth Air Force in England; after the war, he had also served as a consultant to the USSBS.[195] Paxson shared Weaver and Williams' conviction that a scientifically rigorous theory of warfare was within reach, and he set out to propel Project Air Force toward this end by mathematizing every component of a single atomic attack on the Soviet Union's military-industrial base.

The result was complexities on top of complexities. For example, Paxson recruited members of the Airborne Vehicles Section to develop equations interrelating shape, size, strength, weight, engines, and bomber performance to determine an optimum aircraft. But preliminary work indicated that the interdependencies of these variables would require the calculation of well over a hundred equations and relationships.[196] Similarly, the efforts of Paxson and his assistant Edward Quade to apply game theory to air combat duels between U.S. bombers and

190 Ibid., pp. 24–25.

191 Ibid., p. 26.

192 Ibid., pp. 25.

193 Ibid., pp. 25–26, 288.

194 Ibid., p. 28.

195 Ibid., p. 28.

196 Ibid., p. 29.

Soviet interceptors only highlighted the thirtyfold gap in interceptor effectiveness between the results of their theoretical model and the empirical results from World War II.[197]

As Paxson's study expanded, consuming ever more of Project Air Force's resources, so did the complications: "With each apparent solution researchers uncovered whole new levels of complexity and uncertainty."[198] There seemed to be no end in sight. This prompted Collbohm to use an Air Force request in June 1949 for RAND to recommend a preferred diameter for the TX-5 (later the Mark-5) atomic bomb to prod Paxson's team into producing more tangible results and meet the Air Force's January 1950 deadline for delivering the bombing systems analysis.[199] In response Paxson re-defined the problem as follows: "Given a fixed amount of fissile material and a fixed sum of money with which to procure, operate, and maintain a strategic striking force at strength for a four-year period, specify the atomic bombs and aircraft which will maximize damage of an initial bombing strike."[200] This refocusing of Paxson's bombing systems study sought to constrain mathematical complexity. Nevertheless, when the *Strategic Bombing Systems Analysis* (R-173) was finally completed in March 1950, it "contained no fewer than 400,000 different bomb-bomber combinations, and calculated the effects of dozens of different variables, many of which were interdependent."[201]

Like U.S. war plans through early 1949, R-173 appears to have assumed that a future war with the USSR would begin with Soviet conventional offensives in Europe, the Middle East, and the Far East. In Western Europe, U.S. forces would withdraw to the Rhine River, eventually holding at the Pyrenees Mountains. In the meantime, the TROJAN war plan the Joint Chiefs of Staff approved on January 28, 1949, envisioned using 133 atomic bombs to attack a broad range of industrial facilities in 70 Soviet cities.[202] At the time, Strategic Air Command (SAC) was rapidly acquiring the capability to execute an atomic strike of this magnitude. By mid-January 1950 SAC had more than 200 nuclear-capable B-29s, B-50s, and B-36s, and by the year's end, the U.S. stockpile had grown from 170 to 299 Mark-3 and Mark-4 atomic

197 Fred Kaplan, *The Wizards of Armageddon* (New York: Simon & Schuster, 1983), p. 88. Paxson and Quade figured that when the interceptor and bomber were in roughly the same geometric position, the interceptor would have a 60 percent kill probability. But in World War II the probability of a kill in this situation was only 2 percent.

198 Jardini, *Thinking Through the Cold War*, p. 30.

199 The TX-5 was a 92-point implosion device. It served as the primary for the first U.S. thermonuclear test in November 1952. Fielded as the Mark-5, this bomb was in service from 1952 to 1963. The Mark-5 had a diameter of 39 inches compared to Fat Man's 60-inch diameter. The four variants of the Mark-5 are reported to have yields of 6, 16, 55, 60, 100, and 120 kilotons.

200 Project RAND, *Staff Report Summary: Strategic Bombing Systems Analysis*, R-173S (Santa Monica, CA: The RAND Corporation, March 1, 1950), p. 1.

201 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 13; and Jardini, *Thinking Through the Cold War*, p. 32.

202 Kunsman and Lawson, *A Primer on U.S. Strategic Nuclear Policy*, p. 23.

bombs.[203] It seems plausible, then, to presume that R-173's single atomic strike was comparable to that envisioned in the TROJAN war plan. That said, however, the limited numbers of atomic weapons available and their allocation to the highest priority Soviet targets was by no means SAC's only problem in the late 1940s. The most recent targeting data available to U.S. planners consisted of German aerial photos from 1942–1943.[204] As a practical matter at this early stage in the Cold War, SAC aircrews could not even be certain of locating Russian cities in the vastness of the USSR, much less precise aim points within them, especially at night or in bad weather.[205]

R-173's fundamental goal was to provide a rigorous, quantitative basis for selecting the Air Force's next-generation nuclear bombs and bombers. Regarding the preferred weapons, the study assumed that as late as 1960 the U.S. nuclear stockpile would consist solely of atomic bombs with yields equivalent to thousands of tons of trinitrotoluene (TNT).[206] Most likely this decision was made before Truman's January 30, 1950, announcement that the United States would pursue thermonuclear (fusion) weapons. Given Collbohm's deadline and the complexity of R-173's analysis, the study was undoubtedly too far along by mid-1949 to consider hydrogen bombs with yields equivalent to one million to 25 million tons of TNT.[207] Nonetheless, the United States detonated its first thermonuclear device in November 1952, and the first production H-bomb, the Mark-15, entered active service with SAC in 1955.

As for the preferred bomber, R-173 based its recommendation on three explicit decision criteria: (1) the ratio of the maximum damage inflicted (without regard for losses) divided by a fixed cost; (2) the ratio of damage inflicted per pound of aircraft lost; and (3) the ratio of damage inflicted per aircrew (airplane) lost.[208] Maximizing these ratios led Paxson to recom-

203 David Alan Rosenberg, "U.S. Nuclear Stockpile, 1945 to 1950," *The Bulletin of the Atomic Scientists*, May 1983, p. 30; and Department of Defense and Department of Energy, "Summary of Declassified Nuclear Stockpile Information: Declassified Stockpile Data 1945 to 1994," *U.S. Department of Energy OpenNet*, June 1994, available at https://www.osti.gov/opennet/forms.jsp?formurl=document/press/pc26tab1.html. Both the Mark-3 and Mark-4 weighed over 10,000 pounds. Carey Sublette, "Complete List of All U.S. Nuclear Weapons," *Nuclear Weapon Archive*, October 14, 2006, at http://nuclearweaponarchive.org/Usa/Weapons/Allbombs.html.

204 David Alan Rosenberg, "The Origins of Overkill: Nuclear Weapons and American Strategy, 1945–1950," *International Security*, Spring 1983, p. 15.

205 The limited stockpile of nuclear weapons available to the Strategic Air Command in 1949–1950 was quickly overcome. The stockpile grew to over 12,000 weapons in calendar year 1959 and reached over 22,000 in 1961. By 1956 SAC planning for 1959 envisioned employing thermonuclear bombs with yields of 1.7 to 9 megatons against "Air Power targets" and atomic bombs with yields up to 160 kilotons against lower priority "Systematic Destruction targets" (cities). William Burr, ed., *U.S. Cold War Nuclear Target Lists Declassified for First Time*, National Security Archive Electronic Briefing Book No. 538 (Washington, DC: National Security Archive, George Washington University, December 22, 2015), available at http://nsarchive.gwu.edu/nukevault/ebb538-Cold-War-Nuclear-Target-List-Declassified-First-Ever/.

206 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 17.

207 The first calculations for a hydrogen bomb were run on ENIAC (Electronic Numerical Integrator and Computer) at the University of Pennsylvania from December 1945 to January 1946. George Dyson, *Turing's Cathedral: The Origins of the Digital Universe* (New York: Pantheon, 2012), pp. 74–76.

208 Project RAND, R-173S, pp. 1, 6. In 1960 Hitch and McKean pointed out, in a discussion of criterion errors, that using ratios as decision metrics could overlook the absolute size of gain or cost. Hitch and McKean, *The Economics of Defense in the Nuclear Age*, p. 166.

mend unrefueled turboprop bombers—ones "that had a good range, cost little, and could perform just well enough to bomb the Soviet Union's key industries," and nothing else in a single atomic strike.[209] In the case of the first two criteria the recommended turboprop would have an initial gross weight of about 107,000 pounds; maximizing the third ratio required a heavier bomber with an initial gross weight of about 170,000 pounds.[210]

In using these sort of ratios to calculate the military worth (or utility) of various bomb-bomber combinations, Paxson and his team were following Warren Weaver; in 1946 Weaver had envisioned military worth as optimization: producing "the largest margin of profit—the largest excess of return over cost" relative to the force employed and the results achieved.[211] Unfortunately, the recommendation that the Air Force abandon jet bombers for turboprops encountered resistance on both theoretical and practical grounds. In 1948 the philosopher Abraham Kaplan had observed that the problem of military worth was "one of staggering complexity" and could not be based on military technology alone.[212] Insofar as military worth was to be a basis for decisions, it could "not be formulated without a consideration of the whole social process within which military action takes place, and of the total objectives of that action."[213]

One could argue that R-173's interpretation of military worth as cost-effectiveness did not entirely ignore these sorts of broader considerations. The Truman administration's proposed federal budget for Fiscal Year (FY) 1950 (July 1949 through June 1950), which Truman presented to Congress in January 1949, limited national defense spending to $14.3 billion (in then-year dollars).[214] At the time, the president's main concerns were to reign in deficits and shift funds to other government programs by holding the line on defense. In this pre-Korean War fiscal environment, there were some in the fledgling Air Force—notably the Senior Officers Board charged with producing a long-range plan for USAF research, development, and procurement—who sympathized with R-173's budget-minded analysis.[215] But North Korea's invasion of the South in June 1950 abruptly altered the fiscal environment: for FY 1951 the Defense Department's budget authority grew to over $47 billion, and for FY 1952 it exceeded $60 billion.[216]

209 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, pp. 13, 17.

210 Project RAND, R-173S, p. 6.

211 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 9; and Fred Kaplan, *The Wizards of Armageddon*, p. 89.

212 Abraham Kaplan, *The Concept of Military Worth*, RM-37 (Santa Monica, CA: The RAND Corporation, May 7, 1948), pp. 4, 38.

213 Ibid., p. 3. Emphasis in the original.

214 Harry S. Truman "Annual Budget Message to the Congress: Fiscal Year 1950," January 3, 1949, available at http://www.presidency.ucsb.edu/ws/?pid=13434.

215 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 14.

216 Office of the Under Secretary of Defense (OUSD) (Comptroller), *National Defense Budget Estimates for FY 2015* (Washington, DC: DoD, April 2014), p. 136.

The pragmatic objection to R-173's recommendation that the Air Force be satisfied with turboprop bombers was that SAC under General Curtis LeMay was already well down the road to developing turbojet bombers. In November 1948 the Air Force had ordered 87 B-47Bs, the USAF's first all-jet (albeit medium-range) bomber.[217] While it took several years to deliver these aircraft and fix their teething problems, SAC had four wings equipped with 45 B-47s each by 1952; by 1958 SAC fielded 28 B-47 wings with four additional RB-47 wings.[218] But for intercontinental air warfare, SAC also needed a long-range bomber to replace the B-36. In October 1948 Boeing and the Air Force settled on the all-jet design that became the iconic B-52.[219] The B-52 entered service in 1955, and 76 of the 102 B-52Hs built during 1961–1962 remain in the Air Force's active inventory today.[220] These bombers have repeatedly been modified and upgraded since the early 1960s. The latest upgrade includes a digital architecture, machine-to-machine data transfer, and increased carriage of precision munitions.[221] The Air Force expects these planes to remain in service till at least 2040, and the B-52's longevity certainly goes far to vindicate those USAF leaders in the early 1950s who rejected R-173's recommendation to reverse course in favor of larger numbers of cheaper turboprops.

Perhaps the most strident critic of this recommendation of Paxson's *Strategic Bombing Systems Analysis* was General George Kenney at the Air University. Besides expressing his dismay over R-173's "budget" bombers in a June 1950 letter to Air Force chief of staff General Hoyt Vandenberg, Kenney had the Air University (AU) staff prepare a detailed critique. That analysis argued that R-173's conclusion about preferred platforms depended heavily on the assumption that the Air Force would be constrained by comparatively small budgets; however, the staff analysis suggested that with higher funding levels jets would be more efficient than turboprops.[222] In addition, R-173's single-strike scenario ignored the "salvage" value of bombers that survived the first strike and could fly a second or third mission. "A more realistic, multiple-strike study, the AU Staff asserted, was likely to favor the jet."[223] The single-strike scenario itself also constituted an implicit metric in that it presumed a single, all-out atomic strike was the proper operational context within which to maximize the study's ratios of damage to costs. Alternative scenarios, such as a limited initial atomic strike that might

217 Walter J. Boyne, *Boeing B-52: A Documentary History* (Washington and London: Smithsonian Institution Press, 1981), p. 36.

218 Ibid., p. 37. The Air Force eventually procured over 2,000 B-47 variants.

219 Ibid., pp. 49–50, 53.

220 "B-52 Stratofortress," U.S. Air Force Fact Sheet, current as of December 16, 2015, available at http://www.af.mil/AboutUs/FactSheets/Display/tabid/224/Article/104465/b-52-stratofortress.aspx. 58 B-52Hs remain in operational units at Barksdale AFB, LA, and Minot AFB, ND, and they retain both conventional and nuclear roles. There are another 18 B-52Hs in a reserve unit at Barksdale AFB.

221 Kris Osborn, "Air Force Begins Massive B-52 Overhaul," *DoD Buzz*, July 12, 2013, available at http://www.dodbuzz.com/2013/07/12/air-force-begins-massive-b-52-overhaul/.

222 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 2, p. 15.

223 Ibid., Chapter 2, p. 15.

incentivize Moscow to halt its offensive in Europe, or focusing U.S. bombers on Soviet nuclear forces instead of cities, were ignored.

In hindsight, then, R-173's choice of decision criteria, together with its scenario and budget assumptions, produced a recommendation for the USAF's preferred bombing system that the Air Force was right to reject. R-173 ignored some of the broader aspects of value that RM-28, Lipp's air defense study, had highlighted. To mirror Abraham Kaplan's warning, if the military value was to be a basis for decisions, then the choice of appropriate criteria had to consider higher-level aspects of the problem, including intra-service competition for funds, overall defense strategy, and national objectives. Further, as Schlesinger observed in 1967, taking these higher-level considerations into effect inevitably introduces a substantial but irreducibly subjective element into the choice of decision metrics.

Arguably, the more fundamental error in Paxson's 1950 *Strategic Bombing Systems Analysis* was the drive to create a mathematically rigorous science of warfare, a goal RAND's management abandoned in 1952.[224] The fact is that despite prodigious efforts, no such science along the lines envisioned by Warren Weaver in 1946 has emerged in the decades since R-173 appeared. Contrary to the hopes of many, better data, vastly greater computational power, and more advanced models have yet to come remotely close to eliminating the frictions and uncertainties of war that were so central to Carl von Clausewitz's *Vom Kriege*.[225] Or, to recast the point in the more technical terms of Alan Turing's 1936 paper "Computable Numbers and the *Entscheidungsproblem* [Decision Problem]," there "is no algorithmic process to determine the future—whether it's the future of a computer program, a thought process of the human mind, or the universe as a whole."[226]

Ironically, despite the Air Force's rejection of R-173's recommendations about both future bombs and bombers, RAND's "systems analysis methodology was rapidly embraced by both the Senior Officers Board" and the Air Force generally.[227] Recall that in April 1949 defense secretary Louis Johnson had canceled the Navy's supercarrier (CVA-58) just days after its keel was laid. The decision effectively gave the intercontinental atomic attack mission to the Air Force and the long-range B-36.[228] In the resulting struggle between the Navy and the Air Force over roles and missions, many airmen saw the utility of "the apparent scientific

224 Even before Paxson's *Strategic Bombing Systems Analysis* was completed, RAND began a comprehensive air defense study headed by Edward J. Barlow. Barlow's initial design envisioned 54 separate component analyses for the first phase of the study. Jardini, *Thinking Through the Cold War*, p. 35. Although Barlow's design team later reduced the number of subprojects to seven, after his air defense study was completed in 1951 RAND management concluded that future research should concentrate "on selecting key component problems in which large payoffs may exist," and no further systems analyses of the breadth of those led by Paxson and Barlow "should be contemplated." Ibid., p. 37.

225 Barry D. Watts, *Clausewitzian Friction and Future War*, revised edition, McNair Paper 68 (Washington, DC: Institute for National Strategic Studies, National Defense University, 2004), pp. 68–78, 85–90.

226 Charles Petzold, "Turing Machines That Run Forever," *Petzold Book Blog*, May 18, 2008, available at http://www.charlespetzold.com/blog/2008/05/Turing-Machines-That-Run-Forever.html.

227 Jardini, *Thinking Through the Cold War*, p. 37.

228 Phillip S. Meilinger, "The Admirals' Revolt of 1949: Lessons for Today," *Parameters*, September 1989, pp. 87–88.

rigor and rationality of systems analysis" in competing with the other services for funding, notwithstanding the limitations of systems analysis.[229]

Working Hypotheses and a Caveat

First, the judgments and conclusions reached in this chapter's three cases all depend on the choice of analytic criteria. Second, it is difficult to discern many commonalities across the various criteria and metrics examined. The need for appropriate criteria for judgment and decision in any specific case appears to be ubiquitous. But third, the criteria themselves are tied to the specifics of each individual case. The criterion of gaining daylight air superiority over the Luftwaffe prior to the Normandy landings is incommensurate with merchant/U-boat exchange ratios in the Battle of the Atlantic or R-173's ratios of damage inflicted to various kinds of costs (dollars expended for bombs and bombers, the pounds of aircraft lost in a single all-out atomic blitz, and the number of aircrews lost). Finally, the sheer diversity of the criteria in these three cases supports the suspicion that there may be no universal methodology for selecting appropriate criteria for any strategic decision, historical judgment, or assessment.

In fairness to those who developed systems analysis at RAND after World War II, it must be said that recommending instrumentalities and techniques for intercontinental air warfare to the U.S. Air Force was a considerably harder problem than those faced by wartime operations analysts. After all, analysts like P. M. S. Blackett were working ongoing operational problems, such as deciding whether trans-Atlantic convoys should be large or small and whether British Coastal Command bombers flying at night should be painted black or white. Such problems had very near-term time horizons, and combat operations provided immediate feedback on the efficacy of larger convoys and painting bombers white. By contrast, RAND's systems analysis method sought to make recommendations about weapons and force postures that the Air Force would, if accepted, have to live with for 10–15 years or longer, and the absence of feedback from ongoing combat operations confronted the systems analysts with much greater uncertainty. Early analyses, such as Paxson's massive bombing systems analysis, unquestionably had their flaws. But the problems they sought to address were much harder and more abstract than those faced by World War II operational researchers.

229 Jardini, *Thinking Through the Cold War*, p. 38.

CHAPTER 2

Practical Theory

> *I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely, in your thoughts, advanced to the state of science, whatever the matter may be.*
>
> Lord Kelvin, 1883[1]

> *Science is measurement.*
>
> The Cowles Commission Motto, 1932[2]

> *[A]nything that has more upside than downside risk from random events (or certain shocks) is antifragile; the reverse is fragile.*
>
> *Fragility is quite measurable, risk not so at all, particularly risk associated with rare events.*
>
> Nassim Taleb, 2012[3]

This chapter delves into some of the theoretical aspects of the general problem of selecting appropriate analytic measures. Central to these more theoretical concerns is the development of what came to be known at RAND as the "criterion problem." Succeeding chapters return to documenting the role, variability, and ubiquity of analytic measures in a series of empirical cases.

---

1 Sir William Thomson, *Popular Lectures and Addresses*, Vol. I, *Constitution of Matter* (London and New York: Macmillan, 1889), pp. 73–74.

2 Carl F. Christ, "History of the Cowles Commission, 1932–1952," in *Economic Theory and Measurement: A Twenty Year Research Report* (Chicago: Cowles Commission for Research in Economics, 1952), p. 61. The motto currently on the foundation's crest is "theory and measurement." The original motto was obviously inspired by Lord Kelvin.

3 Nassim Nicholas Taleb, *Antifragile: Things That Gain from Disorder* (New York: Random House, 2012), pp. 5, 8.

Measurement Theory and Systemic Value

As the statements about the importance of measurement from Lord Kelvin and the Cowles Commission suggest, the measurement of physical quantities and phenomena has been a central, if not definitive, feature of both the physical and social sciences. It seems pertinent to wonder what measurement theory may have to say about the selection of appropriate analytic measures and decision criteria? The short answer is that the analytic measures and criteria that are the focus of this report are fundamentally about the systemic value of outcomes. For this reason, neither the choice of systems of units and standards nor measurement theory appears to have much to say about the selection of appropriate analytic measures for judging historical outcomes or making strategic choices among alternative systems or courses of action.

Measurement is essentially about associating numbers with physical quantities and phenomena found in the real world based upon rules that preserve the correspondence between real-world attributes and numbers. While there is evidence that other early civilizations devised measurement standards or scales, the Egyptian royal cubit (20.62 inches) is generally recognized as the most ubiquitous standard of length in the ancient world. Counting and arithmetic relations, such as addition and greater than, are intimately entwined with standard units for length, weight, area, etc. For example, two bushels of wheat plus three bushels equals five bushels just as laying a stick 2 cubits in length end-to-end with a 3-cubit-length stick yields a total length of 5 cubits. Similarly, a 3-cubit-long stick is longer than a 2-cubit-long stick.

Patrick Suppes traces modern views on quantitative measures to Isaac Newton's *Universal Arithmetick*. Newton asserted that the relationship between any two quantities of the same kind can be expressed by a positive real number.[4] Subsequently, two complementary research traditions on measurement arose. One begins with Herman von Helmholtz and continues with the work of Otto Hölder and Norman Cambell on axiomatics and structure-preserving morphisms; the other tradition is the work undertaken by the psychologist Stanley Smith Stevens and his school on scale types and transformations.[5] However, it was not until 1951 that Patrick Suppes provided an entirely adequate formal basis for a system of extensive quantities, meaning those "to which numbers can be assigned uniquely up to a similarity transformation (that is, multiplication by a positive constant)."[6]

4 Patrick Suppes, "A Set of Independent Axioms for Extensive Quantities," *Portugaliae Mathematica* 10, no. 4, 1951, p. 163. Newton's *Universal Arithmetick* was first edited and published in 1707 by William Whiston, Newton's successor as the Lucasian professor of mathematics at Cambridge, based on Newton's notes. Newton was so unhappy with Whiston's edition that he refused to have his name appear on the book.

5 José A. Díez, "A Hundred Years of Numbers. An Historical Introduction to Measurement Theory 1887–1990," *Studies in History and Philosophy of Science* 28, no. 1, 1997, p. 167.

6 Suppes, "A Set of Independent Axioms for Extensive Quantities," p. 164.

In 1962 Suppes and Joseph Zinnes noted that while psychologists were paying homage to measurement theory, students of the subject were confronted with a bewildering, often conflicting, set of rules.

> To cite just one peculiar, yet uniformly accepted example, as elementary science students we are constantly warned that it "does not make sense" (a phrase often used when no other argument is apparent) to add together numbers representing distinct properties, say, height and weight. Yet as more advanced physics students we are taught, with some effort no doubt, to multiply together numbers representing such things as velocity and time, or to divide distance numbers by time numbers. Why does multiplication make "more sense" than addition?[7]

The first fundamental problem of measurement theory is to show that certain aspects of the arithmetic of numbers have the same relationships as those among the empirical objects or events being measured. For example, measurements of human heights should assign larger numbers to correspondingly taller individuals.

The early history of mathematics shows how difficult it was to divorce arithmetic and algebra from empirical structures. Up until the development of non-Euclidian geometries in the nineteenth century, most observers interpreted Euclid's "flat" geometry as describing the geometry of physical space. As Suppes and Zinnes explained in 1962,

> The ancient Egyptians could not think of 2 + 3, but only of 2 bushels of wheat plus 3 bushels of wheat. Intellectually, it is a great step forward to realize that the assertion that 2 bushels of wheat plus 3 bushels of wheat equal 5 bushels of wheat involves the same mathematical considerations as the statement that 2 quarts of milk plus 3 quarts of milk equal 5 quarts of milk.
>
> From a logical standpoint there is just one arithmetic of numbers, not an arithmetic for bushels of wheat and a separate arithmetic for quarts of milk.[8]

Moreover, the language of arithmetic has no meaning in and of itself:

> The theoretical mathematician deals entirely with the realm of the formal language and is concerned with the structure and relationships within the language. The applied mathematician or statistician, on the other hand, is concerned not only with the language, but the relationship of the symbols in the language to real-world objects and events.[9]

It is this concern about the meaning of numbers when applied to the real world that raises a concern about measurement.

S. S. Stevens' theory of measurement scales presumes an isomorphism between the numbers and the physical quantities or phenomena they represent or quantify. In his seminal 1946

7 Patrick Suppes and Joseph L. Zinnes, *Basic Measurement Theory*, Technical Report No. 45 (Stanford, CA: Institute for Mathematical Studies in the Social Sciences, Applied Mathematics and Statistics Laboratories, March 15, 1962), p. 1, available at https://suppes-corpus.stanford.edu/techreports/IMSSS_45.pdf.

8 Ibid., p. 3.

9 David W. Stockburger, *Introductory Statistics: Concepts, Models, and Applications*, 3rd Web Edition (Springfield, MO: Missouri State University, 1998), Chapter 4, available at http://www.psychstat.missouristate.edu/IntroBook3/sbk04.htm.

article, he described four distinct kinds of scales: (1) nominal, (2) ordinal, (3) interval and (4) ratio.[10] Nominal scales use numbers to label the members of a type or class. In the case of objects being grouped into subgroups with no object being in more than one subgroup, the assignment of a unique numeral to all members of each subgroup renders it "meaningless to find means, standard deviations, correlation coefficients, etc."[11]

Ordinal scales rise from assigning numbers that preserve the rank ordering of the real-world objects or events. A classic example is the hardness of minerals.[12] Permissible statistical calculations for ordinal scales are medians and percentiles.

Interval scales not only preserve the order between real-world objects or events, but also the magnitude of the differences between them. Things are assigned numbers such that the differences between the numbers reflect the differences of the attribute being measured. Most psychological measurement aspires to create interval scales and sometimes succeeds.[13] Interval scales are "quantitative" in the ordinary sense of the word. But if an arbitrary zero point is agreed upon, as in the Centigrade and Fahrenheit (F) temperature scales, then it is meaningless to say, for instance, that today's temperature of 60°F is twice that of yesterday's 30°F. So long as there is no arbitrary zero point, all the usual statistics (mean, standard deviation, rank-order correlation, etc.) can be applied to interval scales.

Ratio scales are those most commonly encountered in physics. They are possible only:

> when there exist operations for determining all four relations: equality, rank-order, equality of intervals, and equality of ratios. Once such a scale is erected, its numerical values can be transformed (as from inches to feet) only by multiplying each value by a constant. An absolute zero value on some scales (e.g., Absolute Temperature) may never be produced. All types of statistical measure are applicable to ratio scales, and only with these scales may we properly indulge in logarithmic transformations such as are involved in the use of decibels.[14]

Stevens has noted that "Both nominal and ordinal scales have at times been called intensive, and both interval and ratio scales have sometimes been labeled extensive."[15] The distinction appears to have to do with the extent to which a relatively full range of statistical methods are permissible because of the isomorphism between the numerical relations and those of the physical objects or events being measured. In 1951 Suppes articulated the distinction as follows:

> Intensive quantities are those which can merely be arranged in a serial order; extensive qualities are those for which a "natural" operation of addition or combination can also be specified.

10 S. S. Stevens, "On the Theory of Scales of Measurement," *Science* 103, no. 2648, June 7, 1946, p. 678.

11 Stockburger, *Introductory Statistics: Concepts, Models, and Applications*, Chapter 4.

12 Stevens, "On the Theory of Scales of Measurement," p. 679.

13 Ibid., p. 679.

14 Ibid., pp. 679–680.

15 Ibid., p. 678.

> Another, more exact way of making a distinction of this order is to say that intensive quantities are quantities to which numbers can be assigned uniquely up to a monotone transformation, and extensive quantities are quantities to which numbers can be assigned uniquely up to a similarity transformation (that is, multiplication by a positive constant).[16]

Suppes then went on to "present a formally adequate system of axioms for extensive quantities" that remedied certain defects in earlier attempts by Hölder and others to put the analysis of extensive quantities on a solid axiomatic foundation.[17]

For present purposes, we need not delve further into the complexities of Suppes' axiomatization of extensive quantities. On the one hand, José Díez has argued that Suppes' 1951 paper marks the beginning of mature measurement theory, which later developed rapidly, especially during the 1960s.[18] On the other hand, Suppes was right to highlight two further problems his axiomatization did not address. First, his formal system had to contain an infinite number of elements, which flagrantly violated the obvious finitistic requirements of empirical measurement; second, it assumed perfect measurement, whereas real-world measurements are neither infinitely sensitive nor error free.[19] Thus Suppes' axiomatic theory of extensive quantities, as he emphasized in 1951, required profound alteration to address these two problems, and they lay behind much of the later development of measurement theory.

From physics and chemistry to economics and psychology, the sciences need measurement. As George Furnas observed in reviewing the second volume of the seminal trilogy *Foundations of Measurement* by Patrick Suppes, David Krantz, Duncan Luce, and Amos Tversky, "Measurement elevates the qualitative to the quantitative."[20] But, as Furnas added, the practice of measurement in the real word is "never as pristine as the formal definitions and derivations of *Foundations*" and "must be viewed through a veil of noise."[21]

The inevitable question at this juncture is as follows: what does measurement theory have to do with the choice of efficacious analytic criteria? In answer, consider first the criteria—whether appropriate or not—that came up in Chapter 1's discussions of RM-28 and R-173. The latter—Paxson's *Strategic Bombing Systems Analysis*—used ratios of effects to costs in selecting the optimum combination of atomic bombs and bombers for executing a single

16 Suppes, "A Set of Independent Axioms for Extensive Quantities," pp. 163–164. A monotone (or monotonic) transformation is one that preserves the inequalities of its arguments. If T is a monotone transformation, then if and only if x > y, then Tx > Ty; and, conversely, if and only if x < y, then Tx < Ty.

17 Ibid., p. 164.

18 Díez, "A Hundred Years of Numbers. An Historical Introduction to Measurement Theory 1887–1990," p. 167.

19 Suppes, "A Set of Independent Axioms for Extensive Quantities," p. 172.

20 George W. Furnas, "Volume II: Geometrical, Threshold, and Probabilistic Representations," book review, *Applied Psychological Measurement* 15, no. 1, March 1991, p. 103. The first volume of *Foundations of Measurement*, subtitled *Additive and Polynomial Representations*, appeared in 1971. The second volume, *Geometrical, Threshold, and Probabilistic Representations*, appeared in 1989, and the third volume, *Representation, Axiomatization, and Invariance*, appeared in 1990. Dover published unabridged but slightly corrected versions of all three volumes in 2007.

21 Ibid., p. 103.

strike. Regardless of the shortcomings of these particular criteria, the unmistakable intent was to judge the overall or *systemic value* of the various alternatives. The Air Force's rejection of the recommendation about what bombers to develop and field argues that there were other, non-quantitative aspects of the situation that needed to be taken into account (but were not). In the broadest sense, then, systemic value does not appear to be amenable to precise quantification. Systemic value is more nebulous—especially when trying to judge the operational, strategic, and social utility of consequential choices under uncertainty.

The same conclusion follows in the case of trying to discern the most or more important causal drivers of historical outcomes. Many have judged Alan Turning and W. Gordon Welchman's development of a machine that could decrypt naval Enigma messages as the decisive contribution to the Allies winning the Battle of the Atlantic.[22] But, to repeat Bennett's point, intelligence is useless without the means to take advantage of it, and the Allies' growing ASW capabilities in 1943 provided the wherewithal to do precisely that.

The answer to the question about measurement theory and analytic criteria, then, seems clear. The quantitative metrics that are the concern of measurement theory are one thing. Analytic criteria for assessing systemic value are qualitatively something else altogether. Indeed, this is precisely why this report is about analytic criteria rather than the empirical measures and metrics central to hard sciences such as physics.

RAND's Criterion Problem

The earliest formulation of the criterion problem this author is aware of is a January 1951 RAND paper entitled *What Is the Best System?* (D-860).[23] Its introductory formulation of the problem remains relevant in the second decade of the twenty-first century:

> RAND's official purpose is to recommend preferred instrumentalities of air warfare between the United States and other continents. Our systems analyses, in seeking this end, combine the individual contributions of many specialists. However, specialized knowledge and extensive computations are insufficient to indicate which systems are better or worse than others. The best of many alternative systems cannot be determined unless a valid criterion is employed.
>
> The purpose of this paper is partly negative and partly positive. Negatively, it can be shown that certain "criteria" are logically incorrect and will not discriminate between good systems and bad, except by accident. . . . More positively, this paper indicates the nature of logically correct criterion, but it also explains some of the practical difficulties that prevent its pure employment.
>
> The essence of the criterion problem is that we do not know how to define and measure what might be loosely termed the utility and disutility functions. Thus, in actual practice, we can never

22 Hodges, *Alan Turing: The Enigma*, pp. 182–183.

23 The author is grateful to Andrew May for providing a copy of *What Is the Best System?* (D-860) in April 2015 from documents Andrew Marshall left in the Office of Net Assessment after he retired in January. Abraham Kaplan, of course, had dealt with many of the complexities of measuring military worth as a basis for decisions bearing on conducting future wars and sustaining peace, but he did not use the term "criterion problem."

> distinguish with confidence the optimum optimorum. However, there are both reasonable and practical ways of approximating this ideal that are more logical and justifiable than other standards that might have been used. It is these realistic optima that RAND must discover from its systems analyses and recommend to [the] Air Staff.[24]

Given the widespread infatuation with scientific methods of solving military problems that prevailed after World War II, it should come as no surprise that RAND's *What Is the Best System?* exhibited a distinctly mathematical approach. After stating categorically that the selection of preferred weapons or techniques for intercontinental air warfare required a "Social Utility function" that we neither possess nor know how to derive, the authors turned to "suboptimizing procedures which give incomplete answers, but are consistent—so far as they go—with the optimal function."[25] In this vein they invoked the mathematical principle of separation of variables as a metaphor for identifying, measuring, and separating at least some of the variables and coefficients underlying Social Utility into two additive functions: an objectives function (for example, Soviet targets destroyed by a strategic bombing campaign) and a cost function (for example, the cost of the corresponding offensive bombing system in dollars).[26] "In the complete Social Utility function there are determining variables which have positive coefficients [i.e., positive first derivatives of Social Utility with respect to these variables] and there are variables with negative coefficients."[27] The objectives and cost functions reflect social utility and disutility, respectively.

In the case of selecting an objectives function, D-860 stressed that "there can be no escape from the need for judgment." The fundamental political goals for which a nation prepares for or undertakes war "will usually be categorical and unamenable to numerical analysis."[28] And even the dollars allocated in defense budgets to a military system or capability "is a generalized claim upon the current resources (e.g., labor, services of capital, etc.) and the existing stock (e.g., petroleum and U^{235} reserves) of a nation."[29] Ideally, the true economic cost of a bombing system must also reflect the alternative uses to which resources could have been assigned, as well as the lives lost in the operation or campaign.[30] Thus, even dollar costs omit hard-to-quantify aspects of social disutility and generally ignore the salvage value of bombing systems that survive the initial strike or operation.

Recall that Paxson's *Strategic Bombing Systems Study* (R-173) maximized quantifiable or "engineering" criteria as a basis for recommending to the Air Force preferred bomber-weapon

24 Alchian et al., D-860, pp. 1–2. The Latin term *optimum optimorum* is usually translated as "the best of the best." It is used when there are several criteria that are locally optimal to refer to the best among these.

25 Ibid., p. 3.

26 Ibid., pp. 4–5, 43. Separation of variables is a technique for solving for ordinary and partial differential equations.

27 Ibid., p. 3.

28 Ibid., p. 9.

29 Ibid., p. 14.

30 Ibid., pp. 17–20.

combinations for atomic attack. The 107,000-lb turboprop was recommended by two criteria: the ratio of the maximum damage inflicted on the USSR's war industry (without regard for losses) over a fixed cost; and the ratio of damage inflicted per pound of aircraft lost. The six authors of D-860 concluded, however, that the "economic criterion" reflecting social utility will not select the same "best" system as engineering criteria, such as those used in R-173:

> Different hardware criteria, if applied with vigor over the widest possible range of performance characteristics, are likely to lead to the selection of widely different systems. The maximization or minimization of the ratio of some physical output to some physical input will normally result in the selection of an economically inefficient system. There are two major reasons why this will be so. First, diminishing physical returns are an almost universal fact of life; holding all other inputs constant, one can hardly ever indefinitely redouble output by redoubling one input. Second, more than one input will almost certainly be an economic good or service; it will then be impossible to attain maximum output per some input without wasting other scarce goods and occasioning unnecessarily higher total costs.[31]

The implication seems clear. While engineering criteria, such as the ratio of the damage an atomic strike would inflict on the Soviet Union's war industry per pound of SAC aircraft lost, can be precisely defined and quantified, systemic social utility and disutility cannot. In other words, precisely defining, much less quantifying, social utility and disutility give every indication of being an impossible problem.

Here it is worth remembering that impossible problems do exist. As Charles Petzoid wrote in 2008:

> There is no way to trisect an angle with ruler and compass, no way to square the circle, and no way to prove Euclid's fifth postulate from the first four. There are no integer solutions to $x^n + y^n = z^n$ for n greater than 2, there is no way to establish the consistency of arithmetic within the system, and there is no general decision procedure for first-order logic.[32]

Granted, the impossibility of precisely defining and quantifying functions for social utility and disutility, even within the narrow context of intercontinental air warfare, almost certainly cannot be proved by strictly logical and mathematical methods. But D-860's statement that we neither possess nor know how to derive these functions remains as unchallenged today as it was in 1951. As Charles Hitch and Roland McKean summarized the situation in 1960:

31 Ibid., p. 30.

32 Charles Petzold, *The Annotated Turing; A Guided Tour through Alan Turing's Historic Paper on Computability and the Turing Machine* (Indianapolis, IN: Wiley Publishing, 2008), p, 279. Pierre Wantzel proved algebraically the impossibility of trisecting an arbitrary or general angle with only ruler and compass in 1836. "Angle Trisection," *Wolfram MathWorld*, available at http://mathworld.wolfram.com/AngleTrisection.html. The development of logically consistent non-Euclidian geometries by Bernhard Riemann, János Bolyai, and Nikolai Lobachevsky established the independence of Euclid's fifth postulate. Andrew Wiles and his student Richard Taylor proved Pierre de Fermat's "last theorem"—the conjecture that there are no integer solutions to $x^n + y^n = z^n$ for $n > 2$—in 1995. In 1931 Kurt Gödel proved that there is no way to prove the consistency of arithmetic within the system of arithmetic. Finally, Alan Turing and Alonzo Church proved there is no general decision procedure for the first-order predicate logic in 1936–1937.

> The choice of an appropriate economic criterion is frequently the central problem in designing a systems analysis. In principle, the criterion we want is clear enough: the optimal system is the one which yields the greatest excess of positive values (objectives) over negative values (resources used up, or costs). But . . . this clear-cut ideal is seldom a practical possibility in military problems. Objectives and costs usually have no common measure: there is no generally acceptable way to subtract dollars spent or aircraft lost from enemy targets destroyed. Moreover . . . there may be multiple objectives or multiple costs that are incommensurable.[33]

Subsequent accounts of the criterion problem by RAND economists have generally not been as technical or mathematical as *What Is the Best System?* For example, Hitch and McKean's 1960 *The Economics of Defense in the Nuclear Age* devoted an entire chapter to the problem of selecting criteria or tests of preferredness for deciding among alternative systems. However, nowhere in this chapter does one find the explicitly mathematical tone evident in *What Is the Best System?* Generally absent from later accounts are the direct appeal in *What Is the Best System?* to the separation of variables and the association of social utility and disutility with, respectively, positive and negative first derivatives of whatever determining variables can be identified.

True, *The Economics of Defense in the Nuclear Age* does contain an appendix that details mathematical methods for maximizing the efficiency with which limited defense resources are used. But these methods are all prefaced with the assumption that the criterion problem has been solved when in reality it remains unsolved to this day.[34] And given how inappropriate—if not downright awful—many of the criteria chosen by RAND engineers were in the early days of systems analysis, the economists under Hitch had every incentive to highlight more prominent mistakes. The chapter on the criterion problem in *The Economics of Defense in the Nuclear Age* discusses the following errors:

- Overlooking the absolute size of gain or cost;
- Setting the wrong size of gain or cost;
- Neglecting "spillovers" (external economies or diseconomies); and
- Using the wrong concepts of cost or gain.[35]

In 1964, McKean offered an expanded list of the most frequently occurring errors in choosing analytic criteria:

- Ignoring the absolute scale of objective or cost;

33 Hitch and McKean, *The Economics of Defense in the Nuclear Age*, p. 120. Attempts to define and maximize a "social welfare function" for national security "is completely impractical and sterile. We have to break economic problems, like so many others, into manageable pieces before we can make a good beginning at finding solutions." Ibid., pp. 3–4.

34 Ibid., p. 362.

35 Ibid., pp. 165–173.

- Setting the wrong objectives or scale of objective;
- Ignoring uncertainty;
- Ignoring effects on other operations;
- Adopting the wrong concepts of cost;
- Ignoring the time dimension;
- Trying to use an "overdetermined" test; and
- Applying "good" criteria to the wrong problems.[36]

It will not be necessary to explore all these errors in detail. Instead, it will suffice to focus on the first. Both lists start with the same error—overlooking or ignoring the absolute magnitude of gain or cost—and its exposition concentrates on the ratio criteria used in Paxson's 1951 *Strategic Bombing Systems Study*. In 1964 McKean explained the potential pitfall of maximizing ratios of gain to cost without regard to the magnitude of gain or cost as follows:

> One common test is the ratio of effectiveness to cost, that is, the ratio of achievement-of-objective to cost. For instance, in the selection of bombing systems, the test might be the maximum ratio of targets destroyed to cost, that is, maximum targets destroyed per dollar cost. This sounds like a reasonable criterion, yet it could let the scale of the system wander freely from a $500 billion system to cheap weapons of low effectiveness. As a matter of fact, it would probably favor existing weapons of early vintage. Suppose, for instance, that one bombing system, already on hand and relatively simple to maintain, would destroy 10 targets and cost $1 billion (a ratio of 10 to 1), while another system would destroy 200 targets and cost $50 billion (a ratio of 4 to 1). Should we choose the former—a weapons system which might merely invite and lose a war inexpensively? The answer is no; we cannot afford to ignore the scale of the activity—the absolute amount of damage the system could do, or the absolute amount of its cost.[37]

In the case of Paxson's *Strategic Bombing Systems Study*, the recommendation that the Air Force forego jet bombers in favor for turboprops arguably neglected the spillover consequences of failing to embrace the state-of-the-art jet propulsion technology that powers Air Force fighters and bombers to this day.

Wohlstetter's Strategic Air Base Study

To this point, discussion of the criterion problem has focused on the difficulties of choosing appropriate criteria and the more common mistakes in making such choices. However, even in the 1950s, there were systems analyses in which RAND analysts selected appropriate decision criteria. Perhaps the most famous case is the strategic air base study that Albert Wohlstetter

36 R. N. McKean, "Criteria," in E. S. Quade, ed., *Analysis for Military Decisions*, R-387-PR (Santa Monica, CA: the RAND Corporation, November 1964), pp. 85–90.

37 Ibid., p. 85.

and his colleagues completed in 1954. So outstanding was this study that Edward Quade chose it in 1964 to help clarify the scope and methods of systems analysis. As he pointed out, *Selection and Use of Strategic Air Bases* (R-266) not only resulted in a more secure posture for Strategic Air Command's bomber force but also saved the Air Force a billion dollars in planned construction costs for overseas bases.[38]

Originally conceived as an exercise to determine the best locations for overseas bases for SAC bombers in terms of terrain, climate, and so forth, R-266 "became in the end a study of U.S. strategic deterrent policy."[39] Initially, the Air Force's request for an overseas basing study looked so mechanical and tedious that Wohlstetter, at the time a consultant to RAND, nearly declined to work on it. Upon reflection, though, he realized that as Soviet capabilities grew, overseas bases would be increasingly vulnerable, which meant that the real question was what sort of basing system would offer the best trade-offs across a range of variables such as the strike distance from base to target; vulnerability to Soviet attack; attrition to the USSR's territorial air defenses; and, of course, cost.[40]

As Wohlstetter's understanding of the basing problem expanded, it became clear that he could not adequately address the complexities of the issue on his own. But in the wake of the follow-on to Paxson's massive *Strategic Bombing Systems Analysis*, the *Multiple-Strike Study* conducted by engineer Richard Schamburg and mathematicians Edward Quade and Robert Specht, RAND's managers were disinclined to provide Wohlstetter with the people needed for a second large systems analysis.[41] Wohlstetter, therefore turned to persuasion, eventually interesting aeronautical engineer Robert Lutz and economists Henry Rowen and Fred Hoffman in the problem. Each then worked on a number of different issues resulting in a group effort that was more cross-disciplinary than any previous RAND systems analyses.[42] When *Selection and Use of Strategic Air Bases* was published in 1954, Wohlstetter included Hoffman, Lutz, and Rowen as co-authors. He also acknowledged the contributions of 15 other analysts as well as "customer" feedback from Air Force officers at SAC and Headquarters Air Force.[43] Over the course of the basing study, the Air Force received some 92 briefings of evolving results as the team investigated issues ranging from the feasibility of

38 E. S. Quade, "The Selection and Use of Strategic Air Bases: A Case History," in Quade, R-387-PR, p. 24. Wohlstetter began the analysis in the spring of 1951. An essentially complete version of the study was briefed to SAC and Air Force headquarters in mid-1953, and a full report was published in April 1954.

39 Ibid., p. 62.

40 May, "The RAND Corporation and the Dynamics of American Strategic Thought," Chapter 4, p. 14. Each chapter in the electronic files of May's dissertation starts with page 1.

41 Ibid., Chapter 2, p. 18; Chapter 4, p. 15.

42 Ibid., Chapter 4, p. 15.

43 A. J. Wohlstetter, F. S. Hoffman, R. J. Lutz, and H. S. Rowen, *Selection and Use of Strategic Air Bases*, R-266 (Santa Monica, CA: the RAND Corporation, April 1954), p. iii. R-266 was originally classified TOP SECRET. The declassified version is over 380 pages.

nuclear-powered bombers to aerial tanker requirements, increasing Russian capabilities, and the cost of a summer campaign to destroy 80 percent of the Soviet Union's war industry.[44]

When Wohlstetter began the basing study in May 1951, Congress had authorized $3.5 billion for airbase construction in Fiscal Year 1952, of which $1.5 billion was earmarked for overseas operating bases for SAC's bomber force.[45] The objectives that motivated the basing study were straightforward:

> The major danger was considered to be the possibility of a Soviet attack on Western Europe. The general U.S. objective was deterrence (of such an attack) based on our unilateral atomic capability. If deterrence failed, the SAC war objective was to destroy the Soviet Union's industrial base for waging conventional war.[46]

Preliminary analysis showed that "The important problem was not how to acquire, construct, and maintain air bases in foreign countries but *where and how to base the strategic Air Force and how to operate this force in conjunction with the base system chosen*."[47] To bound the problem, Wohlstetter decided to limit the analysis to recommending a basing scheme for the SAC bomber force programmed for the 1956–1961 timeframe rather than a theoretical optimum force that had little chance of being procured. The point of departure chosen for the analysis was the basing scheme the Air Force had programmed for 1956: it envisioned basing SAC's bombers in the United States in peacetime but, in time of war, deploying all the medium-range B-47s and some of the heavy bombers to advanced overseas bases and operating them from there.[48] Including this starting point, Wohlstetter and his colleagues eventually settled on four alternatives:

1. Bombers based on advanced overseas operating bases in wartime;
2. Bombers based on intermediate overseas operating bases in wartime;
3. U.S.-based bombers operating intercontinentally with the aid of air refueling; and
4. U.S.-based bombers operating intercontinentally with the help of ground refueling at overseas staging bases.[49]

---

44 Quade, "The Selection and Use of Strategic Air Bases: A Case History," pp. 28, 34, 63.

45 A. J. Wohlstetter, *Economic and Strategic Considerations in Air Base Location: A Preliminary Review*, D-1114 (Santa Monica, CA: the RAND Corporation, December 29, 1951), Section 1, available at http://www.rand.org/about/history/wohlstetter/D1114.html.

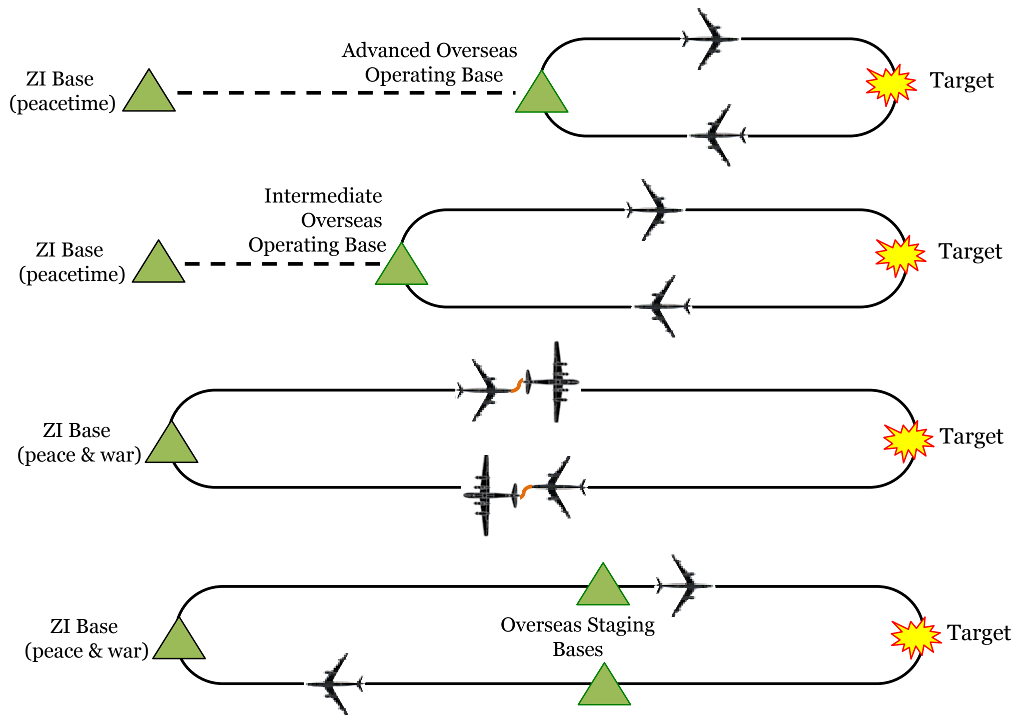
46 Quade, "The Selection and Use of Strategic Air Bases: A Case History," p. 25. At this point Air Force base planning had not yet taken into account the possibility of the Soviet Union having nuclear weapons in quantity.

47 Ibid., p. 26. Emphasis in the original.

48 Ibid., p. 27.

49 Wohlstetter, Hoffman, Lutz, and Rowen, R-266, p. vii.

FIGURE 4: R-266 BASING OPTIONS



The next step was to choose appropriate variables for assessing these alternatives. As the analysis proceeded and understanding of the basing problem by Wohlstetter and his team deepened, the critical factors turned out to be four distances: "from the proposed bases to (1) the targets, (2) the favorable entry points [for SAC bombers] into enemy defenses, (3) the source of base supply, and (4) the points from which the enemy could attack these bases."[50] *Selection and Use of Strategic Air Bases* used these critical distances to analyze their joint effects "on the costs of extending bomber radius; on how the enemy may deploy his defenses and the numbers of our bombers lost to enemy fighters; on logistics costs; and on base vulnerability and our probable loss of bombers on the ground."[51] On the basis of this analysis, the preferred basing schemes were those associated with the fourth alternative in Figure 4: U.S.-based bombers operating intercontinentally with the help of ground refueling at overseas staging bases. Overseas operating base systems, whether advanced bases closest to enemy targets or intermediate bases further back, were too vulnerable, and, "while a U.S.-based

50 Quade, "The Selection and Use of Strategic Air Bases: A Case History," p. 29. It should be stressed that the four critical distances are variables that would contribute to systemic social utility and disutility functions, if only we knew how to define and compute those functions.

51 Wohlstetter, Hoffman, Lutz, and Rowen, R-266, p. v.

air-refueled system would buy lower base vulnerability," it would do so "at so high a cost that total striking power would be drastically reduced."[52]

Several points warrant emphasis. First, it took some thoughtful analysis to develop a robust and realistic set of basing alternatives. These alternatives were, of course, bounded by the state of aerodynamic and propulsion technologies in the 1950s. For example, to have built an intercontinental radius capability into a bomber of the B-47 type would have made the aircraft enormous in size, costly, and vulnerable.[53] But the more technologically realistic solution of using air refueling to increase the radius of the B-47 to 3,600, 4,200, and 5,200 nautical miles (nm), respectively, increased basing system costs three, five, and ten times compared to the cost of the B-47's unfueled radius of 1,750 nm.[54]

In the end, however, R-266's fundamental decision criterion emerged from "examining a very large number of potential conflicts and the interdependent choices of both sides in these conflicts."[55] These "tests" focused on whether enough SAC bombers survived a Soviet surprise atomic attack to mount an effective retaliatory strike.[56] When the fourth alternative (U.S.-based bombers operating intercontinentally with the help of ground refueling at overseas staging bases) was compared with the other three, "It excelled in extending the bomber force's round-trip radius more cheaply; enabling bombers to bypass Soviet defenses and interceptors and reach enemy targets more effectively; decreasing logistical and operational costs; and increasing the quality and time interval of tactical warning, as well as lowering the vulnerability of bases and bombers on the ground to attack by the Soviet Union's growing stockpile of aircraft-delivered atomic bombs."[57]

Second, the importance of the critical distance variables that emerged during the study, much less their influence on total costs, was not self-evident at the outset. Instead, they emerged over time as Wohlstetter and his colleagues immersed themselves in the complexities of the basing problem. Here an important step was separating the locality and location costs of the alternative basing systems.

There are two kinds of bombing system costs traceable to basing decisions. Those influenced by site characteristics such as weather, terrain, availability of construction industry, existing defense, etc., may be called *locality* costs; those which pertain to such critical general base relationships as the routes from the United States to base, from base to target (including the

52 Ibid., p. vii.

53 Ibid., p. xiii.

54 Ibid., p. xiv.

55 Albert Wohlstetter, "Theory and Opposed-System Design," in Robert Zarate and Henry Sokolski, eds., *Nuclear Heuristics: Selected Writings of Albert and Roberta Wohlstetter* (Carlisle, PA: Strategic Studies Institute, January 2009), p. 151.

56 Barry D. Watts, telephone conversation with Andrew W. Marshall, September 9, 2016. Marshall, who was blessed with a near photographic memory, is adamant that SAC survivability eventually emerged as Wohlstetter's main analytic criterion for assessing alternative basing schemes during R-266.

57 Robert Zarate, "Introduction," in Zarate and Sokolski, *Nuclear Heuristics*, p. 15.

path through enemy defenses), and risks of enemy attack may be termed *location* costs.[58]And while R-266 took into account many locality effects, the analysis focused primarily on variations in SAC survivability occasioned by base location—first and foremost because base location decisions affected the strategy of employing bombers and, ultimately, nuclear deterrence of Soviet aggression.

Third, although R-266 was an extraordinarily successful and influential study, it did not employ complicated mathematical models featuring astronomically large numbers of machine computations or operations research techniques such as linear programming. Nor did R-266 attempt to determine a sharp optimum.[59] Instead, the aim was robustness (if not antifragility[60])—selecting basing schemes that would work reasonably well in widely divergent situations and perform satisfactorily in the event of a major catastrophe at an acceptable cost.

Wohlstetter's approach paid off, both for the Air Force customer and nuclear deterrence. After appraisal by an ad hoc Air Staff Group, R-266's principal recommendation to go with U.S.-based bombers operating intercontinentally with ground refueling overseas was accepted as Air Force policy. This basing policy reduced SAC vulnerability by cutting sharply the number of functions performed at overseas bases and stimulated research on strategic warning, airborne alert, and long-endurance bombers.[61]

The importance of selecting appropriate decision criteria cannot be overemphasized. As Hitch observed, big, broad problems have to be cut down to size by focusing on the variables that are especially important for the decision the analysis is intended to support and suppress the rest. "This takes good judgment; it amounts to no less than deciding in the designing of the analysis what is important and what is not."[62] In hindsight, Wohlstetter's *Selection and Use of Strategic Air Bases* did so, whereas Paxson's *Strategic Bombing Systems Analysis* did not. As Andrew Marshall later said, Wohlstetter's "crucial invention" was to recognize the importance of remaining open to the possibility that an analyst's initial intuitions about the appropriate decision criteria for an analysis might be wrong and need revision as the problem became better understood.[63]

58 Wohlstetter, Hoffman, Lutz, and Rowen, R-266, p. xii. Emphasis in the original.

59 Quade, "The Selection and Use of Strategic Air Bases: A Case History," pp. 62–63.

60 "Wind extinguishes a candle and energizes fire." Taleb, *Antifragile: Things that Gain from Disorder*, p. 3. The candle is fragile; the fire, which feeds on wind, is antifragile.

61 Quade, "The Selection and Use of Strategic Air Bases: A Case History," p. 62.

62 C. J. Hitch, "Analysis for Air Force Decisions," in Quade, R-387-PR, p. 15.

63 Andrew W. Marshall, "Strategy as a Profession in the Future Security Environment," in Zarate and Sokolski, *Nuclear Heuristics*, pp. 629–630.

Limited Resources and Strategic Choice

What Is the Best System? drew a distinction between "economic criteria" reflecting broad social utility on the one hand and narrow "engineering" criteria, such as the ratios of Soviet industrial targets destroyed to the various costs used in Paxson's *Strategic Bombing Systems Analysis*, on the other. The primary motivation for this distinction was the deep-seated appreciation of RAND economists of the fact that resources are always limited, always constraining our choices.

Early in Wohlstetter's basing study he became aware of a paper on basing issues by General Harold R. Maddux, who had originally asked RAND to do the basing study.[64] In his paper "Air Base Considerations Affecting an Expanding Air Force," Maddux suggested that basing decisions should be considered "from a purely military point of view not restricted by considerations of cost."[65] Costs—by which Maddux appears to have meant locality costs only—could be introduced later in the planning process. Wohlstetter's emphasis on SAC survivability obviously took a much more comprehensive approach to the basing problem by not just focusing on location costs, but by recognizing how location choices affected the employment and deterrent value of SAC's bomber force.

There is an intimate connection between resource limitations and choice, particularly in the case of higher-level or more strategic choices. Perhaps the most compelling expression of this connection can be found early in Hitch and McKean's *The Economics of Defense in the Nuclear Age*:

> Resource limitations are our starting point because in all problems of choice we strive to get the most out of what we have. To put it another way, we try to use the resources that are available to us so as to maximize what economists call "utility." Resources are always limited in comparison with our wants, always constraining our action. (If they did not, we could do everything, and there would be no problem of choosing preferred courses of action.)[66]

Or, as Hitch and McKean wrote in introducing defense as an economic problem, "Strategy and cost are as interdependent as the front and rear sights of a rifle."[67]

64 Kaplan, *The Wizards of Armageddon*, p. 103. In 1951 Maddux, who later retired from the Air Force as a major general, was the assistant for air bases in the Air Base Division within the Deputy Chief of Staff for Operations, Headquarters, Air Force. In January 1952 he became chief of the Air Base Division.

65 Wohlstetter, "Economic and Strategic Considerations in Air Base Location: A Preliminary Review," endnote 1.

66 Hitch and McKean, *The Economics of Defense in the Nuclear Age*, p. 23. Hitch and McKean introduced the idea that economics is "concerned with allocating resources—choosing doctrines and techniques—so as to get the most out of available resources" in the opening pages of *The Economics of Defense in the Nuclear Age*. They went on to say, "In our view the problem of combining limited quantities of missiles, crews, bases, and maintenance facilities to 'produce' a strategic air force that will maximize deterrence of enemy attack is just as much a problem in economics (although in some respects a harder one) as the problem of combining limited quantities of coke, iron ore, scrap, blast furnaces, and mill facilities to produce steel in such a way as to maximize profits. In both cases there is an objective, there are budgetary and other resource constraints, and there is a challenge to economize." Ibid., p. 2.

67 Ibid., p. 3.

This interdependence is explicit in the following conception of strategy. Whether in business, war, or other competitive situations, strategy is "fundamentally about identifying or creating asymmetric advantages that can be exploited to help achieve one's ultimate objectives despite resource and other constraints, most importantly the opposing efforts of adversaries or competitors and the inherent unpredictability of strategic outcomes."[68] As Ross Harrison has observed:

> This high level definition captures both the inward and outward faces of strategy. It creates an understanding that strategy involves overcoming challenges that come from both inside and outside an organization. The focus on resource constraints and the need to create asymmetric advantages or capabilities points us toward some of the internal challenges faced when preparing and executing strategy. The focus on exploiting advantages against adversaries and competitors, and the unpredictability of strategic outcomes addresses the external challenges strategy needs to overcome.[69]

In sum, strategic choice and strategy itself are consequences of constraints.

In this context, the notion of constraints should be understood as being vastly broader than, say, the direct monetary costs of a military capability, current operation, or a business enterprise. The availability of weapons-grade fissile material (uranium enriched to 90 percent ^{235}U or plutonium that is 93 percent ^{239}Pu) is a physical constraint on the production of nuclear weapons. At the other extreme are political and social constraints. As of this writing, the gross debt of the U.S. federal government was over $19 trillion and growing day by day. The failure of Congress and the two most recent administrations to address the federal government's annual deficits is surely a constraint on the nation's strategic choices, as is the American public's war-weariness in light of the results to date of the American interventions in Afghanistan and Iraq. In short, strategic choices are always and inevitably subject to a wide range of constraints.

68 Andrew F. Krepinevich and Barry D. Watts, *Regaining Strategic Competence* (Washington, DC: Center for Strategic and Budgetary Assessments, 2009), p. 19. This notion of strategy emerged from a September 2007 workshop sponsored by Andrew Marshall's Office of Net Assessment (ONA) and held at the Center for Strategic and Budgetary Assessments (CSBA) in Washington, DC. Besides CSBA participants, the seminar included Marshall and Andrew May from ONA, James Schlesinger, David Abshire, Lieutenant General (ret.) David Barno, and Richard Rumelt, who published *Good Strategy Bad Strategy: The Difference and Why It Matters* in 2011. Marshall's motivation for holding the seminar stemmed from a conversation with Peter Feaver, who was then on the National Security Council (NSC). Feaver's concern was the seeming lack of real interest within the NSC in the hard work of crafting good strategies.

69 Ross Harrison, *Strategic Thinking in 3D: A Guide for National Security, Foreign Policy, and Business Professionals* (Washington, DC: Potomac Books, 2010), p. 3. Harrison is Professor in the Practice of International Affairs and Chair of the International Commerce and Business concentration in the Master of Science in Foreign Service program at the Walsh School of Foreign Service at Georgetown University.

CHAPTER 3

The Rise of the West

[J]ust how did the West, initially so small and deficient in most natural resources, become able to compensate for what it lacked through superior military and naval power?

Geoffrey Parker, 1988[1]

[G]iven its scientific traditions and apparent technological superiority over Western Europe until the seventeenth century, how is it that China "failed to give rise to distinctly modern science" despite its "having been in many ways ahead of Europe for some fourteen previous centuries"?

Joseph Needham (quoted by Toby Huff), 1993[2]

How did the rich countries get so rich? Why are the poor countries so poor? Why did Europe ("the West") take the lead in changing the world?

David Landes, 1998[3]

Historians of democratic ages will tend to determinism: . . . "they wish to show that events could not have occurred otherwise." But they could have indeed. And therefore the very meaning of events that actually happened involves a consideration of what could have happened—a plausible and relevant potentiality.

John Lukacs, 2011[4]

1 Geoffrey Parker, *The Military Revolution: Military Innovation and the Rise of the West 1500–1800* (Cambridge, UK: Cambridge University Press, 1988 and 1996), p. 4.

2 Toby E. Huff, *The Rise of Early Modern Science: Islam, China and the West*, 2nd ed. (Cambridge, UK: Cambridge University Press, 2003), p. 34.

3 David S. Landes, *The Wealth and Poverty of Nations: Why Some Are So Rich and Some So Poor*, (New York and London: W. W. Norton, 1998), p. xxi.

4 John Lukacs, *The Future of History* (New Haven and London: Yale University Press, 2011), pp. 41–42.

How did the West—meaning Western Europe, the United States, Canada, Australia, Japan, and a few other places—rise to wealth and global dominance after 1500?[5] What were the criteria that explain why Western civilization alone became rich and dominant while the civilizations that were wealthier and more advanced in the early sixteenth century—notably in China and the Arabic-Islamic world—fell further and further behind after 1500? The main argument of this chapter is that the long temporal span and causal complexity of the West's rise make it impossible to identify with certainty the necessary and sufficient conditions that produced this outcome. The ascendance of the West took centuries, and the complex skein of causes and influences that gave rise to its global ascendance cannot be definitively untangled. In the case of very long-term historical changes as consequential as the rise of the West, it is simply not possible to unravel the vast array of interwoven circumstances that, at various times, helped to give birth to Western science, market capitalism, and industrialization.

Europe's escape from the poverty that had heretofore been the fate of the majority of human societies was a low probability, contingent outcome: things need not necessarily have turned out as they did. The West's rise, the Sinologist Etienne Balazs wrote in 1964, was a "miracle," "a freak of fortune, one of history's privileged occasions, in this case granted solely to that tiny promontory of Asia, Europe."[6] As David Landes later wrote, anyone who looked at the world around 1,000 A.D. would have put the probability of eventual European global dominance "somewhere around zero. Five hundred years later, it was getting close to one."[7] From the wider perspective of world history, the West's rise since 1500 has appeared to many "as a vast explosion" that, within three centuries, reversed "the millennial land-centered balance among Eurasian civilizations"—a view that downplays how gradual Western economic growth was prior to about 1750.[8]

The economic historian Eric Jones basically agrees with these characterizations of how miraculous the West's ascendance was. In 1981 he not only cited the passage from Balazs quoted above but titled his comprehensive effort to shed light on why very long-term economic growth occurred in the West after 1500 *The European Miracle*.[9] Insofar as Balazs and Jones were justified in using the term "miracle," a complete account of the critical criteria and the

5 Historians have somewhat arbitrarily chosen 1500 A.D. as the year that marks the end of the Middle Ages and "the advent of the modern era, in the world as well as in European history." William H. McNeill, *The Rise of the West: A History of the Human Community with a Retrospective Essay* (Chicago: University of Chicago Press, 1991), p. 565. However, Douglass North and others emphasize that the developments in Europe that eventually culminated in the rise of the West can be traced back to as early as the tenth century. Douglass North, *Understanding the Process of Economic Change* (Princeton and Oxford: Oxford University Press, 2005), pp. 127–128.

6 Etienne Balazs; H. M. Wright, trans., and Arthur F. Wright, ed., *Chinese Civilization and Bureaucracy: Variations on a Theme* (New Haven and London: Yale University Press, 1964), p. 23.

7 Landes, *The Wealth and Poverty of Nations*, p. 29.

8 McNeill, *The Rise of the West*, pp. 565, 567.

9 Eric Jones, *The European Miracle: Environments, Economics and Geopolitics in the History of Europe and Asia*, 3rd ed. (Cambridge, UK: Cambridge University Press, 2003), p. 202. The first edition of *The European Miracle* appeared in 1981, and a second appeared in 1987.

causal interactions that led to the West's economic, military and political emergence, while explaining why the other three Eurasian empires remained static or regressed, may simply not be possible. Among other reasons, many of the small, incremental advances in European technology and organization that preceded England's industrialization in the nineteenth century "will have gone unrecorded," and historians can only deal with those things that did not "steal unnoticed across the historical record."[10] In addition, there is the possibility—even likelihood—that Western growth had different causes in different historical periods, and more significant effects only became manifest decades after their causes had been forgotten.[11]

For these reasons, one should be cautious about calling the West's emergence *literally* a miracle. The occasional occurrence of highly improbable, hard to predict (in advance), yet consequential outcomes is not a rejection of causality. Such events are precisely what led Nassim Taleb to introduce the concept of a "Black Swan." Black Swan events are rare, carry extreme impact, and are only predictable after the fact by most observers.[12] They are not objective phenomena like rain or a car crash but simply events that were "not expected by a *particular* observer" or observers.[13] For most observers, England's industrialization from about 1750 to 1880, like the Great Depression of the 1930s, illustrate Taleb's concept of a Black Swan event.

But, again, Black Swans do not require us to abandon causality. Chaos theory (nonlinear dynamics) has demonstrated that, in the very long term, the states of even simple mathematical and physical systems that exhibit sensitive dependence on the tiniest differences in initial conditions can widely diverge, even though at every step each system's next state is rigidly determined through a feedback mechanism by its prior state.[14] Rather than abandoning causality, labeling the West's rise a "miracle" highlights the limits of both economic theory and historical knowledge. Possible causes of Western ascendance range from environmental circumstances (geography and the availability of resources) and population size to social

10 Ibid., pp. 65, 66.

11 Nathan Rosenberg and L. E. Birdzell, Jr., *How the West Grew Rich: The Economic Transformation of the Industrial World* (New York: Basic Books, 1986), pp. 8–9.

12 Nassim Nicholas Taleb, *The Black Swan: The Impact of the Highly Improbable*, 2nd ed. (New York: Random House, 2010), pp. xxi–xxii. Taleb has cited the Lebanese civil war that began in 1975, World War I, and the unexpected success of Yevgenia Nikolayevna Krasnova's book *A Story of Recursion* as examples of Black Swans. Ibid., pp. 7, 14, 23–25. He has also categorized the computer, the Internet, and the laser as Black Swans. All three were unplanned, unpredicted, initially underappreciated, and ultimately consequential. Ibid., p. 135.

13 Taleb, *The Black Swan*, p. 339.

14 See Edward N. Lorenz, *The Essence of Chaos* (Seattle: University of Washington, 1993), pp. 3–24; James A. Dewar, James J. Gillogly, and Mario L. Juncosa, *Non-Monotonicity, Chaos, and Combat Models*, R-3995-RC (Santa Monica, CA: the RAND Corporation, 1991), pp. 14–16; and Ian Stewart, *Does God Play Dice? The Mathematics of Chaos* (Oxford: Basil Blackwell, 1989), pp. 64–72. Deterministic chaos has been found in nonlinear circuits, hydrodynamics, acoustic and optical turbulence, solid-state physics, and biology. Hao Bai-Lin, *Chaos* (Singapore: World Scientific, 1984), pp. 67–73. Lorenz originally discovered chaos by reducing the number of decimal points in the input data for a weather model. Marcus du Sautoy, *What We Cannot Know: Explorations at the Edge of Knowledge* (London: 4th Estate, 2016), pp. 44–45. Henri Poincaré first discovered chaos in 1890, but its significance was not widely known until Lorenz rediscovered it in 1963.

institutions, governance, and ideas. The natural conceit is that if we could just identify the complete configuration of causes and influences behind the West's ascent, then the conditions that produced this long-term change would be laid bare before our eyes. But such complete knowledge of the past inexorably eludes us even in the information age; some determinants of the West's rise only acted at certain times. And one cannot always be certain about which factors were causes and which were effects. For example was the diffusion of technology in the late Middle Ages "a response to population growth or its cause"?[15] Compared to the Ottoman Empire in the Near East, the Mughal Empire in India, and the Ming and Manchu empires in China during the sixteenth, seventeenth, and eighteenth centuries, "Europe was an innovative, decentralized, yet stable, aberration."[16] Hence comes the impossibility of reaching absolute certainty, much less widespread scholarly consensus, on the principal drivers that produced the West's rise.

The present chapter will do four things to build a case for this conclusion. First, it will briefly recount some of the more obvious economic, military, and social changes manifested in the West's rise—particularly between the late Middle Ages and the nineteenth century. Here the challenge will be to avoid inferences about the causes of these changes. Second, even today "substantial disagreement" persists regarding the contours, causes, and timing of the West's divergence from the rest of the world.[17] Keeping constantly in mind that no single cause or gross statistic can definitively explain the divergence between the West and the "Rest," the contrast between Jared Diamond's narrow focus on environmental endowments and Deepak Lal's emphasis on European "cosmological" beliefs will be used to illustrate just how little agreement there is, even today, on how and why the West's ascendance happened.[18] Third, given this lack of consensus on the causes of the West's rise, some of the better known explanations will be described, starting with Max Weber's emphasis on the Protestant ethic but also including the later views of scholars such as William McNeill, Angus Maddison, Eric Jones, Douglass North, and Deirdre McCloskey. Lastly, this chapter will address the absence of either plausible theories or simple models of *very* long-term economic change due to the non-ergodic nature of reality.

What Happened in Europe but Nowhere Else

As a point of departure, the second, third, and fifth chapters of Nathan Rosenberg and L. E. Birdzell's *How the West Grew Rich* cover three of the outward manifestations of the West's rise without delving too far into the how and why behind them. These changes are:

15 Jones, *The European Miracle*, p. 49.

16 Ibid., p. 225.

17 Angus Maddison, *Growth and Interaction in the World Economy: The Roots of Modernity* (Washington, DC: American Enterprise Institute Press, 2005), p. 1.

18 Rosenberg and Birdzell, *How the West Grew Rich*, p.vi.

1. The decline of the manor system in rural Europe during the late Middle Ages accompanied by the growing political and economic autonomy of urban centers such as Genoa, Venice, Florence, the Hansa towns of the North Sea and the Baltic, and the Dutch cities.

2. The expansion of European trade and markets from the middle of the fifteenth century to the middle of the eighteenth century, as well as legal developments that made the growth of commerce and trade possible.

3. The industrialization of England from 1750 to 1880.

They provide an overview of the transformation that occurred in Europe after 1500.

1. Manorial Decline and Urban Autonomy. To start with the medieval manor system is to emphasize that Europe's industrialization during 1750 to 1880 did "not come out of the blue." Changes on this scale "are invariably well and long prepared."[19] In this case, much of that long preparation harks back to the emergence of urban trading centers in Flanders and northern Italy in the eleventh and twelfth centuries accompanied by "changes that fostered entrepreneurship and abrogated feudal constraints on the purchase and sale of property."[20]

After the fall of the Roman Empire, the lords or proprietors of land in Europe generally lived in fortified castles on their own estates or manors in the midst of their own tenants and dependents.[21] Manors (or agricultural *fiefs*) were land holdings that sovereigns (kings, independent princes, or other lords) granted to *seigneurs* (lords) to hold in fealty in return for various services, including military service.[22] Most people in the Middle Ages did not live in castles, towns, cathedrals, or inns, but were peasants (*villeins*), the majority of whom were bound to the land as serfs and owed their lords two or three days of labor a week as well as dues; their hereditary status was assumed at birth and excluded any right to select a more attractive occupation.[23] The manorial system of serfdom was "deeply oppressive," as indicated by the desperation of peasant uprisings, which were repeated over and over again no matter how bloodily suppressed.[24] "With status fixed by birth, upward mobility implied flight from the manors by some such device as escape to a town or joining a military unit." The manor system,

---

19 Landes, *The Wealth and Poverty of Nations*, p. 194.

20 Maddison, *Growth and Interaction in the World Economy*, p. 19.

21 Adam Smith; Edwin Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. I (Chicago: University of Chicago, 1976), Book II, Chapter 3, p. 420. Cannan's version of *The Wealth of Nations* was published in London by Methuen & Co. in 1904. It is also available online at http://www.econlib.org/library/Smith/smWN.html.

22 Rosenberg and Birdzell, *How the West Grew Rich*, p. 41. The manor system was "a hierarchy of land-tenure relationships" that paralleled the hierarchy of military relationships. Ibid. It has been calculated that 15 to 30 peasant families were needed to support one knightly household in the eleventh century. Harold J. Berman, *Law and Revolution: The Formation of the Western Legal Tradition* (Cambridge and London: Harvard University Press, 1983), pp. 302–303.

23 Rosenberg and Birdzell, *How the West Grew Rich*, p. 38; and *Understanding the Process of Economic Change*, p. 129.

24 Rosenberg and Birdzell, *How the West Grew Rich*, p. 45.

in combining economic and political authority in the lord of the manor, produced burdens that "impelled revolt and escape to the towns, to the Crusades and to the marauding armies."[25]

Medieval manors were inward-oriented and generally self-sufficient. Besides food they typically provided other services to the 10–20 percent of the population that lived outside the manor. These included mills to grind grain, ovens to bake bread, sawmills, smiths, and so forth.[26] While exchange within the manor was not mediated by money, transactions with those outside the manor—especially for services and luxury items such as arms, jewels, and higher quality clothing—did involve money.

The primary aim of the manor was to produce food. The precariousness of the food supply in medieval Europe is indicated by the small size of cities even late in the Middle Ages. Cologne, for example, supported perhaps 20,000 inhabitants in the fifteenth century, "though it was located at the meeting point of two branches of the Rhine and was thus far more favorably situated for the transportation of food than most medieval towns."[27] In contrast to the more intermittent rainfall and lighter soils of the Mediterranean rim, northwestern Europe was favored with abundant rainfall, thick forests, heavy soils suited to livestock, and the production of cereals (wheat, barley, rye, and oats).[28] While agricultural improvements diffused slowly in medieval Europe, they included the transition from oxen to horses aided by the horse-collar, the heavy plow with wheels, and the shift from the two-field to the three-field system of crop rotation.[29]

In much of medieval Europe, "A poor crop, even locally, meant hunger, malnutrition and greater vulnerability to disease, and a crop failure meant starvation."[30] In some regions of Europe, the Great Famine of 1315–1317, which was triggered by torrential rains in 1315 and again in 1317, killed a tenth of the population.[31] Famine was not the only calamity that affected Europeans in the fourteenth century. The arrival in 1347 of Genoese trading ships at the Sicilian port of Messina after journeying through the Black Sea introduced into Europe the

25 Ibid., p. 40.

26 Ibid., p 47.

27 Ibid., p. 40.

28 North, *Understanding the Process of Economic Change*, p. 128.

29 Ibid., p. 130.

30 Rosenberg and Birdzell, *How the West Grew Rich*, p. 40.

31 Amy Davidson, "The Next Great Famine," *The New Yorker*, January 11, 2016, p. 17; and William Chester Jordan, *The Great Famine: Northern Europe in the Early Fourteenth Century* (Princeton: Princeton University Press, 1996), pp. 23–24. Jordan estimates the population of the German-speaking lands, England, France (using modern boundaries), Denmark, Norway, and Finland to have been 39–42 million people in 1300 A.D. Ibid., p. 12. These are, of course, educated guesses and should be taken with several grains of salt.

Black Death in both its bubonic and pneumonic forms. The plague is estimated to have killed over 20 million people—almost one-third of the Europe's population.[32]

Medieval towns and cities, by contrast, were not self-sufficient. If the townspeople were to eat, they needed to import food from the countryside as well as the raw materials for urban craft guilds (wood, coal, peat for fuel, leather, iron, etc.).[33] These transactions were mediated by money, and before the Middle Ages ended, European sovereigns and lords began to realize that mercantile prosperity based on the buying and selling of goods by merchant guilds meant revenue, which was convertible into pleasure and power.[34] But mercantile activity could only flourish if buyers and sellers could negotiate prices, contracts could be enforced, and profits retained. Hence came the development of mercantile law between 1000 and 1200, A.D., which rendered commercial transactions "substantially more objective and less arbitrary, more precise and less loose."[35] After 1300 the continuing growth of trade fueled the further growth of towns and the settlement of merchants in them. As would be expected, geography and the high costs of land transport influenced which urban locations would prosper: for example, "at the head of a gulf (Bruges), where a road crossed a river (Maastricht), near the confluence of two rivers (Ghent), or at a breakpoint in transportation (Brussels)."[36]

Over time the economic incentives for granting greater autonomy—political as well as economic—to the townsmen (*bourgeois*) inevitably undermined manorial authority:

> The late history of the decline of the manor and the rise of the nation-state is, in part, a history of the decay of the lord's role as a political intermediary between the rank and file and the sovereign. There came to be substituted direct rights and obligations between kings and the inhabitants of the manors—a conversion to national citizenship from manorial citizenship and an expansion of access to the king's courts.[37]

The military revolutions of the late Middle Ages also served to erode the manorial system. The tenth to the sixteenth century in northwestern Europe "was a period of endless warfare at every level," and changes in military technology affected not only the conduct of warfare but increased the viable size of political units in order to bear the costs of military operations.[38] In the field, the battles of Crecy (1346), Poitiers (1356), Agincourt (1415), and countless

---

32 Landes, *The Wealth and Poverty of Nations*, p. 40. Landes estimates European deaths from bubonic and pneumonic plague at over one-third of the population. Jared Diamond puts the death toll at "one quarter of Europe's population between 1346 and 1352, with death tolls ranging up to 70 percent in some cities." Jared Diamond, *Guns, Germs, and Steel: The Fates of Human Societies* (New York and London: W. W. Norton, 1997), p. 202.

33 Rosenberg and Birdzell, *How the West Grew Rich*, p. 50.

34 Landes, *The Wealth and Poverty of Nations*, pp. 37, 363.

35 Berman, *Law and Revolution*, p. 341.

36 North, *Understanding the Process of Economic Change*, p. 130; and Smith; Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. I, Book III, Chapter 3, pp. 426–428.

37 Rosenberg and Birdzell, *How the West Grew Rich*, p. 43. Communes—"governments of merchants, by the merchants, and for the merchants"—appeared in Europe but nowhere else. Landes, *The Wealth and Poverty of Nations*, pp. 36–37.

38 North, *Understanding the Process of Economic Change*, pp. 131, 140–141.

lesser encounters during the Hundred Years' War (1337–1453) "confirmed that a charge by heavy cavalry could be stopped by archery volleys," and headlong charges and hand-to-hand combat gave way to exchanges of firepower.[39] The development of powerful siege guns in the mid-1400s ended the era of "vertical defense" provided by high castle walls.[40] But engineers eventually developed better defenses. In the last decades of the fifteenth century, a number of Italian states began building defensive fortifications known as the *trace italienne* that could withstand artillery bombardment.[41] Other military trends in the late Middle Ages included the "shift to professional armies" and the development of firearms. Facilitating the effective coordination of pikemen, bowmen, cavalry, and musketeers required skill and proficiency that exceeded what was available from serfs and even part-time soldiers.[42] The introduction of firearms, according to Adam Smith, "enhanced still further both the expense of exercising and disciplining any particular number of soldiers in time of peace, and that of employing them in time of war."[43] All told, these developments hastened the end of the manorial system and encouraged the emergence of nation-states in Europe.

39 Parker, *The Military Revolution*, p. 69.

40 Ibid., pp. 7–8; Jones, *The European Miracle*, p. 130.

41 Parker, *The Military Revolution*, p. 9. In the 1440s the architect and humanist Leon Battista Alberti proposed building defensive fortifications in uneven lines, like the teeth of a saw, to provide mutually supportive fields of fire. Ibid., pp. 8, 11. But Alberti's *De re aedificatoria*, written in 1440, remained unpublished until 1485. The main drawback to the *trace italienne* fortifications was their stunning cost. Ibid., p. 12.

42 Rosenberg and Birdzell, *How the West Grew Rich*, p. 63.

43 Smith; Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. II, Book V, Chapter 1, Part I, p. 230. "A musquet is a more expensive machine than a javelin or a bow and arrows; a cannon or a mortar than a balista or a catapulta. The powder which is spent . . . is lost irrecoverably, and occasions a very considerable expence. . . . The cannon and the mortar are not only much dearer, but much heavier machines than the balista or catapulta, and require a greater expence, not only to prepare them for the field, but to carry them to it. As the superiority of the modern artillery too over that of the ancients is very great, it has become much more difficult, and consequently much more expensive, to fortify a town so as to resist even for a few weeks the attack of that superior artillery. In modern times many different causes contribute to render the defence of the society more expensive." Ibid.

FIGURE 5: *TRACE ITALIENNE* FORTIFICATIONS, FORT MCHENRY, BALTIMORE HARBOR, MARYLAND

U.S. Government photos courtesy of the National Park Service and the FOMC social media team.

2. The Growth of European Commerce and Trade. Rosenberg and Birdzell put the population of Western Europe[44] around 1300 at more than 70 million inhabitants.[45] If the great famine of 1315–1317 shrunk Europe's population by ten percent, and the arrival of the Black Death in 1347 reduced it by more than a third, then the post-plague population of Europe was around 40–45 million inhabitants. It is generally believed that Europe's population did not return to the 1300 level—some 70 million inhabitants—until the early 1600s. Of course, these population figures are, at best, educated guesses, and Angus Maddison's gross domestic product (GDP) estimates prior to 1820 even more so.[46]

Nevertheless, it is apparent that population growth after the Black Death in the latter half of the fourteenth century, accompanied by the growth of cities and improvements in transportation, facilitated the expansion of interurban trade in Europe during the late Middle Ages.[47] The monarchs of Europe's relatively small political units (compared to China) increasingly acquiesced to the bulk trade of basic commodities (grain, meat, fruit, wine, etc.), choosing to profit from taxation rather than from pillaging or outright seizure.[48] Goods from Scandinavia flowed to the mouth of the Rhine, then to Switzerland and

44 Using modern political boundaries, Western Europe consists of the United Kingdom, France, Belgium, Switzerland, Austria, Denmark, Finland, Germany, Italy, the Netherlands, Norway, Sweden, Ireland, Greece, Portugal, and Spain.

45 Rosenberg and Birdzell, *How the West Grew Rich*, p. 75. Rosenberg and Birdzell's source is M. K. Bennett, *The World's Food* (New York: Harper, 1954). They also cite Colin Clark, although his population estimate for Europe is about 10 million greater than Bennett's. Colin Clark, *Population Growth and Land Use* (New York: Macmillan, 1977), p. 110.

46 For the medieval period, Maddison's European database only gives population, GDP, and GDP per capita estimates for 1000 and 1500 A.D. See Angus Maddison, "Historical Statistics of the World Economy: 1–2008 AD," *World Economics*, December 2010, available at http://www.worldeconomics.com/Data/MadisonHistoricalGDP/Madison%20Historical%20GDP%20Data.efp.

47 Rosenberg and Birdzell, *How the West Grew Rich*, p. 74.

48 Deepak Lal, *Unintended Consequences: The Impact of Factor Endowments, Culture, and Politics on Long-Run Economic Performance* (Cambridge, MA, and London: The MIT Press, 1998 and 2001), p. 79.

northeast Italy in return for produce from the Mediterranean; at the same time, Flanders gathered North Sea produce, moved it through northern and eastern France and thence to the Mediterranean and northwestern Italy.[49]

Perhaps the greatest stimulus to trade was the discovery and exploitation of new lands. In the fourteenth century, the Genovese and other Italians, along with the Portuguese and Basques, discovered the Near-Atlantic islands: the Azores, Madeiras, and Canaries.[50] The two more southerly of these islands—the Madeiras and Canaries—were superbly suited to the cultivation of sugar cane, and sugar plantations employing slave labor emerged in the quarter century before Christopher Columbus discovered the Americas. At its peak, the Portuguese crown was taking one-third of the gross from its sugar islands in fees, contracts, and taxes.[51]

European voyages of discovery into the Atlantic and down the west coast of Africa required better ships. This included the production of "carvels, longer and sleeker, rather than broad, cargo-bearing cogs; stern rudders; [and] a mix of square and lateen sails," as well as a marriage of shipbuilding techniques from the Mediterranean and the Atlantic.[52] Eventually the increasing size and sturdiness of European ships "made it possible to transform seagoing vessels into gun platforms for heavy cannon."[53] In the decisive battle for control of the Arabian Sea in 1509, the superior numbers of the Muslim fleet simply provided the Portuguese with more targets. The old tactics of sea fighting—ramming, grappling, and boarding—were almost useless against cannon fire, which had effective ranges of as much as 200 yards.[54]

The European age of discoveries that culminated with Columbus' discovery of the Americas (1492) and Vasco da Gama's opening of a sea route to India (1498) was initiated by the Portuguese Duke of Viseu, better known as Henry the Navigator (1394–1460). He established a marine research station that directed decades of inquiry into sailing and navigating on the high seas. Besides steady advances in shipbuilding, ocean sailing depended on the compass for direction, the astrolabe and cross staff for measuring the altitudes of celestial bodies (latitude in the case of the sun), and sand clocks for timing and speed estimation. Each Portuguese voyage "built on the ones before it; each time, they went a little farther;

49 Lal, *Unintended Consequences*, p. 80; and North, *Understanding the Process of Economic Change*, p. 130.

50 Landes, *The Wealth and Poverty of Nations*, p. 67.

51 Ibid., pp. 68–70.

52 Ibid., p. 87.

53 McNeill, *The Rise of the West*, p. 571.

54 Ibid., p. 571.

each time they noted their latitude, changed their maps, and left a marker of presence."[55] In 1488 Bartolomeu Dias brought back the coordinates of the southern tip of Africa. Ten years later Vasco da Gama landed in the Indian city of Calicut, opening a sea route to the subcontinent and gaining access to spices (essential to preserving meat) that circumvented the traditional routes across the Red Sea and the Spice Road that had been blocked by the territorial expansion of the new Islamic civilization.[56] Thus the Portuguese arguably led the way in linking "the Atlantic face of Europe with the shores of most of the earth."[57]

In the east, the Portuguese were soon joined by the Dutch and English, who fought harder and sailed better.[58] By the mid-sixteenth century, Portuguese power was on the wane, while that of the Spanish and Dutch was on the rise. In the 1560s the province of Holland alone possessed some 1,800 seagoing ships. In the 1580s Portugal was joined under the Spanish crown, and in 1605, "The Dutch took Amboina and drove the Portuguese from other bases in the Moluccas (Spice Islands)."[59]

In terms of new lands for trade, the greatest opening of new markets flowed from Christopher Columbus' discovery of the Americas. The Portuguese discovered Brazil in 1500 and began settling the country in the 1530s. Driven by the Spanish ruthlessness and a prodigious lust for gold and silver, Hermán Cortés destroyed the Aztec state during the years 1519–1521, and Francisco Pizarro crushed the Incan empire during the years 1531–1535. With the establishment of New France along the St. Lawrence River by the French (1530s), Jamestown in Virginia by the English (1607), and New York by the Dutch (1614), the Europeans began colonizing North America while the Spanish and Portuguese focused on Mexico and South America. From 1500 to 1700 the "formidable combination of European warlikeness, naval technique, and comparatively high levels of resistance to disease transformed the cultural balance of the world," although the impact of this shift had yet to be seriously felt in Moslem, Hindu, and Chinese lands.[60] Quantitatively, Western progress in shipping and navigation produced a twentyfold growth in world trade from 1500 to 1820.[61]

55 Landes, *The Wealth and Poverty of Nations*, p. 87; and McNeill, *The Rise of the West*, p. 570. In the Atlantic the decision of Portuguese navigators to sail west, almost to the coast of South America, before turning east was "the most inventive and audacious" step of all, showing tremendous confidence in their ability to navigate on the high seas. In da Gama's case his fleet was out of sight of land for 96 days and traveled about 4,500 miles between landfalls compared with Columbus' 2,600 miles and 36 days out of sight of land. Ibid., pp. 570, 573.

56 Lal, *Unintended Consequences*, p. 71; and Landes, *The Wealth and Poverty of Nations*, p. 130.

57 McNeill, *The Rise of the West*, p. 565.

58 Landes, *The Wealth and Poverty of Nations*, p. 131.

59 Ibid., p. 127.

60 McNeill, *The Rise of the West*, p. 574. "According to one calculation, a central Mexican population of about 11 million in 1519 had shrunk to a mere 2.5 million by 1600, and [disease-driven] depopulation continued until a low point of about 1.5 million was reached in 1650." Ibid., p. 601.

61 Maddison, *Growth and Interaction in the World Economy*, p. 2.

That the West would escape from poverty in the nineteenth and twentieth centuries was not at all obvious as late as 1700. Rosenberg and Birdzell have rightly stressed how gradual the West's year-by-year economic growth was from the early Middle Ages to as late as 1820. [62] Similarly, Landes has pointed out that for the century from 1760–1860, which spans the core of Britain's industrial revolution, England's per capita GDP—admittedly a dubious measure of individual wealth—only grew 1 or 2 percent a year when deflated for population growth.[63]

One of the reasons why the root causes of the West's ascension have been so controversial—especially among economists insistent on hard, quantitative data—stems from the low average annual growth rates in the West from one period to the next up to 1870. Maddison's best estimates of annual average compound growth rates for the West and the Rest starting in 1000 A.D. are shown in Table 5.[64] Of course since the concept of GDP only dates back to the 1940s, Maddison's earlier GDP estimates are reconstructions rather than statistics based on contemporaneous data collection.[65] And the further one looks back in time the greater the uncertainty in Maddison's estimates of per-capita GDP, population, and GDP growth rates for the West versus the Rest.

62 Rosenberg and Birdzell, *How the West Grew Rich*, p. 6.

63 Landes, *The Wealth and Poverty of Nations*, p. 196.

64 Suffice it to say that GDP estimates before the mid-1930s are retrospective reconstructions. The definitions of GDP we use today "date back to two seismic events in modern history, the Great Depression of the 1930s and World War II." Diane Coyle, *GDP: A Brief but Affectionate History* (Princeton and Oxford: Princeton University Press, 2014), p. 12. The first American Gross National Product (GNP) statistics were published in 1942. Ibid., p. 15. GDP = consumption + investment + government spending + (exports – imports) within a country's borders by its nationals and foreigners calculated over the course of one year. GNP is the total value of goods and services produced by all the nationals of a country whether within or outside the country. GDP is discussed in more depth in Chapter 4.

65 Coyle, *GDP: A Brief but Affectionate History*, p. 4.

TABLE 5: GROWTH RATES OF PER-CAPITA GDP, POPULATION, AND GDP, 1000–1973 (ANNUAL AVERAGE COMPOUND GROWTH RATES) [66]

	1000–1500	1500–1820	1820–1870	1870–1913	1913–1950	1950–1973
Per-capita Gross Domestic Product (GDP)						
West	0.13%	0.14%	1.06%	1.57%	1.17%	3.72%
Rest	0.04%	0.02%	0.06%	0.82%	0.65%	2.83%
West/Rest	3.25	7.00	17.67	1.91	1.80	1.31
Population						
West	0.15%	0.27%	1.86%	1.07%	0.78%	1.05%
Rest	0.09%	0.27%	0.29%	0.72%	0.98%	2.15%
West/Rest	1.67	1.00	6.41	1.49	0.80	0.49
Gross Domestic Product						
West	0.27%	0.41%	1.93%	2.66%	1.96%	4.81%
Rest	0.13%	0.29%	0.39%	1.54%	1.63%	5.04%
West/Rest	2.08	1.41	4.95	1.73	1.20	0.95

Nevertheless, while Maddison's growth rates for the West's GDP and per-capita GDP do not exceed 2 percent per year through 1870, they are impressive, if not spectacular, compared to those of the rest of the world up to that time. From 1820 to 1870 the West's average annual GDP growth was nearly five times that of the Rest, and the West's per-capita GDP growth rate was 17 times greater. As Maddison summarized the quantitative divergence between the West and the Rest in 2005:

> Real per-capita income of the West increased by nearly threefold between 1000 and 1820 and twentyfold from 1820 to 2001. In the rest of the world, income rose much more slowly, by a third from 1000 to 1820 and sixfold since then. The West had 52 percent of the world GDP in 2001, but only 14 percent of the world population. Average income was about $22,500 (in 1990 purchasing power). The Rest, by contrast, with 86 percent of the world population, had an average income of less than $3,400.[67]

66 Maddison, *Growth and Interaction in the World Economy*, p. 10. These average annual compound growth rates were calculated using the GDP and population data in Maddison's Table 2 and the formula $(V_{t_n}/V_{t_0})^{(1/(t_n-t_0))} - 1$. Ibid., p. 7. Table 6 below displays a selection of Maddison's GDP and population data. The West includes Western Europe, the United States, Canada, Australia, New Zealand, and Japan. The Rest consists of Asia (excluding Japan), Latin America, Eastern Europe and the former Soviet Union, and Africa. Maddison adds: "Western Europe's economic ascension began about the year 1000, continued to 1820, and accelerated thereafter. Western Europe caught up with Chinese income levels in the fourteenth century. By 1950, European per capita levels were ten times higher than the Chinese levels. The experience of China in the past half century, and particularly since the 1978 economic reforms, shows clearly that divergence is not inexorable. China's economic resurrection has involved a very significant element of catch-up." Ibid., p. 8.

67 Maddison, *Growth and Interaction in the World Economy*, p. 8.

In addition, Maddison and others are adamant that the foundations of the West's rise preceded Britain's economic takeoff by centuries. By the seventeenth century, "European statesmen had firmly grasped the notion that a flourishing commerce and industry was a direct asset to the state."[68] By then European commerce was nourished and protected by the "law of the merchant" (*lex mercatoria*), which Deepak Lal traces back to the papal revolution of 1075 when Pope Gregory VII declared the papacy's supremacy over the entire church and secular matters.[69] From the eleventh to thirteenth centuries *lex mercatoria* developed the institutional protections for trade and commerce that, in the late eighteenth century, established key infrastructure for the first industrial revolution in England. This infrastructure included the development of negotiable bills of exchange and promissory notes; the invention of the mortgage of movables, bankruptcy law, bills of lading, and other transportation documents; the development of joint venture stock companies in which the liability of each investor was limited to the amount of his investment; the floating of public loans secured by bonds; and the development of deposit banking.[70] To this list, one could also add the invention of double-entry bookkeeping, which Rosenberg and Birdzell describe as "an actualization of the profit-seeking firm as a truly autonomous . . . unit."[71] The upshot was a "social and legal revolution that radically transformed the nature of medieval society, in fact laying the foundations of modern society and civilization."[72]

Besides being sheltered and encouraged by the emergence of *lex mercatoria*, European voyages of discovery found new lands that grew into new markets for European trade and commerce over time. As Adam Smith famously explained, market expansion permitted greater division of labor, and the division of labor, in turn, was the source of the "greatest improvement" in labor's productive powers.[73] In the opening chapter of *An Inquiry into the Nature and Causes of the Wealth of Nations* Smith cited the manufacture of pins to illustrate the productive power of dividing labor into sequences of simpler tasks. All told, Smith argued that the manufacture of pins could be broken into about 18 distinct operations: drawing out the wire, straightening it, cutting the straightened wire, sharpening one end of the cut wires, grinding their tops to receive heads, making heads (which required two or three distinct operations), soldering on the heads, etc. Smith had seen some pin manufactories[74] where ten men divided the 18 steps among them. True, some workers still had to perform two or three of the distinct operations. But Smith estimated that, with the division of labor among ten men,

68 McNeilll, *The Rise of the West*, pp. 583, 652.

69 Lal, *Unintended Consequences*, p. 81.

70 Berman, *Law and Revolution*, pp. 349–350.

71 Rosenberg and Birdzell, *How the West Grew Rich*, p. 127.

72 Huff, *The Rise of Early Modern Science*, p. 317.

73 Smith; Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. I, Book I, Chapter 1, p. 7.

74 The factory system that the English began introducing in the mid-eighteenth century not only brought workers under supervision that integrated the division of labor but used a central source of power such as coal-fired steam engines. Without central power, one has a *manufactory* rather than a *factory*. Landes, *The Wealth and Poverty of Nations*, p. 186.

they could produce upwards of 48,000 pins of a middling size in a day, whereas ten men each working separately and independently to perform all 18 steps would have been unlikely to produce even 20 pins for a total of 200.[75] Thus ten workers in a pin factory with their labor divided into a sequence of simpler tasks were some 240 times more productive than ten men individually producing pins without the division of their labor.

As for the relationship between markets and the division of labor, Smith argued that the division of labor was limited by the extent of the market.[76] The larger the market, the greater the division of labor that was likely to occur. The discovery and colonization of the Americas by Europeans inevitably expanded old markets and created new ones. The "Columbian" biological exchange between the New World and the Old provided the Europeans with "new peoples and animals, but above all, new plants—some nutritive (maize [Indian corn], cocoa [cacao], potato, sweet potato), some addictive and harmful (tobacco, coca), some industrially useful (new hardwoods, rubber)."[77] The new foods altered diets on a global scale.

> Corn, for example became a staple of Italian (polenta) and Balkan (mamaliga) cuisine; while potatoes became the main starch of Europe north of the Alps and the Pyrenees, even replacing bread in some places (Ireland, Flanders).[78]

In return, the Europeans brought to the New World "new plants—sugar, cereals, and new fauna—the horse, horned cattle, sheep, and new breeds of dog."[79]

Rosenberg and Birdzell argue that the expansion of European trade and commerce from the late Middle Ages to 1750 was a natural response to the comparative advantage of regional specialization inside and outside Europe. During this period,

> Western Europe created an active merchant class and marked out an area where it could trade with enough freedom to leave it a world of opportunities readily to hand. It created also a network of markets, commercial and financial relationships, and economic institutions which needed only to be scaled upward in order to process a much larger volume of trade; no substantial generic modifications were necessary for them to exploit the technology of the Industrial Revolution.[80]

75 Smith; Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. I, Book I, Chapter 1, pp. 8–9.

76 "When the market is very small, no person can have any encouragement to dedicate himself entirely to one employment, for want of the power to exchange all that surplus part of the produce of his own labour, which is over and above his own consumption, for such parts of the produce of other men's labour as he has occasion for." Ibid., Vol. I, Book I, Chapter 3, p. 21.

77 Landes, *The Wealth and Poverty of Nations*, p. 169.

78 Ibid.

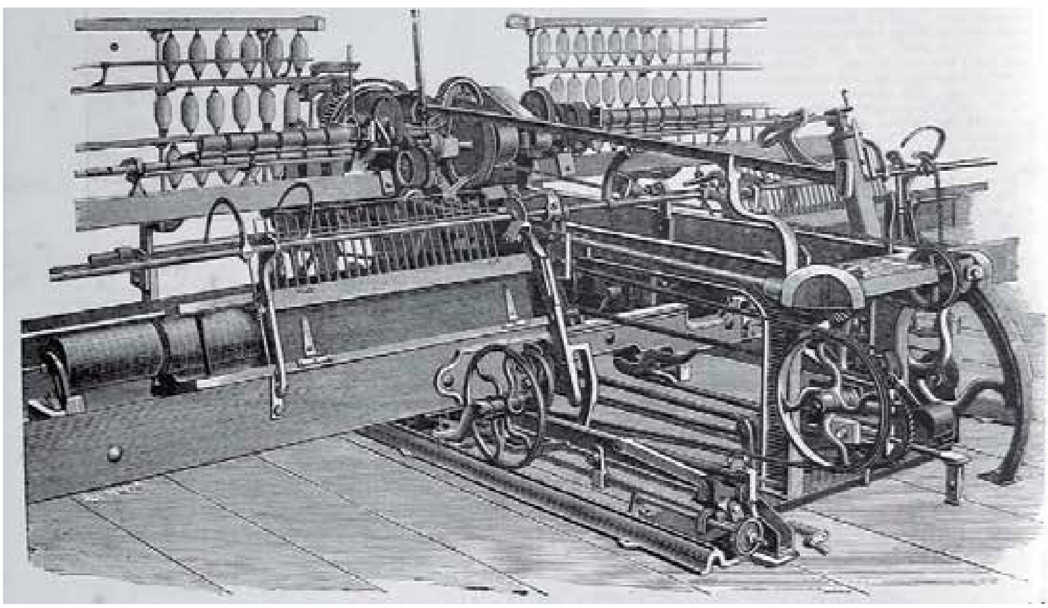
79 Far worse for the Amerindians, who died in great numbers, were the pathogens that the Europeans and the black slaves they brought with them introduced into the New World: smallpox, measles, yellow fever, malaria, diphtheria, typhus, and tuberculosis. Ibid.

80 Rosenberg and Birdzell, *How the West Grew Rich*, p. 109.

3. The industrialization of England from 1750 to 1880. This first industrial revolution stemmed from a series of inventions that transformed British manufacturing by substituting machine labor for human labor, exploiting water and (increasingly) coal-fired steam power in factories, and substituting iron and steel for wood in fabricating machinery.[81] A major component of this revolution was the industrialization of the British textile industry. In the early 1600s, the English East India Company had begun importing Indian cotton fabrics to Europe. By the end of the seventeenth century, Indian cottons had transformed the dress of Europe and its overseas offshoots.

> Lighter and cheaper than woolens, more decorative (by dyeing or printing), easier to clean and change, cotton was made for a new wide world. Even in cold climes, the suitability of cotton for underwear transformed the standards of cleanliness, comfort, and health. In the American plantations, it answered perfectly; as some Jamaica traders put it (1704): ". . . the said island being situated in a hot climate, much of the clothing of the inhabitants is stained callicoes, which being light and cheap and capable of often washing contributes very much to the keeping them clean and in health." Here was a commodity of such broad and elastic demand that it could [and did] drive an industrial revolution.[82]

FIGURE 6: ROBERTS' SELF-ADJUSTING SPINNING MULE



Public Domain image.

81 Landes, *The Wealth and Poverty of Nations*, p. 186; Rosenberg and Birdzell, *How the West Grew Rich*, p. 146.

82 Landes, *The Wealth and Poverty of Nations*, p. 154.

Of course, the rise of the British textile industry was by no means the whole story of the first industrial revolution. Entwined with the mechanization of spinning and weaving was the development of coal-fired steam engines. The first device to use steam to create a vacuum was patented in England by Thomas Savery in 1698; the first steam engine proper (with a piston) was built by Thomas Newcomen in 1705.[83] But Newcomen's engine was best used to pump water out of coal mines. It was another 60 years before James Watt, in 1768, invented an engine with a separate condenser whose fuel efficiency made steam power profitable away from the mines, thereby providing the motive power for England's new industrial centers, and "It took another fifteen years to adapt the machine to rotary motion, so that it could drive the wheels of industry."[84]

FIGURE 7: A WATT DOUBLE-ACTING STEAM ENGINE BUILT IN 1859

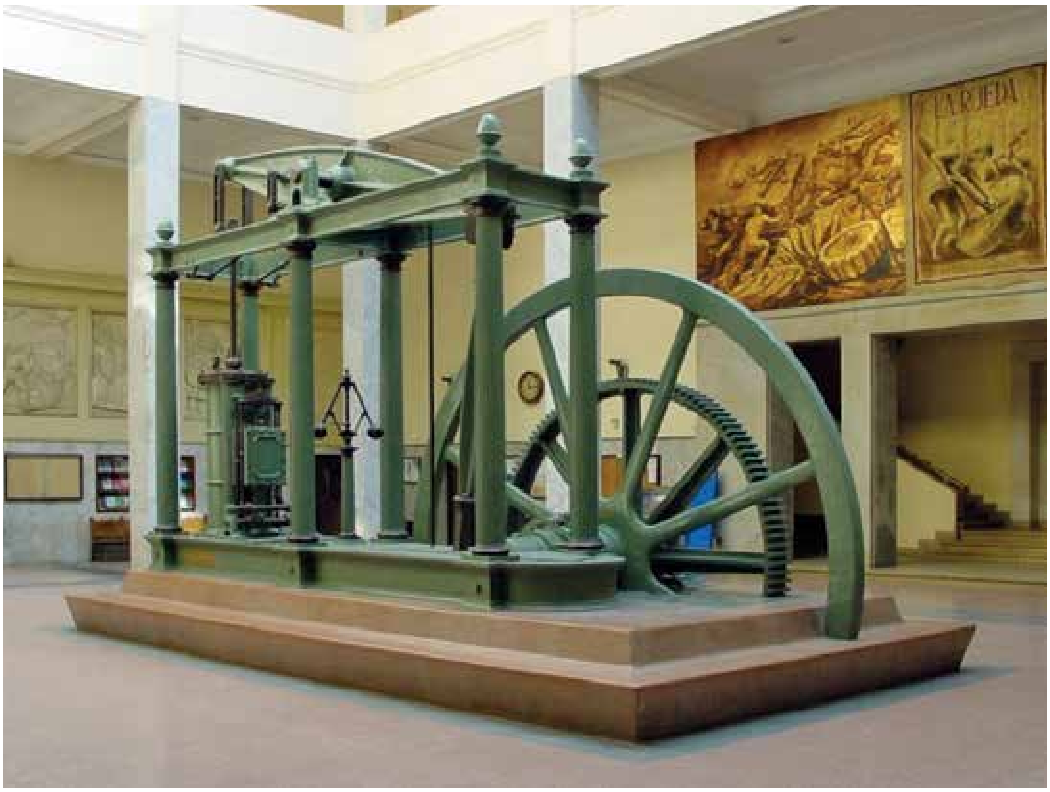


Photo credit: Nicolás Pérez under the Creative Commons license.

83 Ibid., p. 187.

84 Ibid., p. 188. For Maddison's data on the growth in the dollar value of machinery and equipment per capita in Britain, the United States, and Japan from 1820 to 1973, see the discussion following Table 6 below.

Along the way engineers and mechanics had to solve numerous manufacturing and maintenance problems. For example, "Making cylinders of smooth and circular cross section, so that the piston would run tight and not leak air to the vacuum side, required care, patience, and ingenuity."[85] To this end, in 1774 the ironmaster John Wilkinson patented a new method of boring cylinders out of solid castings. Thus steam power for machines such as Roberts' self-adjusting spinning jenny grew out of both British inventiveness and the emergence in the eighteenth century of an iron and steel industry.[86] The industrialization of England therefore involved the interaction of numerous components: maritime trade; domestic transportation (canals and railroads); coal; textiles; iron and steel production; machinery (which included steam engines, spinning and weaving equipment, clocks, guns, carriages, and machines to build other machines); ceramics; and furniture.[87]

Factor Endowments versus Institutional and Cultural Drivers

To this point, the aim has been to describe the West's gradual escape from poverty after 1500, which culminated in England's industrial revolution. To the greatest extent possible, description has sought to avoid explanations of why these changes occurred in Europe when they did and not elsewhere. In particular, no mention has been made so far of plausible explanatory developments such as the European Renaissance in art and literature that began in Italy; the Protestant Reformation initiated by Martin Luther in 1517 that eventually broke the dominance of the Roman Catholic church; the rise of modern science epitomized by the discoveries and experiments of Galileo Galilei (1564–1642); and the Westphalia treaties of 1648 that ended the Thirty Years' War between Catholic and Protestant states, thereby entrenching the idea of the sovereign nation-state in Europe.[88] But separating description from explanation is easier said than done. Landes, for example, has argued that the Europeans' cultivation of invention—the constant search for and pleasure in the new and better—had roots in the

85 Ibid.

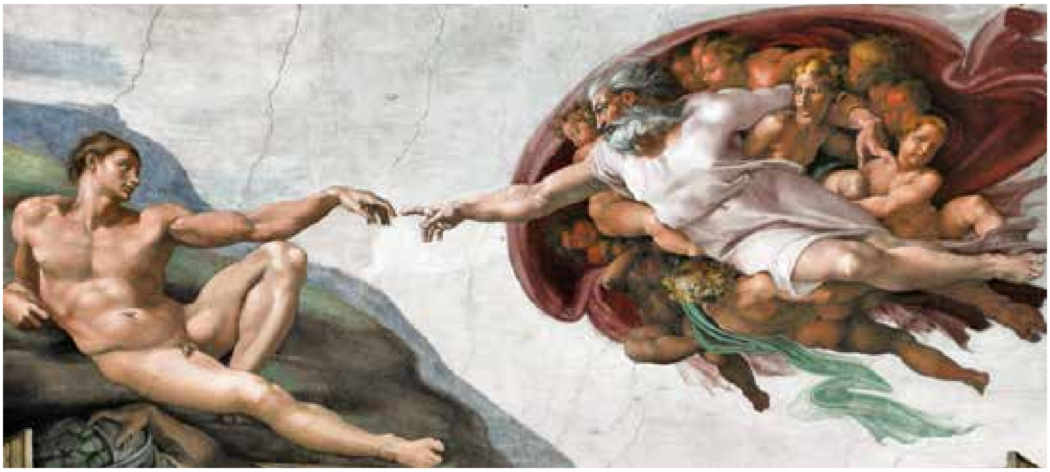
86 Rosenberg and Birdzell, *How the West Grew Rich*, pp. 156–158; Landes, *The Wealth and Poverty of Nations*, pp. 189–190.

87 Landes, *The Wealth and Poverty of Nations*, p. 191.

88 "The Reformation without Luther, the Jesuits without Loyola, or modern science without Galileo are really unthinkable. It is a defect of historical essays such as this that the unique individual career and the strategic moment of personal thought or action can be easily obliterated by the weight of inexact generalizations." McNeill, *The Rise of the West*, p. 599.

Europe of the Middle Ages and that this cultivation underlays the cumulative improvements in the productivity of labor at the heart of the European miracle.[89]

FIGURE 8: MICHELANGELO'S "CREATION OF ADAM," SISTINE CHAPEL CEILING (1508–1512)



Nevertheless, even today not everyone accepts the view that cultural factors such as the inventiveness of European societies could have been anything other than a proximate, derivative cause of Europe's rise that can be fully explained on the basis of strictly material environmental factors such as geography and climate. Nearly two and a half centuries after Adam Smith's seminal inquiry into the wealth of nations, there is scant consensus among scholars regarding the causal criteria that produced the European miracle and explain why the West alone escaped from poverty. Jared Diamond and Deepak Lal illustrate just how divergent views of the ultimate explanations for the West's ascension can be.

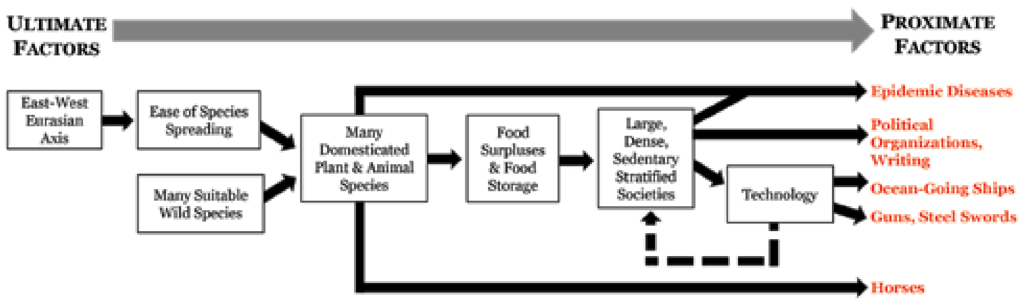
89 Landes, *The Wealth and Poverty of Nations*, pp. 45, 50, 58. The Chinese invented the compass, the mechanical clock, block printing, paper, and gunpowder well ahead of the Europeans. For example, the Chinese invented a large mechanical clock in the eleventh century but, apparently, without the escapement mechanisms that the Europeans later added. Huff, *The Rise of Early Modern Science*, p. 322. The Europeans, on the other hand, invented eyeglasses (extending the productive careers of craftsmen), the telescope, the microscope, and the mechanical clock, which led to a new (digital) reckoning of time that was not replicated in China or the Middle East. Ibid., pp. 49, 51; and Rosenberg and Birdzell, *How the West Grew Rich*, p. 149. Johannes Gutenberg's development in the 1430s of movable metal type for his printing press facilitated an explosion of information and ideas in Europe, greatly aided by the introduction of water-powered paper mills. The Europeans also learned in the sixteenth century how to "corn" their gunpowder, leading to more powerful explosives.

"Why," Diamond asked in 1997, "did wealth and power become distributed as they are now, rather than in some other way?"[90] In answer, he maintained that,

> . . . the striking differences between the long-term histories of peoples of different continents have been due not to innate differences in the peoples themselves but to differences in their environments. I expect that if the populations of Aboriginal Australia and Eurasia could have been interchanged during the Late Pleistocene, the original Aboriginal Australians would now be the ones occupying most of the Americas and Australia, as well as Eurasia, while the original Aboriginal Eurasians would be the ones now reduced to downtrodden population fragments in Australia.[91]

Or, stated more succinctly and precisely, "History followed different courses for different peoples because of differences among peoples' environments, not because of biological differences among peoples themselves."[92]

FIGURE 9: FACTORS UNDERLYING THE BROADEST PATTERN OF WORLD HISTORY[93]



As Figure 9 indicates, Diamond's account of the West's rise to wealth and global military dominance is unabashedly based on material factors: climate, geography, and the availability of plants and animals that could be domesticated (that is, factor endowments). His argument for this conclusion proceeds as follows. The earliest human civilization emerged soon after 4,000 B.C. in the valleys of the lower Tigris and Euphrates rivers; this Sumerian civilization consisted of agricultural settlements that relied on priests to organize large-scale water engineering (irrigation) to increase food production.[94] We know, Diamond maintains, that the Sumerians produced sufficient food to give rise to dense, sedentary human populations; stored food surpluses; and maintained the wherewithal to feed non-farming specialists. We also

90 "For instance, why weren't Native Americans, Africans, and Aboriginal Australians the ones who decimated, subjugated, or exterminated Europeans and Asians?" Diamond, *Guns, Germs, and Steel*, p. 15.

91 Ibid., p. 405.

92 Ibid., *Guns, Germs, and Steel*, p. 25.

93 Ibid., *Guns, Germs, and Steel*, p. 87. For two alternative views of the primary causes that drove the West's rise see Figures 10 and 11.

94 McNeill, *The Rise of the West*, pp. 29–34, 36. The domestication of plants and animals in the Fertile Crescent began around 8500–8000 B.C. Diamond, *Guns, Germs, and Steel*, p. 362.

know that there were other fertile and highly suitable areas of the world—notably California, temperate Australia, Europe, New Guinea, and the Argentine pampas—"where [agricultural] food production never developed indigenously at all."[95] Diamond's core argument—which he also extends to animal husbandry—is that factor endowments alone can adequately explain why indigenous food production sufficient to produce the first civilization that initially arose in the Fertile Crescent and failed to arise in many other places. There is no reason to appeal to the cultures or institutions of the various peoples involved to explain the divergent outcomes between the West and the Rest.[96]

Diamond's argument is fairly persuasive with respect to the emergence of early agricultural societies. It is not, however, very persuasive regarding later developments in the West, such as the rise of early modern science. Factor endowments may suffice for surplus food production, early animal husbandry, and population growth. But they are woefully inadequate to explain the development of science in the West. To cite one example among many, there seems to be considerable merit to Toby Huff's argument that Galilean-Newtonian science "was in fact the outcome of a unique combination of cultural and institutional factors that are, in essence, nonscientific."[97]

Part of the problem with Diamond's unabashedly materialistic line of argument is his use of ultimate versus proximate causes.[98] The logical problem with using these concepts as he does is the presumption that any causal sequence must have a first term from which all other causes are derivative.[99] This "fallacy of absolute priority" makes it easy to ignore necessary conditions over and above the factor endowments handed out by accidents of geography and environment. At least some of the factors that influenced Britain's industrial revolution were emergent phenomena. Here again, modern science comes to mind. One is hard-pressed to see the causal linkage between geographical and environmental factors and the emphasis of Galileo and other early European scientists on evidence and experiment. Timur Kuran has made much the same point regarding the role of Islamic contract law in holding Islam back from modernity. Referring explicitly to Diamond, he observed in 2011 that a "stable climate cannot explain the development of Islamic contract law in the early Islamic centuries, or the adoption, a millennium later, of the French commercial code."[100]

95 Ibid., p. 153.

96 Ibid., pp. 138–143, 147, 150, 152–156.

97 Huff, *The Rise of Early Modern Science*, p. 10.

98 In the behavioral sciences, ultimate explanations are concerned with whether a particular trait or behavior is (or is not) selected; proximate explanations are concerned with the mechanisms that underpin the trait or behavior—that is, how it works. Thomas C. Scott-Phillips, Thomas E. Dickens, and Stuart A. West, "Evolutionary Theory and the Ultimate-Proximate Distinction in the Human Behavioral Sciences," *Psychological Science* 6, no. 1, 2011, p. 38.

99 Timur Kuran, *The Long Divergence: How Islamic Law Held Back the Middle East* (Princeton and Oxford: Princeton University Press, 2011), p. 15.

100 "In the proverbial relationship between chicken and egg, there is no absolute starting point. Each . . . [causal factor] serves as both source and product, making the relationship bidirectional." Ibid.

One other aspect of Diamond's position warrants mention. He has a more personal, if not politically correct, motivation for insisting on a strictly materialistic explanation of the West's rise. Diamond maintains that attributing any part of the West's rise to European cultures or institutions is a "racist conclusion," which leads him to categorize as derivative causes any and all other conditions that may have in fact contributed to or influenced the divergence between the West and the Rest.[101]

Deepak Lal, like Huff and others, differs from Diamond's strict materialism. Lal does not deny that the material endowments (geography, climate, and other resources) of China, India, Central Asia, and Europe were major determinants of each civilization's economic growth from medieval times to the mid-1950s.[102] But he does deny that these material factors alone are sufficient to explain what he terms Europe's "Promethean" growth and the absence of comparable growth in the Chinese, Islamic, and Hindu civilizations:

> That merely technological and scientific developments were insufficient to deliver the Industrial Revolution is borne out by the failure of Sung China [960–1279 A.D.] to do so, although it had these ingredients. It was a "package" of cosmological beliefs, political decentralization, and the application of the "inquisitive Greek Spirit" that uniquely led to Promethean growth and the ascendancy of the West.
>
> The failure of both the Abbasids [who ruled the third Islamic caliphate from 750 to 1258 A.D.] and the Sung to put together such a package is plausibly ascribed to the relative closing of their peoples' minds by the victories of their respective orthodox clergies.[103]

Lal characterizes a society's cosmological beliefs as "those related to understanding the world around us and humankind's place in it which determine how people view their lives—its purpose, meaning, and relationship to others."[104] His key point is that the West's cosmological beliefs were decidedly individualist as compared with the communal outlooks that have long persisted in the Chinese, Islamic, and Hindu civilizations.[105] The crux of his argument is that the *package* of cosmological beliefs, political decentralization, and inquisitive scientific spirit that emerged in early modern Europe did not arise in the other great civilizations that by 1500 possessed the resources and technology to escape from poverty.

---

101 Diamond, *Guns, Germs, and Steel*, p. 300.

102 Lal, *Unintended Consequences*, p. 173. For a particularly compelling account of the influence exerted on different societies by "nature's inequalities"—geography, climate, diseases, insects, pests, and water—see Landes, *The Wealth and Poverty of Nations*, pp. 3–16. But Landes is also adamant that these endowment or environmental factors should not be seen as destiny. Ibid., p. 15.

103 Lal, *Unintended Consequences*, p. 173. Later, during the Ming dynasty (1368–1644), the central Chinese state "attempted to prohibit all overseas trade." David S. Landes, "Why Europe and the West? Why Not China?" *Journal of Economic Perspectives* 20, no. 2, Spring 2006, p. 7.

104 Lal, *Unintended Consequences*, pp. 7–8. "In most fields, agriculture being the chief exception, Chinese technology stopped progressing well before the point at which a lack of scientific knowledge had become a serious obstacle." Landes, "Why Europe and the West? Why Not China?" p. 6.

105 Lal, *Unintended Consequences*, p. 12.

Take the case of China. As Joseph Needham (1900–1995) concluded on the basis of his seminal research into Chinese science and civilization, between the first century B.C. and the fifteenth century A.D. "Chinese civilization was much *more* efficient . . . in applying human natural knowledge to practical human needs" than was European civilization.[106] But in the fifteenth century, China's rulers turned inward and closed their minds to further progress down the individualistic path that led to Britain's industrial revolution and, in the Western view, modernity.

A striking illustration of the closing of the Chinese mind was the central government's decision in the fifteenth century to cease the naval expeditions initiated by the eunuch admiral Zheng He. These voyages explored the waters of Indonesia, the Indian Ocean, and the seas of Asia's east coasts. Between 1405 and 1430 seven armadas of as many as 62 junks carrying 37,000 soldiers sailed as far west as Zanzibar in east Africa and as far to the northeast as Kamchatka.[107] But these expeditions did not lead to an age of discovery comparable to what occurred in Europe. By 1436 the new emperor in Beijing sided with the mandarins (bureaucrats) who distrusted commerce as well as eunuchs like Zheng. The upshot was an imperial decision not just to cease maritime exploration but "to erase the memory of what had gone before lest later generations be tempted to renew the folly" of these extravagant voyages.[108] By contrast, in the fractious competition between the states of a decentralized Europe, such an arbitrary decision by a single central ruler was not possible. The European states-system "was very different from the unitary empires of Asia and the Middle East."[109] Recall that Christopher Columbus' efforts to secure funding for his expedition to what turned out to be the New World rather than the Orient was rejected by the Portuguese crown (twice), Genoa and Venice, and the English crown before the King Ferdinand II and Queen Isabella I of Spain finally decided in 1489 to support his expedition.

At one extreme, Diamond maintains that the great divergence in wealth and power between the West and the Rest after 1500 can be completely explained by differences in the initial material endowments of different societies. At the other extreme, Lal maintains that cultural and political factors—including the individualist or cosmological beliefs that were unique to European civilization—are necessary to explain Europe's ascendance. What does this explanatory gulf between their respective positions suggest about analytic criteria for explaining the European miracle? The most obvious implication is that the analytic criteria each of

106 The Needham quote is from Rosenberg and Birdzell, *How the West Grew Rich*, p. 87. Their source is Needham's "Science and Society in East and West" in Joseph Needham, *The Grand Titration* (London: George Allen & Unwin, 1969), p. 190. Needham was a biochemist and self-taught Sinologist. Starting in 1954, Cambridge University Press published seven volumes of Needham's monumental *Science and Civilization in China*. Wang Ling assisted or collaborated on the first four volumes; others contributed to later ones. As of 2016 Needham's line of research continues at the Needham Research Institute under Christopher Cullen.

107 Jones, *The European Miracle*, p. 203.

108 Landes, *The Wealth and Poverty of Nations*, pp. 95–97. The resumption of maritime voyages by China was proposed in 1480 but was promptly quashed. Jones, *The European Miracle*, p. 205.

109 Jones, *The European Miracle*, p. 245.

them chose to emphasize drives and determines their assessments as to the why and how of the West's rise. Diamond, bothered by the possibility of drawing racist conclusions, tries to hang everything on material factors (climate, geography, and the availability of plants and animals whose domestication led to agriculture). Lai, on the other hand, highlights the necessary role played by non-material factors. In reflecting on nothing more than the role played by early modern science in the West's rise, one cannot help but suspect—strongly suspect—that Diamond's extreme materialism leaves out non-material factors that were necessary to England's industrial revolution. Similarly, neither *lex mercatoria* nor Westphalian nation-states emerged in the Chinese and Islamic civilizations. One could argue that Europe's decentralized states system was encouraged by Europe's geography, but *lex mercatoria*, the market economy, and Galilean-Newtonian science all appear better understood as emergent phenomena. They may have been influenced by material factors but not completely determined by them.

The other ramification that emerges from the extreme positions on the West's ascendancy held by Diamond and Lal is that the choice of what analytic criteria to emphasize in explaining historical outcomes inevitably contains a substantial and irreducibly subjective element.[110] One reason, as explained in the discussion of the past versus history at the end of this chapter, is that we cannot hope to possess more than a tiny fraction of the vast amount of information that would be needed to reconstruct an exhaustive, top-to-bottom, causal account of the West's rise. Indeed, Douglass North (1920–2015) has articulated persuasive reasons for thinking that we do not currently possess a dynamic theory of very long-term economic change and suggests that such a theory may well be impossible as squaring the circle with ruler and compass constructions.[111]

Other Prominent Explanations: Weber and McNeill to Maddison, Jones, and North

Before delving deeper into these limits to human understanding, we need to review some of the other analytic criteria that scholars from Max Weber and William McNeill to Angus Maddison, Eric Jones, and Douglass North have highlighted to explain the divergence between the West and the Rest. The various criteria these writers have emphasized generally fall somewhere between the extreme positions espoused by Diamond and Lal. As we shall see, though, the dominant explanations to which Weber, McNeill, and most scholars have given pride of place do suggest one point of general agreement: that purely material explanations of the West's rise will not suffice. Ideas, institutions, politics, and cultures have also played a role.

This observation provides further evidence that while the past is fixed, history is not, and the long-term future is unpredictable. Expanding on a point Ralph Bennett made in trying to

110 Schlesinger, P-3464, p. 4.

111 North, *Understanding the Process of Economic Change*, pp. vii, 69.

explain why the Allies won the Battle of the Atlantic, in none of the privileged explanations reviewed in this chapter—including Diamond's—is there any suggestion that a single causal factor could adequately explain a historical outcome as unexpected, complex, and protracted over centuries as Europe's ascendance. "Monocausal explanations will not work."[112] Instead, the broad explanatory thrust in each case is best understood as drawing attention to analytic criteria that appear to have been important (if not necessary) conditions that, in conjunction with many others, produced the rise of the West after 1500.

One of the first to hazard an overarching explanation for the European miracle was Karl Emil Maximilian "Max" Weber (1864–1920). His seminal essay *The Protestant Ethic and the Spirit of Capitalism* originally appeared as a two-part article in 1904–1905. These two articles immediately provoked critical debate in which Weber himself participated actively, and which has not subsided to this day.[113]

The bulk of Weber's two articles is devoted to articulating what he sees as the ethos or spirit (*Geist*) of "modern" or "rationalized" Western capitalism. Early in Part I, he explicitly rejected the notion that capitalism is even remotely identical with unlimited greed for gain or acquisition by force. The article defines "a capitalistic economic action as one which rests on the expectation of profit by the utilization of opportunities for exchange, that is on (formally) peaceful chances of profit."[114] A man imbued with the ethics of modern capitalism does not pursue ever greater economic success in order to enjoy what material wealth can buy. Rather he "exists for the sake of his business, instead of the reverse."[115] "He gets nothing out of his wealth for himself, except the irrational sense of having done his job well."[116] In this sense the spirit of modern capitalism—to make money as long as one is able—"takes on the character of an ethically coloured maxim for the conduct of life."[117]

Weber believed that the Protestant Reformation initiated by Luther (1483–1546) and continued by John Calvin (1509–1564) turned this ethical dimension into a calling. *The Protestant Ethic and the Spirit of Capitalism* highlights the four principal forms of ascetic Protestantism (Calvinism, Pietism, Methodism, and the sect that grew out of the Baptist movement) that, in Weber's assessment, were causal underpinnings of the European miracle.[118] One of the most important results of these influences—especially on Luther's part—was

112 Landes, *The Wealth and Poverty of Nations*, p. 517.

113 Anthony Giddens, "Introduction," in Max Weber, Talcott Parsons, trans., *The Protestant Ethic and the Spirit of Capitalism* (London and New York: Routledge, Taylor & Francis e-Library, 2005), p. vii. This translation is taken from a revised version of Weber's essay that appeared in his *Collected Essays on the Sociology of Religion* in 1920–1921. In 1976 Giddens summarized five of the more telling criticisms of *The Protestant Ethic and the Spirit of Capitalism*. Ibid., pp. xxi–xxiii.

114 Weber, *The Protestant Ethic and the Spirit of Capitalism*, p. xxxii.

115 Ibid., p. 32.

116 Ibid., p. 33.

117 Ibid., p. 17.

118 Ibid., p. 53.

"the moral justification of worldly activity," specifically the conviction that the fulfillment of one's duty in world affairs was the highest form which the moral activity of the individual could assume."[119] The unceasing pursuit and achievement of worldly success came to be seen by ascetic Protestantism as a sign of the individual having been "predestined unto everlasting life" by God even "before the foundation of the world was laid."[120]

In Weber's analysis, ascetic Protestantism was a significant causal contributor to the ascendance of the West relative to other civilizations after 1500. While capitalism existed in China, India, and Babylon, as well as during the classical world and the Middle Ages, the Protestant "ethos was lacking."[121] Weber's bottom line is that Protestantism—specifically its Calvinist branches—had a part in and "promoted the rise of modern capitalism."[122]

Again, this thesis has been severely criticized ever since it first appeared in 1904–1905 in the pages of *Archiv für Sozialwissenschaft und Sozialpolitik,* of which Weber was one of the editors. Anthony Giddens' assessment in 1976 was that the "elements of Weber's analysis that are most definitely called into question . . . are: the distinctiveness of the notion of the 'calling' in Lutheranism; the supposed lack of 'affinity' between Catholicism and regularized entrepreneurial activity; and, the very centre-point of the thesis, the degree to which Calvinist ethics actually served to dignify the accumulation of wealth in the manner suggested by Weber."[123]

One need not reach a final judgment on the validity of these criticisms to see Weber's deeper claim: that ideas embedded in specific cultures can "become effective forces in history."[124] Japan's Meiji restoration began around 1867–1868. It ended the three-century reign of the Tokugawa Shogunate and resulted in the oligarchic council-based "imperial" rule in Japan that persisted to the end of World War II. One effect of the Meiji restoration was to accelerate the country's industrialization in order to catch up with the West economically and militarily. Japan's industrialization was greatly aided by a Japanese version of Weber's Protestant work ethic. As Landes has observed, the "Japanese home worker was able and willing to put up with hours of grinding, monotonous labor that would have sent the most docile English spinner or pinmaker into spasms of rebellion."[125] Whereas a strong, ascetic work ethic was absent in China, the Middle East, and India, it was manifest in late nineteenth-century Japan as well as in Protestant Europe during the sixteenth century.

William McNeill's highly regarded *The Rise of the West: A History of the Human Community* aspires, as the subtitle suggests, to a world history of humankind since the dawn of

119 Ibid., pp. 40–41.

120 Ibid., p. 57. The quoted passage is from the authoritative Westminster Confession of Faith of 1647.

121 Ibid., p. 17.

122 Landes, *The Wealth and Poverty of Nations*, p. 174,

123 Giddens, "Introduction," in Weber, *The Protestant Ethic and the Spirit of Capitalism*, pp. xxiii–xxiv.

124 Weber, *The Protestant Ethic and the Spirit of Capitalism*, p. 48.

125 Landes, *The Wealth and Poverty of Nations*, p. 383.

civilization.[126] His account starts with the rise of the first civilization in Mesopotamia shortly after 4000 B.C. Consequently, it is not until page 569 (of over 800) that McNeill begins to discuss the West's ascendance after 1500. In fact, he does not focus on the West's escape from poverty through industrialization until page 726.

McNeil emphasizes a very different driver behind the West's ascendance than Diamond, Lal, or Weber. First published in 1963, the 1991 edition of *The Rise of the West* contains a retrospective essay that highlights "the principal factor behind historically significant social change" as "contact with strangers possessing new and unfamiliar skills."[127] Cultural diffusion, as developed among American anthropologists in the 1930s, is "the main motor of social change" involving civilizations. Of course, this viewpoint begs the question of who really belongs to a given civilization, and McNeill conceded in his 1991 essay that he had been a bit "fuzzy" on this point.[128] In hindsight, he felt that he had paid inadequate attention to the emergence of world systems, notably the interpenetration between Greece and the Orient from 500 to 146 B.C. as well as the nineteenth-century "contacts and collisions" between the industrializing West led by Great Britain and the major Asian civilizations.[129]

Notwithstanding McNeill's firm commitment in 1991 to cultural diffusion as the motor of significant social change, there is little evidence of this mechanism in his explanation of the causes of the "First or British Phase" of the industrial revolution.[130] McNeill is clearly on solid ground in observing that there were two technical enablers of Britain's industrialization in the nineteenth century: coal and iron.[131] And, as Diamond would insist, both were accidents of geography. Still, there was surely more to what happened than the presence of accidental material endowments. The use of coal for steam power and the replacement of wooden with iron machinery occurred on such a large scale that from 1820 to 1900 the per-capita GDP in the United Kingdom went from $1,706 to $4,492 (in 1990 International Geary-Khamis dollars), even though the population grew from 21 to 41 million.[132] This unprecedented increase in wealth led McNeill to opine that the magnitude of this first industrial

---

126 This is no less than Arnold Toynbee judged *The Rise of the West* to be "the most lucid presentation of world history in narrative form" that he knew of. The book assumes "a decent familiarity with Western history" and emphasizes "matters which have been usually underrated, while passing over more familiar ground with a casual reference or with no mention whatever." McNeill, *The Rise of the West*, p. xxxi.

127 Ibid., p. xvi.

128 Ibid., pp. xx–xxi.

129 Ibid., pp. 277–280, 282, 726–730.

130 The primary actors in the second phase of the West's industrialization (from roughly 1870 to 1914) were Germany and the United States. This second phase centered on "the rise of the electrical, chemical, petroleum, and light metal industries, with their derivative products like automobiles, radios, airplanes, and synthetic textiles." Ibid., p. 737.

131 Ibid., p. 732.

132 Maddison, "Historical Statistics of the World Economy: 1–2008 AD."

revolution was comparable to that of the Neolithic transition from predation to agriculture and animal husbandry.[133]

What, then, were the vital and necessary ingredients of this momentous change? After dismissing as mere possibilities such "metaphysical" factors as "the Protestant ethic" and the "capitalistic spirit," McNeill's suggestion in 1963 was to point to a "pervasive looseness in British society."

> This looseness had been inherited from parliament's angry dismantlement of the aspiring absolutism of the early Stewarts in the mid-seventeenth century. But, between 1756 and 1815, the parliamentary archaism of the English Old Regime was rather thoroughly shaken up by the strains and exigencies of war with the French. Abrupt changes in overseas trading patterns, depending on the momentary correlation of naval forces; equally abrupt changes in government orders for the manifold sinews of a growingly complicated and massive warfare; and drastic fluctuations in money supply and price levels all weakened traditional resistance to economic innovation. Under these conditions, the financial rewards accessible to a lucky or skillful promoter were so conspicuous as to induce hundreds of other ambitious or greedy individuals to attempt to duplicate his feat.[134]

These are quite plausible reasons for the greater looseness of British society, if not also for British inventiveness and the adaptability of the nation's labor force. By the late eighteenth century the British people enjoyed "increasing freedom and security,"[135] and industry itself was "perfectly secure"—or at least "as free or freer than in any other part of Europe."[136] These various conditions go a good way toward explaining why the first industrial revolution occurred when and where it did. On the other hand, one is hard-pressed to discern in them any signs of cultural diffusion between civilizations. By the nineteenth century, the British were not encountering new and unusual skills: they were inventing them.

Perhaps sensing the disconnect between cultural diffusion as the motor of significant social change and the causes of the first industrial revolution, McNeill in effect punted. He suggested a more tightly regulated British society would have prevented the rapid transformation of Britain's industrial plant. But he immediately went on to add:

> No catalogue of general conditions can ever entirely account for a massive change in human affairs. Individual persons, inspired by diverse motives, made decisions and acted in ways that in sum began to transform the life of England. Pride in skillful workmanship and the wish to achieve respect among fellow mechanics may have been as potent a stimulus to invention in some cases as the desire for money was in others.[137]

133 McNeill, *The Rise of the West*, p. 732.

134 Ibid., pp. 733–734.

135 Landes, *The Wealth and Poverty of Nations*, pp. 219–220.

136 Smith; Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. II, Book IV, Chapter 5, p. 50.

137 McNeill, *The Rise of the West*, p. 734.

Thus, the suggestions McNeill marshals, under the rubric of "the looseness of British society" as causes for the first industrial revolution, are just that—suggestions—and they cannot be more than a partial explanation that omits other necessary conditions.

Angus Maddison (1926–2010) was a British economist who described himself as a lover of figures. He took a quantitative, macroeconomic approach to the measurement and analysis of economic growth and development. Regarding the various qualitative accounts of the sources of West's economic rise, he wrote in 2005 that his "judgment of the nature of long-run Western performance and its exceptional character" did "not differ greatly from that of Adam Smith (in 1776), David Landes (in 1998), or [John] McNeill and [William] McNeil (in 2003)."[138] His more fundamental aim, however, was to make a more systematic use of quantitative evidence in understanding how and why the West escaped from poverty. In 1995 he published GDP estimates for 56 countries dating back to 1820, and in 2001 he estimated the world output in 1 A.D. as $105.4 billion in 1990 prices.[139] Table 6 shows the historical GDP and population data Maddison had compiled.

TABLE 6: GDP, POPULATION, AND PER-CAPITA GDP FOR THE WEST AND THE REST, 1000–1973[140]

	1000	1500	1820	1870	1913	1950	1973
Gross Domestic Product (billions of 1990 international dollars)							
West	$14.1	$53.0	$194.4	$504.5	$1,556.9	$3,193.0	$9,398.0
Rest	$102.7	$195.3	$501.0	$608.2	$1,175.2	$2,137.0	$6,626.0
Population (millions)							
West	35	75	175	268	424	565	718
Rest	233	363	867	1,004	1,367	1,959	3,198
Per-capita GDP (1990 international dollars)							
West	$405	$702	$1,109	$1,882	$3,672	$5,649	$13,082
Rest	$441	$538	$578	$606	$860	$1,091	$2,072

138 Maddison, *Growth and Interaction in the World Economy*, p. 3. John R. McNeill is the son of William H. McNeill. Maddison's reference is to their 2003 book *The Human Web: A Bird's-eye View of World History*.

139 "Maddison Counting," *The Economist*, May 29, 2010, available at http://www.economist.com/node/16004937.

140 Maddison, *Growth and Interaction in the World Economy*, p. 7. Table 6 contains the data on which the growth rates in Table 5 were calculated.

As noted in discussing Table 5, Maddison's earlier GDP estimates—especially prior to 1820—should be viewed as educated guesses at best. Much the same can probably be said of his earlier population estimates. Maddison freely conceded that the further back in time he went, the more his numbers depended on historical clues and conjectures. But for good or ill, economists demanded numbers and Maddison provided them.[141]

Nonetheless, Maddison believed that his numbers provided strong evidence for the view that Europe's economic takeoff after 1820 had historical roots that went back centuries. He highlighted four major intellectual and institutional changes in the West before 1820 that had no counterparts elsewhere in the world:

1. Recognition of human capacity to transform the forces of nature through rational investigation and experimentation, a development that included the founding of European universities and the appearance and exploitation of Gutenberg's printing press;

2. The emergence of important urban trading centers in Flanders and northern Italy that fostered entrepreneurship and abrogated feudal constraints on the purchase and sale of property together with the emergence of nondiscretionary legal systems that protected property rights;

3. The adoption of Christianity as a state religion in 380 A.D., which made marriage monogamous, limited inheritance entitlements to close family members, and channeled large amounts of money to the church; and

    1. The emergence of a system of nation-states that had significant trading relations and relatively easy intellectual interchange in spite of linguistic differences.[142]

These developments were, in Maddison's view, the roots of modernity.

There were a number of things associated with these changes that could be quantified. Foremost in Maddison's assessment was "the explosive growth in the stock of machinery and equipment per head" in the United Kingdom, the United States, and Japan.[143] Measured in terms of their monetary value (in 1990 international dollars), from 1820 to 1973 the gross stock of machinery and equipment per capita in the United Kingdom, the United States, and Japan rose by multiples of 67, 129, and 68, respectively.[144] During this same period,

141 "Maddison Counting," *The Economist*. As Landes has noted, not all numbers are plausible, and "Economic historians, especially the self-styled new economic historians, have a visceral love for numbers." David S. Landes, "What Room for Accident in History?" *Economic History Review* 47, no. 4, 1994, p. 648. Landes is not alone in criticizing economists for their "mathiness," "model mania," and tendency to be blinkered by their own theories. See "A Less Dismal Science," *The Economist*, "The World If" supplement, July 16, 2016, p. 12, available at http://worldif.economist.com/article/12134/less-dismal-science.

142 Maddison, *Growth and Interaction in the World Economy*, pp. 18–20.

143 Ibid., p. 14.

144 Ibid., p. 13.

human capital increased by factors of eight in the United States and Japan and nearly six in the United Kingdom (based on the average number of years of education weighted by the level attained).[145] And the value of British and American per-capita exports grew over thirty-fold from 1820 to 1973.[146] Together with advances in transportation, growing energy use, and decreases in the hours worked per head of population, the West's economic ascendance after 1820 is readily documentable using these sorts of quantitative metrics. Granted some of these indices could be attributed to factor endowments such as coal and iron, but the first three developments prior to 1820 emphasized by Maddison seem difficult to attribute to geography or environment even through a long causal chain.

Eric Jones is far more cautious than Weber, or even Maddison, in identifying the factors that he sees as the strongest candidates for explaining the European miracle. In the introduction to the second edition of *The European Miracle* (1987), he stated that there is "not the least reason to suppose that any single factor or relationship had an all-powerful effect on [Europe's] economic development."[147] In an afterword to the third edition (2003), however, he went on to say that while no one factor can be isolated, "The nation-state and the states-system [that arose in Europe] are strong candidates for the first team. Nation–states helped to expand markets and make their operation more certain, establishing unified coinages, railway networks focused on capital cities, national education systems, and national languages."[148] Thus he gives pride of place to European political institutions. The European "balance of power did in practice prevent the rise of monolithic empires [as in China and under the Ottomans] and curb some of the waste of wars among the nation–states. . . . The pith of the European miracle lies somewhere here, in politics rather than economics."[149]

Douglass North's explanation of the West's rise has much in common with those of Jones, Rosenberg and Birdzell, and Landes.

> The rise of the Western world was the ascendancy of a relatively backward part of the world to world hegemony between the tenth and the eighteenth centuries. It involved economic, political, and military changes as well as a shift in the focus of institutional change—from dealing with the uncertainties associated with the physical environment to dealing with those of the increasingly complex human environment. The rise of science and its systematic application to solving problems not only of economic scarcity but also of human well-being was still in its infancy by the eighteenth century but the foundation had been laid for the revolutionary developments of the next two centuries. . . . [At the core of how and why this all came about was] the complex interplay among beliefs, institutions, and other factors such as geography, military technology, and evolving competitive conditions that influenced the process of change.[150]

145 Ibid., pp. 13, 14.

146 Ibid., p. 13.

147 Jones, *The European Miracle*, p. xviii.

148 Ibid., p. 244.

149 Ibid., p. 125.

150 North, *Understanding the Process of Economic Change*, p. 127.

For North, gradual changes in European institutions and organizations laid the foundations for the West's ascendance after 1500. "Institutions," he has written, "are the rules of the game, organizations are the players; it is the interaction between the two that shapes institutional change."[151]

> The rise of the Western world was ultimately a consequence of the kinds of skills and knowledge (not only "productive knowledge" but notably knowledge about military technology) that were deemed valuable to the political and economic organizations of the medieval Western world.[152]

In North's assessment, the developments that influenced the emergence of the specific skills and knowledge that underlay the European miracle range from "the gradual rise of the nation state" to Weber's Protestant ethic, the establishment of "more secure and efficient property rights," and, echoing Lal, the deepening of individualistic belief structures.[153]

True, North only attributes a complementary role to the individualistic belief structures that came to be most pronounced in the Netherlands and England, but these belief structures contrast sharply with the more communal belief structures that prevailed in France and Spain. The English Parliament's 1628 Petition of Right "established for all Englishmen a set of rights protected by law."[154] By the seventeenth century, Englishmen viewed themselves as "freeborn" rather than as serfs, villeins, or other dependent individuals. Combined with the particular conditions that existed in the Netherlands and England, these more individualistic belief structures "induced behavior conducive to both economic growth and to the evolution of freedoms."[155]

What do these other prominent explanations of the West's rise suggest? Landes has summed up the situation as well as anyone. As he wrote at the end of *The Wealth and Poverty of Nations*, if we have learned anything from the history of economic development it is "that culture makes all the difference. (Here Max Weber was right on.)"[156] One need not fully accept Weber's emphasis on the Protestant work ethic to see the folly of denying that culture played a role—if not an important and necessary one—in the West's rise. But as Landes has also noted, culture in the sense of the inner values and attitudes that guide a population has a sulfuric odor of race and inheritance that has frightened many scholars, especially in the current era of political correctness. Anti-Western interpretations of the European miracle have stressed the West's sheer luck or outright misconduct. Explanations highlighting Western misconduct include income and wealth inequality, the exploitation of workers, colonialism, imperialism, and slavery.

151 Ibid., p. 59.

152 Ibid., p. 63.

153 Ibid., pp. 133, 134, 137, 141.

154 Ibid., p. 145.

155 Ibid., p. 145.

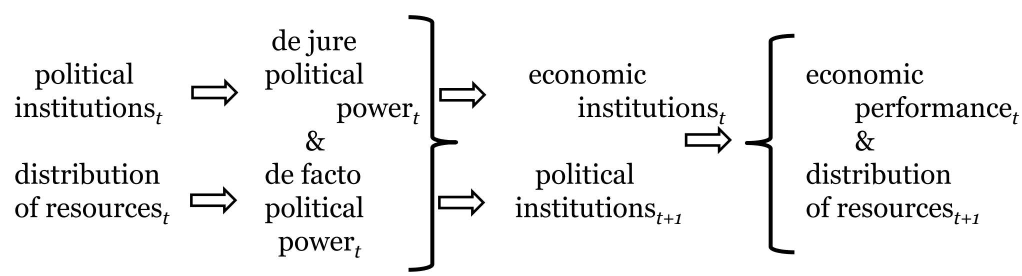
156 Landes, *The Wealth and Poverty of Nations*, p. 516.

Grains of truth can be found in some of these accusations. It may well be, for example, that income and wealth inequality was a necessary condition for the entrepreneurship and risk-taking that promoted Western industrialization after 1750. The main benefit of anti-Western explanations for the European miracle may be to provide a check on Western hubris. However, as Rosenberg and Birdzell's critiques of these alternative explanations for capitalist success suggest, there is not a great deal to be said in favor of "their adequacy as explanations for Western economic growth."[157] And except for pure good luck, anti-Western criteria such as the exploitation of workers, colonialism, and slavery only add to the growing list of inherently cultural explanations.

Very Long-term Growth, the Unchanging Past versus Historical Narratives, and the Unpredictable Future

Ever since Adam Smith argued that the accumulation of capital through savings was key to economic growth, the sheer breadth and variety of the diverse criteria that have been advanced over the years to explain the West's ascendancy put paid to the possibility of a monocausal explanation of Europe's escape from poverty through industrialization. Nonetheless, in 2004 Daron Acemoglu, Simon Johnson, and James Robinson argued that institutions are "the fundamental cause of income differences and long-run [economic] growth."[158] Figure 10 shows a schematic (and simplistic) representation of their framework for explaining economic growth (at times *t* and *t+1*). Intuitively their emphasis on economic and political institutions makes more sense historically than Diamond's view that they arose entirely from geographical and environmental factors. Besides, present-day policy decisions can affect institutions but not the original distribution of geographical and environmental factors.

FIGURE 10: ACEMOGLU, JOHNSON, AND ROBINSON'S FRAMEWORK FOR EXPLAINING ECONOMIC GROWTH[159]



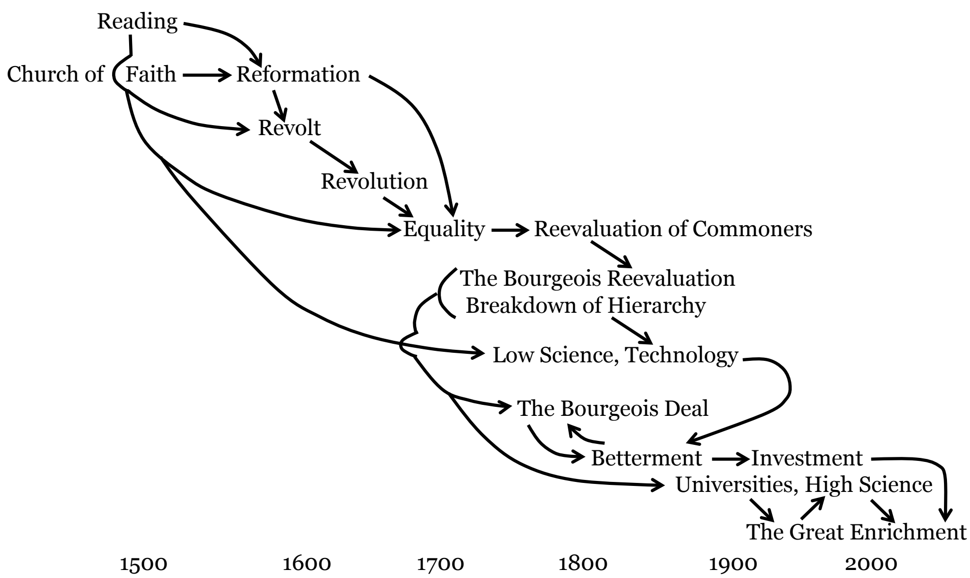
157 Rosenberg and Birdzell, *How the West Grew Rich*, p. 13.

158 Daron Acemoglu, Simon Johnson, and James Robinson, *Institutions as the Fundamental Cause of Long-run Growth*, Working Paper 10481 (Cambridge, MA: National Bureau of Economic Research, May 2004), p. 29.

159 Ibid., pp. 5–6.

Yet as tempting as it may be to conclude that institutions are the elixir of economic growth, there is no consensus among historians on this view of how the West escaped from poverty. Deirdre McCloskey argues in her third book on bourgeois inequality that Smith and North (and by implication Acemoglu, Johnson, Robinson, and the World Bank) did not get it quite right: "Capital became productive because of ideas for betterment—ideas enacted by a country carpenter or a boy telegrapher or a teenage Seattle computer whiz."[160] Ideas sparked economic growth first in Holland and then in England because of the dignity, liberty, and equality the law provided the citizens of these societies. "Liberated people, it turns out, are ingenious."[161] The ideas McCloskey has in mind start with the spread of reading in Europe followed by the reformation of the church, revolts against hierarchy, political revolutions, and the reevaluation of the bourgeois. These ideas were the "original and sustaining causes of the modern world."[162] They, rather than material factors, produced the West's industrial revolution and its follow-on, the Great Enrichment. Perhaps most striking are the divergent causal explanations for the West's rise in Figures 9 (Diamond), 10 (Acemoglu, Johnson, and Robinson), and 11 (McCloskey). If nothing else, the three figures underscore the lack of scholarly consensus on why and how the West got rich.

FIGURE 11: THE IDEAS THAT CAUSED THE GREAT ENRICHMENT[163]



160 Deirdre N. McCloskey, "How the West (and the Rest) Got Rich," *The Wall Street Journal*, May 20, 2016, available at http://www.wsj.com/articles/why-the-west-and-the-rest-got-rich-1463754427. This essay was adapted from Deirdre McCloskey, *Bourgeois Equality: How Ideas, Not Capital or Institutions, Enriched the World* (Chicago: University of Chicago Press, 2016).

161 McCloskey, "How the West (and the Rest) Got Rich."

162 McCloskey, *Bourgeois Equality*, p. xxxi.

163 Ibid., p. xxxi.

While many other things (including capital, the rule of law, labor, and hydrocarbon energy) were necessary, McCloskey argues that the ideas of Reading, Reformation, Revolt, Revolution and Reevaluation sufficed.[164] But as this chapter has labored to show, there were so many causal influences that acted at different times and in different ways that even a list containing dozens of candidate criteria seems unlikely to be causally exhaustive and sufficient. The complexity of what happened in Europe both before and after 1500—and especially in Britain after 1750 (but nowhere else prior to the twentieth century)—is too great. There are simply too many of the causal influences and details that we do not know at present and are unlikely ever to know.

The logic behind this last conclusion stems from a rather simple but obvious observation made by Rosenberg and Birdzell: "[The] origins of Western economic institutions cannot be traced to the wisdom of any identifiable human beings. They are the product of history, unintended consequences of action taken for entirely different reasons."[165] The West's global economic, military, and political ascendancy was an emergent phenomenon—a Black Swan in terms of both its consequential effects and unexpectedness. The first industrial revolution was the culmination of countless choices made by millions of Europeans over a period of centuries in the pursuit of their own ends. To paraphrase Adam Smith, the net result of all these decisions and actions was as if each individual, seeking only his own gain, was guided by "an invisible hand" that, in the aggregate, promoted the common good of British society through economic growth.[166]

In trying to explain how an extended economic order could emerge spontaneously from the chaos of economic activities by self-interested individuals and corporations, Friedrich Hayek (1889–1992) harked back to Smith's invisible hand.

> Adam Smith was the first to perceive that we have stumbled upon methods of ordering human economic cooperation that exceed the limits of our knowledge and perception. His 'invisible hand' had perhaps better have been described as an invisible or unsurveyable pattern. We are led—for example by the pricing system in market exchange—to do things by circumstances of which we are largely unaware and which produce results we did not intend. In our economic activities we do not know the needs which we satisfy nor the sources of the things which we get. Almost all of us serve people whom we do not know, and even of whose existence we are ignorant; and we in turn constantly live on services of other people of whom we know nothing. All this is possible because we stand in a great framework of institutions and traditions—economic, legal, and moral—into which we fit ourselves by obeying certain rules of conduct that we never made, and which we have never understood in the sense in which we understand how things that we manufacture function.

164 Ibid., pp. xxxiv–xxxvi.

165 Rosenberg and Birdzell, *How the West Grew Rich*, p. 12.

166 Smith; Cannan, ed., *An Inquiry into the Nature and Causes of the Wealth of Nations*, Vol. I, Book IV, Chapter 2, p. 477–478. Smith did not invoke the metaphor of the invisible hand elsewhere in *The Wealth of Nations*. He did use the phrase once in his 1759 *The Theory of Moral Sentiments*. There he stated his opinion that the selfish rich are led by an invisible hand to help the poor and serve the interests of society at large. Adam Smith; D. D. Raphael and A. L. Macfie, eds., *The Theory of Moral Sentiments* (Indianapolis, IN: Liberty Fund, 1984), Part IV, Chapter 1, pp. 184–185.

> Modern economics explains how such an extended order can come into being, and how it constitutes an information-gathering process, able to call up, and to put to use, widely dispersed information that no central planning agency, let alone any individual, could know as a whole, possess or control. Man's knowledge, as Smith knew, is dispersed.[167]

A century and a half after Charles Darwin published *On the Origin of Species by Means of Natural Selection* there is good reason to accept Hayek's view that markets can use dispersed and unsurveyable information to "form super-individual patterns."[168] In the case of the West's rise, the overarching pattern that emerged after 1750 was substantial and sustained economic growth relative to the rest of the world. In the case of biological evolution, the "blind, unconscious, automatic process that Darwin discovered, and which we now know is the explanation for the existence and apparently purposeful form of all life, has no purpose in mind. It has no mind and no mind's eye. It does not plan for the future."[169]

Given the kinds of processes that can give birth to emergent phenomenon, what are the prospects for a complete (and, I would add, predictive) theory of economic growth? Presumably, such a theory would enable us to reconstruct a detailed and causally exhaustive account of the West's rise. But as previously argued, this would require massive amounts of information that we neither have nor are likely to be accessible any longer. A standard distinction among professional historians is between the past, which is fixed, and history, which is not. Histories in this sense are the narratives we construct about the fixed past to explain what happened. But if we are honest with ourselves, it seems highly unlikely that we have more than a tiny fraction—even as much as five percent—of all the information needed to construct a complete history of eras that go as far back into the past as the fifteenth or sixteenth centuries.[170] This information deficit also undermines our efforts to predict "large-scale historical developments": they are, in general, "too complex to be forecast. The task is simply impossible."[171] Hence, the prospect of a complete, predictive theory of the economic growth underlying the West's rise appears dim at best.

167 F. A. Hayek; William W. Bartley III, ed., *The Collected Works of F. A. Hayek*, Vol. I, *The Fatal Conceit: The Errors of Socialism* (Chicago: University of Chicago Press, 1989), p. 14. *The Fatal Conceit*, published in 1988, was Hayek's last book. There is much debate over how much Bartley, as editor, influenced this book after Hayek fell ill in 1985.

168 Hayek, *The Collected Works of F. A. Hayek*, Vol. I, p. 15.

169 Richard Dawkins, *The Blind Watchmaker: Why the Evidence of Evolution Reveals a Universe without Design* (New York and London: W. W. Norton, 1987), p. 5.

170 The 5 percent figure is simply a guess. But thoughtful historians such as Alan Beyerchen and Andrew May agree that the actual percentage is small. From a psychological perspective, we "understand the past less than we believe we do." Daniel Kahneman, *Thinking, Fast and Slow* (New York: Farrar, Straus and Giroux, 2011), p. 201. Citing Taleb's narrative fallacy, Kahneman argues that "we humans constantly fool ourselves by constructing flimsy accounts of the past and believing they are true." Ibid., p 199. Hence comes the title of Taleb's 2001 book *Fooled by Randomness*.

171 Daniel Kahneman and Gary Klein, "Conditions for Intuitive Expertise: A Failure to Disagree," *American Psychologist*, September 2009, p. 520.

Douglass North reached the same conclusion but for more esoteric reasons. One profound difference he noted in 2005 between biological evolution and economic change is that the latter "is for the most part a deliberate process shaped by the perceptions of the actors about the consequences of their actions."[172] Stemming from this observation North concluded that the study of the process of economic change must begin "by exploring the ubiquitous efforts of human beings to deal with and confront uncertainty in a non-ergodic world."[173] To explain what he meant by saying that our economic world is non-ergodic, North cited a 1991 article by Paul Davidson in which Davidson defined an ergodic stochastic process as one in which "averages calculated from past observations cannot be persistently different from the time average of future outcomes."[174] A non-ergodic process, then, is one in which the time averages of future outcomes *can* be (or *are*) persistently different from those calculated from past observations. In an ergodic world, the future is determined by existing parameters and statistically reliable estimates of the probability of future events can be calculated from past and current probability distributions of market data; in a non-ergodic world our economic future may be causally determined but is not predictable.[175]

According to Davidson, the reason the future is not "calculable" in a non-ergodic world is the presence of uncertainty.[176] Davidson attributed this insight to John Maynard Keynes (1883–1946). In 1937 Keynes offered the following characterization of what he termed uncertain knowledge:

> By "uncertain" knowledge . . . I do not mean merely to distinguish what is known for certain from what is only probable. . . . The sense in which I am using the term is that in which the prospect of a European war is uncertain, or the price of copper and the rate of interest twenty years hence, or the obsolescence of a new invention, or the position of private wealth-owners in the social system in 1970. About these matters there is no scientific basis on which to form any calculable probability whatever. We simply do not know.[177]

172 "The perceptions come from the beliefs of the players—the theories they have about the consequences of their actions—beliefs that are typically blended with their preferences." North, *Understanding the Process of Economic Change*, p. viii.

173 Ibid., p. 5.

174 Ibid., p. 19; and Paul Davidson, "Is Probability Theory Relevant for Uncertainty? A Post Keynesian Perspective," *Journal of Economic Perspectives* 5, no. 1, Winter 1991, p. 132. Recently *The Economist* compared the annualized returns for 928 hedge funds from the years 2005–2009 with their returns from the years 2009–2015. The conclusion: "Past performance is not a guide to future returns." "Law of Averages," *The Economist*, August 27, 2016, p. 57.

175 Paul Davidson, "A Response to John Kay's Essay on the State of Economics," *blog*, The Institute for New Economic Thinking, October 5, 2011.

176 Davidson, "Is Probability Theory Relevant for Uncertainty?" pp. 130–131.

177 John M. Keynes, "The General Theory of Employment," *The Quarterly Journal of Economics*, February 1937, pp. 213–214.

Keynes' formulation obviously envisioned both certain and probable knowledge along with uncertain knowledge. Davidson, therefore, acknowledged in 1991 that many economic processes could be ergodic. In those cases, the "ergodic hypothesis" (or axiom) would be perfectly acceptable as well as more tractable mathematically, putting economics on a par with physics and chemistry. That is why in 1969 Paul A. Samuelson (1915–2009) declared the ergodic hypothesis to be "the *sine qua non* of the scientific method in economics."[178]

Nevertheless, North and Davidson are convinced that our economic world is fundamentally non-ergodic.[179] Because economic change depends on human perceptions and intentions, and, in a non-ergodic world, "We cannot know today what we will learn tomorrow which will shape our tomorrow's actions," the prospects for a dynamic theory of economic change are poor, if virtually nonexistent. On present knowledge, "No such theory that could be useful is likely to evolve."[180] This conclusion appears to be an unmistakable manifestation of limits to human knowledge that will not be overcome by gathering more information or increasing our computational capabilities. Human rationality, as Herbert Simon observed in the 1950s, is bounded and "falls far short of the ideal of 'maximizing' postulated in economic theory."[181]

What, then, can we conclude about identifying the necessary and sufficient causal criteria behind the West's rise while also explaining the absence of industrial revolutions in other civilizations during 1750–1880 or earlier? This chapter has reviewed a long list of candidate criteria. But given the limits to our knowledge of the past and the absence of a dynamic theory of very long-term economic growth, the importance and weight we ascribe to any particular criterion are a matter of subjective, fallible human judgment. Indeed, we cannot even be certain that we have identified all the major factors and not included any irrelevant ones.

178 Davidson, "Is Probability Theory Relevant for Uncertainty?" p. 133; P. A. Samuelson, "Classical and Neoclassical Theory," in R. W. Clower, ed., *Monetary Theory* (London: Penguin Books, 1969), p. 184; and North, *Understanding the Process of Economic Change*, p. 19.

179 Is Keynes' uncertainty epistemological or a fundamental (ontological) feature of reality? Werner Heisenberg's uncertainty principle in quantum mechanics suggests that Keynes' uncertainty is real. Heisenberg's uncertainty principle states that there is a fundamental limit to the precision with which certain pairs of physical properties of a particle, known as complementary variables, such as position x and momentum p, can be known simultaneously. The formal inequality is $\sigma_x \sigma_{px} \geq \hbar/2$ where σ is the standard deviation and ħ is the reduced Planck's constant h/2π.

180 North, *Understanding the Process of Economic Change*, pp. 19, 126. Richard Bookstaber extends North's point about past experience affecting our economic future: "our continual accumulation of knowledge" changes not only our view of the world but can change the economic system itself. Richard Bookstaber, *The End of Theory: Financial Crises, the Failure of Economics, and the Sweep of Human Interaction* (Princeton and Oxford: Oxford University Press, 2017), pp. 17–18, 40–42, 58, 84.

181 Herman A. Simon, "Rational Choice and the Structure of the Environment," *Psychological Review* 63, no. 2, 1956, p. 129.

Although one can understand North's claim that our economic world is non-ergodic without deciding whether he or Davidson are right, the issue points to a fundamental schism in economics. Most mainstream economists (John Cochrane, Joseph Stiglitz, Myron Scholes, Milton Friedman, etc.) side with Samuelson and accept the assumption that the world of economics is ergodic.[182] Even Taleb comes down on the side of the ergodic hypothesis. His Black Swans, he admitted in the second edition (2010) of his book of the same name, are merely very rare events, unexpected by at least some (if not many) observers, that result from ergodic processes.[183] If so, then in economics (at least) contingency and free will are illusions. "[The] future path of the economy is already predetermined and cannot be changed by human action today."[184]

It is hard to square this deterministic outlook with the deep-seated sense derived from everyday experience that suggests we really cannot predict or know the long-term future with certainty. On the one hand, chaos theory argues that a difference in initial conditions as small as the flap of a butterfly's wings could lead to differences in the weather as large as a tornado. Yet, on the other hand, the ergodic hypothesis asks us to believe that the global financial crisis of 2008–2009 generated by the collapse of the market for mortgage-backed derivatives could, in theory, have been predicted by better risk management models than the ones used by early twenty-first century financial engineers such as David X. Li.[185] But as Davidson observed in 2011, even in physics, theories are never "conclusively established and can be replaced when events are observed that are deviations from the current existing theory." Thus, the financial crisis of 2008–2009 "should have been sufficient empirical evidence" that the axiomatic basis of mainstream economic theory—in particular, the ergodic axiom—needs to be replaced.[186]

182 Paul Davidson, "Is Economics a Science? Should Economics Be Rigorous?" *Real-World Economics Review*, no. 59, 2012, pp. 1, 2.

183 Taleb, *The Black Swan*, p. 344. Taleb's writing is seldom a paragon of clarity. Earlier in his long postscript to the second edition of *The Black Swan* he describes Black Swans as contingencies. Ibid., pp. 324–315.

184 Paul Davidson, *The Keynes Solution: The Path to Global Economic Prosperity* (New York: Palgrave Macmillan, 2009), p. 37.

185 For a discussion of one model that played a role in the 2008–2009 financial crisis see Felix Salmon, "Recipe for Disaster: The Formula That Killed Wall Street," *Wired*, February 23, 2009, pp. 74–79, 112, available at http://www.wired.com/2009/02/wp-quant/?currentPage=all. The formula discussed in this article is David X. Li's Gaussian copula function. For an overview of the multiple causes of the crisis see "The Origins of the Financial Crisis: Crash Course," *The Economist*, September 7, 2013, pp. 74–75. *The Economist* allocated blame not only to the irrational exuberance of Anglo–Saxon financiers but also to the central bankers and regulators who tolerated the folly of the financiers. The financial engineers who pooled "subprime" mortgages, then used them to back supposedly safe collateralized debt obligations, "claimed to have found a way to banish risk when in fact they had simply lost track of it." Ibid., p. 74.

186 Davidson, "A Response to John Kay's Essay on the State of Economics."

In this regard, chaos theory has been cited to support the unpredictability of the distant future—at least under some initial conditions. But as Marcus du Sautoy has concluded, chaos also undermines our ability to know the past. "[T]rying to work backwards and understand what state our planet was in to produce the present is equally if not more challenging."[187]

We set out in this chapter to discover the main, principal, dominant causal criteria that might explain why the West rose to wealth and global dominance after 1500 while the civilizations of China, the Middle East, and India did not. Sampling the (vast) literature on this issue produced a wide range of candidate causal criteria that have been advanced to explain what happened and why. But we have found neither scholarly consensus on the primary drivers (again, compare Figures 9, 10, and 11) nor a theory of very long-term economic growth that might allow one to converge on an exhaustive list of necessary and sufficient causal factors. In fact, the vast majority of the causal details that would be needed to completely reconstruct the West's rise are no longer accessible, and chaotic processes can irretrievably conceal the precise causes of the present at critical junctures in the past. These results go far to support the conclusion that choosing appropriate analytic criteria is a ubiquitous problem even in the case of understanding the historical past, and that there is no by-the-numbers, mechanistic procedure for making such choices.

187 Du Sautoy, *What We Cannot Know: Explorations at the Edge of Knowledge*, p. 55.

CHAPTER 4

GDP, Gauging the Pace of Business, and Military Effectiveness Metrics

> *GDP is the way we measure and compare how well or badly countries are doing. But it is not a question of measuring a natural phenomenon like land mass or average temperature to varying degrees of accuracy. GDP is a made-up entity.*
>
> Coyle, 2014[1]

> *Gross domestic product (GDP) is increasingly a poor measure of prosperity. It is not even a reliable gauge of production.*
>
> The Economist, 2016[2]

> *I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meager and unsatisfactory kind . . .*
>
> Lord Kelvin, 1883[3]

Chapter 3 focused on the past. It endeavored to show how our choices of analytic measures influence our understanding of the West's rise to global economic, military and political dominance. This chapter turns to how such choices can and do influence our understanding of the present and, consequently, our projections about the future. Gross domestic product has been selected as a point of departure for several reasons. First, since the 1940s GDP has become the most frequently referenced criterion for comparing the economic output and trajectories

1 Coyle, *GDP: A Brief but Affectionate History*, p. 4.

2 "The Trouble with GDP," *The Economist*, April 30, 2016, p. 21.

3 William Thompson, *Popular Lectures and Addresses*, Vol. I, *Constitution of Matter* (New York: Macmillan, 1889), p. 73.

of nations and other economic domains that may be larger or smaller than nations. Second, GDP is devilishly complex, as are the myriad economic activities and interactions that GDP estimates seek to quantify. Third, calculating the actual number for the GDP of an economic territory within a given timeframe is "the product of a vast patchwork of statistics and a complicated set of processes carried out on the raw data to fit them to the conceptual framework" manifested in the prevailing international system of national accounts.[4] Fourth, per capita GDP figures (such as those in Table 6) are frequently taken to be an indicator of material wealth or even well-being. Finally, since World War II GDP numbers and comparisons have often been used to influence the economic decisions and policies of nations and international organizations (especially those concerned with fostering economic development). GDP, then, has been a prominent and consequential analytic criterion for all manner of purposes for decades and remains so to this day.

Another pattern suggested by Chapter 3 is that deeper insight into the how and why of outcomes like the European miracle may be had by considering a robust set of likely causes rather than focusing on any single criterion—even if the list is not causally exhaustive and particular elements acted at different times in different ways. This thought raises the possibility of using competing lists of analytic criteria to settle an argument or support a conclusion. This form of argumentation is often encountered in the pages of *The Economist*, and Chapter 4 examines its use in deciding whether the pace of American business is accelerating as many corporate leaders and observers now believe.

Chapter 1 explored the role various choices of analytic criteria can play in understanding the eventual outcome of the Battle of the Atlantic in World War II as well as the role such choices played in assessing the degree to which the Combined Bomber Offensive could be considered more of a success than a failure. In both cases, the outcomes were analyzed primarily as the results of strategic campaigns. Little attention was paid to the accretion of tactical interactions that eventually made up these two strategic campaigns. Might a focus on tactical engagements provide more insight into the drivers or causal criteria that eventually separate success from failure? The final section of this chapter will examine four plausible candidates for such criteria: casualties, force ratios, firepower scores, and situation awareness. As we shall see, historical data on casualties and force ratios offer little ability to predict likely battle outcomes. The same is true of firepower scores. Situation awareness, however, appears to be a more frequent driver of engagement outcomes, although it is not quantifiable in advance. To foreshadow Chapter 5, the inability to predict combat outcomes has profound implications for the net assessment problem of estimating where one side stands relative to another in a military competition, particularly during peacetime.

4 Coyle, *GDP: A Brief but Affectionate History*, p. 37.

GDP and Wealth

The widespread use of GDP to compare the economic output or activity of countries was an invention of World War II.[5] There were, however, antecedents. As early as 1665 a British scientist and official, William Petty, made the first systematic attempt to measure the income and expenditure, population, land, and other assets of England and Wales.[6] Petty's aim was to show that the country had the resources to take on Holland and France militarily. Much later, during the 1920s and 1930s, Colin Clark in the United Kingdom calculated national income and expenditure statistics on a quarterly rather than annual basis and discussed how to adjust for inflation.[7] In the United States during the Great Depression, President Franklin Roosevelt wanted a clearer picture of the American economy. To provide that picture the National Bureau of Economic Research tasked Simon Kuznets to develop Colin Clark's methods and apply them to the U.S. economy. Kuznets' first report, in January 1934, "showed that American national income had been halved between 1929 and 1932," giving Roosevelt statistical support for the various economic recovery programs that became known as the New Deal.[8]

By 1937 disagreement emerged over what to include in national income. Kuznets saw his task as measuring national economic *welfare* rather than just *output*; he wanted to subtract from national income all expenses on armaments, most outlays on advertising, many financial expenses, and outlays on infrastructures such as subways and expensive housing.[9] Opponents (especially Martin Gilbert at the Commerce Department) wanted to include all government expenditures, including armaments. In a time of war, omitting government expenditures would have shown national income shrinking as consumption declined and spending on military production expanded. But the president would want a measure of the economy that captured its potential to support the war effort. In the end, Gilbert and his colleagues prevailed; GNP (and, later, GDP)[10] would include government expenditures, thereby establishing the federal government's role in the economy as an ultimate consumer of goods and

5 The annual or quarterly economic output of a country or smaller entity is not to be confused with a nation's wealth. The latter would include human capital (the population's demographics, education, skills and productivity); natural capital (including geography, agricultural land, forests, fossil fuels, and minerals); and "manufactured" or physical capital (machinery, buildings, infrastructure, investments, etc.). Partha Dasgupta et al., *Inclusive Wealth Report 2012: Measuring Progress Toward Sustainability* (Cambridge, UK: Cambridge University Press, 2012), pp. 31, 281–287. Expressed in FY 2000 constant dollars, the study calculated that the inclusive wealth index of the United States was nearly six times that of China in 2008. Ibid., pp. 301, 329.

6 Coyle, *GDP: A Brief but Affectionate History*, p. 8.

7 Ibid., p. 12.

8 Ibid., p. 13. Roosevelt's New Deal recovery programs, implemented during 1933–1938, sought to stabilize the U.S. economy by creating long-term employment opportunities, decreasing agricultural supply to drive prices up, and helping homeowners pay mortgages and stay in their homes, which also kept the banks solvent.

9 Ibid., pp. 13–14. Keynes, too, supported the idea that government spending should add to national income rather than reducing it. Ibid., p. 17.

10 For the difference between GNP and GDP, see footnote 64 in Chapter 3.

services.[11] At the height of World War II in 1944, U.S. spending on national defense peaked at 37 percent of GDP (roughly $83 billion in current dollars).[12]

In May 1946 British and American experts met in New York to draw up recommendations on collecting national economic statistics for the newly established United Nations (UN). 13 months later George C. Marshall, then secretary of state, announced that the United States would provide aid to the European nations devastated by the war to support their economic recovery. As early as 1919 Keynes had made his name with a pamphlet arguing that the punitive financial terms the Versailles treaty imposed on Germany after World War I would destabilize the country politically and economically, which was exactly what happened with the rise to power of Adolf Hitler. Marshall's European recovery program (informally the Marshall Plan) sought to avoid repeating this mistake.

The administration of Harry Truman needed comparative economic data such as GNP to monitor implementation of the Marshall Plan. The onset of the U.S.–Soviet Cold War provided a further incentive to be able to compare national economies: in this case, the USSR's centrally planned economy versus America's market economy. Here too GNP was the obvious answer for the kind of standardized measure that was needed.[13]

In 1947 the UN issued a technical report recommending how the underlying calculations for GNP should be conducted. The Organization for European Economic Co-operation (OEEC), a kind of economic brain trust for member countries, took on the role of gathering national account figures for all the member countries and comparing them.[14] In 1953 the UN published its first official *System of National Accounts* (*SNA1953*). Today international agencies, prominently the World Bank, publish GDP comparisons for all the world's countries using *SNA2008*, which was published in 2009 by the International Monetary Fund (IMF), the Organization for Economic Co-operation and Development (OECD),[15] the World Bank, and the UN.

To give an idea of how arcane and complex actually calculating GDP for a country is, *SNA2008* runs some 660 pages.[16] The document is accompanied by an equally daunting double-entry spreadsheet that provides a numerical example. In addition, François Lequiller and Derek Blades have written a widely used explanatory commentary on the *System of National Accounts*; its second edition is over 500 pages. The community of national statisticians with

11 Coyle, *GDP: A Brief but Affectionate History*, pp. 15, 19.

12 OUSD (Comptroller), *National Defense Budget Estimates for FY 2017* (Washington, DC: DoD, March 2016), p. 264; and Bureau of Economic Analysis, "Current-Dollar and 'Real' Gross Domestic Product," Excel spreadsheet, July 29, 2016, available at http://www.bea.gov/national/index.htm#gdp. Current dollars do not take into account price inflation.

13 Coyle, *GDP: A Brief but Affectionate History*, p. 47.

14 Ibid., p. 42.

15 After 1961 the OEEC became the Organization for Economic Co-operation and Development (OECD).

16 *SNA53* had less than 50 pages.

command of all this detail is small, and very few people "truly understand how the regularly published GDP figures are constructed."[17]

One of the elementary aspects of calculating GDP that Lequiller and Blades explain is that there are three distinct approaches to constructing this metric: (1) the output approach (the sum of gross values added); (2) the final-demand approach (consumption + investment + net exports); and (3) the income approach (compensation of employees + gross operating surplus + gross mixed income).[18] Table 7 illustrates all three approaches for the case of Germany in 2012. Note that the different approaches yield the same GDP total in euros.

TABLE 7: GERMANY'S 2012 GDP (BILLIONS OF EUROS)[19]

Gross Domestic Product (input approach)	**€ 2,666**
Gross value added at basic prices, excluding FISIM	€ 2,387
Plus taxes less subsidies on products	€ 280
Gross Domestic Product (expenditure approach)	**€ 2,666**
Final consumption expenditure	€ 2,048
Plus gross capital formation	€ 460
Plus exports of goods and services	€ 1,381
Minus imports of goods and services	€ 1,223
Gross Domestic Product (income approach)	**€ 2,666**
Compensation of employees	€ 1,376
Plus gross operating surplus and gross mixed income	€ 1,016
Plus taxes less substitutes on production and imports	€ 274

The detailed data and analysis underlying each of the major GDP components in Table 7 are complicated and arduous to calculate. The final consumption expenditure for households alone (excluding those for general government and non-profits serving households) contains no less than 14 categories: (1) food and non-alcoholic beverages; (2) alcoholic beverages, tobacco, and narcotics; (3) clothing and footwear; (4) housing, water, electricity, gas, and other fuels; (5) furnishings, household equipment, and maintenance; (6) health; (7) transport; (8) communications; (9) recreation and culture; (10) education; (11) restaurants and

17 Coyle, *GDP: A Brief but Affectionate History*, p. 25.

18 François Lequiller and Derek Blades, *Understanding National Accounts*, Second Edition (Paris: OECD Publishing, 2014), p. 31.

19 Ibid., pp. 23, 31. FISIM stands for financial intermediate services indirectly measured. Table 7 shows "nominal" rather than "real" GDP, meaning that the GDP data have not been adjusted for price inflation.

hotels; (12) miscellaneous goods and services; (13) final consumption expenditures of resident households abroad; and (14) final consumption expenditures of non-resident households on the territory.[20]

A further complication is adjusting GDP components in current prices for inflation. Lequiller and Blades devote an entire chapter in *Understanding National Accounts* to explaining how statisticians distinguish changes in the volume output of an economy from inflationary (or deflationary) changes in prices. Without inflation adjustments, one country's GDP cannot be compared with another's, and within a country over time inflation can be mistaken for real growth. The OECD makes these adjustments using the concept of purchasing power parity (PPP).[21] "Although most price and volume index numbers were developed to measure changes in prices and volumes over time, they can also be adapted to compare levels of prices and volumes between different regions or countries in the same period of time."[22] In terms of comparisons between countries, *SNA2008* defines PPP "*as the number of units of B's currency that are needed in B to purchase the same quantity of individual good or service as one unit of A's currency will purchase in A*."[23] Making inflation adjustments with PPP, whether across borders or over time, is "very complicated, perhaps one of the most challenging of all the methodological statistical issues" in constructing comparable GDP aggregates.[24] Any meaningful comparison of prices between countries or over time must consider an identical basket of goods and services. Not only does doing so entail collecting large amounts of data, but it must deal with the fact that many current products (for example, laptop computers and software) did not even exist in the 1940s when GNP was developed. PPP exchange rates for various countries are generated by the International Comparisons Program's global price surveys. For the 2003–2006 round, each of the roughly 147 participating countries provided national average prices for 1,000 closely specified products.[25] *The Economist*'s Big Mac index provides a simpler approach to understanding exchange-rate theory.[26]

Notice that Table 7's GDP for Germany during 2012 is in euros. To compare Germany's GDP with that of another country such as the United States requires one further step: conversion of Germany's GDP in euros into dollars; the international dollar has long been the currency standard for such comparisons. There are two ways of doing the requisite conversions. One

20 Ibid., pp. 139, 142.

21 Ibid., p. 50.

22 *System of National Accounts 2008* (New York: European Commission, IMF, OECD, UN, World Bank, 2009), p. 295.

23 Ibid., p. 318. Italics in the original.

24 Coyle, *GDP: A Brief but Affectionate History*, p. 31.

25 Tim Callen, "Purchasing Power Parity: Weights Matter," IMF, updated March 28, 2012, available at http://www.imf.org/external/pubs/ft/fandd/basics/ppp.htm.

26 In 1986 *The Economist* invented a lighthearted guide to whether currencies are at their "correct" level using the price of a Big Mac hamburger in various countries as a surrogate for an identical basket of goods and services. Since then the Big Mac index has become widely used to explain exchange-rate theory. "The Big Mac Index," interactive online tool, *The Economist*, updated as of July 13, 2017, available at http://www.economist.com/content/big-mac-index.

is to use the dollar–euro exchange rate that prevailed during the year in question in currency markets. This approach, however, does not take into account the substantial differences in purchasing power among nations, particularly low-income countries whose trade is only a small share of its GDP.[27] The widely used alternative is to use PPP exchange rates to "adjust the actual exchange rate to one that reflects living standards more realistically."[28] For developed countries like Germany, the difference between using dollar–euro exchange rates and PPP is small; Germany's 2012 GDP of 2,666 billion in euros was $3,542 billion using currency exchange rates and $3,557 billion using PPP conversions.

In the case of China and other relatively poor countries, the difference can be much greater. China's 2015 GDP was $10,983 billion using dollar–yuan exchange rates but came to $19,392 billion using a PPP valuation.[29] By comparison, U.S. GDP in 2015 was $17,947 billion in both current dollars and a PPP valuation. Obviously, PPP conversion factors will raise the relative GDP of "poor" countries where non-traded goods and services are cheap.[30] But as large as China's economy has become, the country is still poor compared to the United States in terms of per capita GDP, which many tend (incorrectly) to view as a measure of wealth. According to World Bank data, China's per capita GDP in 2015 was $7, 924 using currency exchange rates, and $14,238 using PPP. The per capita GDP of the United States in 2015, by comparison, was $55,837. These figures suggest some of the reasons PPP has been criticized. Using a PPP valuation, China's GDP surpassed that of the United States in 2015, whereas China's economy was some 38 percent smaller using dollar–yuan conversion rates. Worth recalling is that China's participation in a 2005 survey aimed at refining PPP conversions led the World Bank in 2007 to reduce China's real GNP in PPP terms by a staggering 40 percent.[31]

GDP was originally designed to measure a country's economic output based on the value of the things it produced, net of inputs, in a given year.[32] In the United States after World War II, the economy largely consisted of farms, production lines, and mass markets. By the 1970s, and especially in the United States, the post-World War II rise of consumerism led to the mass production of cars, radios, refrigerators, washing machines, TVs, cameras, lawnmowers, telephones, etc.—the list is long. However, as *The Economist* observed in 2016, in the post-war period it was "tricky enough" to calculate GDP (then GNP) for the American economy; but for today's "rich economies, dominated by made-to-order services

27 Coyle, *GDP: A Brief but Affectionate History*, p. 49.

28 Ibid.

29 Knoema, "World GDP Ranking 2015: Data and Charts," available at https://knoema.com/nwnfkne/world-gdp-ranking-2015-data-and-charts.

30 World Bank, "GDP per capita (current US$)," available at http://data.worldbank.org/indicator/NY.GDP.PCAP.CD; and "GDP per capita, PPP (current international $)," available at http://data.worldbank.org/indicator/NY.GDP.PCAP.PP.CD. Geary-Khamis or international dollars are a hypothetical unit of currency that has the same purchasing power parity as the U.S. dollar has in the United States at a given point in time.

31 Coyle, *GDP: A Brief but Affectionate History*, p. 53.

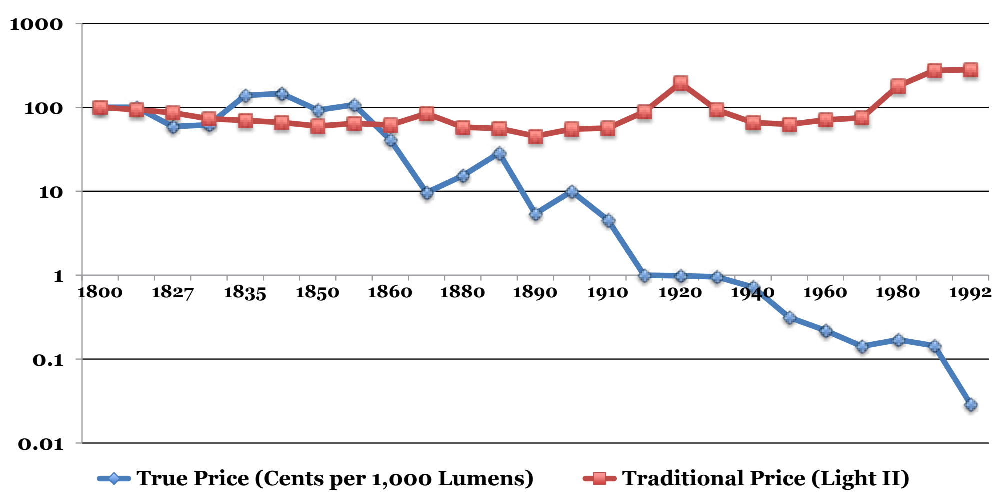
32 Ibid., p. 121.

and increasingly geared to the quality of experience rather than the production of ever more stuff, the trickiness is raised to a higher level."[33]

In 1996 William Nordhaus illustrated this higher level of trickiness by exploring the evolution of artificial illumination or lighting. His primary concern was about GNP's ability to capture technological change. "During periods of major technological change," he wrote at the outset, "the construction of accurate price indexes that capture the impact of new technologies on living standards is beyond the practical capability of official statistical agencies."[34] Toward this end, he used data on lighting efficiency to construct a "true" price of light and compare it with "traditional" price indexes.

Light visible to the human eye is, of course, electromagnetic radiation in the frequency range of 405 terahertz (THz), or red, to 790 THz, or violet.[35] Light flow or flux, measured in lumens, is the total quantity of visible light emitted by a source. A wax candle emits around 13 lumens while a 100-watt filament bulb emits about 1,200 lumens. Nordhaus opted to use lumens per watt to measure the efficiency of lighting devices in terms of light output per unit of energy input, and he settled on the price per 1,000 lumens to reflect the true price of lighting.[36]

FIGURE 12: TRUE VERSUS TRADITIONAL LIGHTING PRICES[37]



33 "The Trouble with GDP," *The Economist*, p. 21.

34 William D. Nordhaus, "Do Real-Output and Real-Wage Measures Capture Reality? The History of Lighting Suggests Not," in Timothy F. Bresnahan and Robert J. Gordon, eds., *The Economics of New Goods* (Chicago: University of Chicago Press, January 1996), p. 29.

35 One THz = 1012 Hertz or cycles per second.

36 Nordhaus, "Do Real-Output and Real-Wage Measures Capture Reality?" p. 31, 45–48.

37 Ibid., pp. 46–47. 1800 = 100.

These metrics are the ones that would naturally occur to physicists rather than to economists focused on price indexes. Traditional price indexes for lighting reflect the prices consumers pay to buy lighting (bulbs, lamps, etc.). By contrast, Nordhaus constructed the true price of lighting as a physicist would: by calculating the cost in cents per 1,000 lumen-hours from 1800 to 1992.[38] Using price indexes he concluded that the traditional price of light had "risen by a factor of between three and five in nominal terms" from 1800 to 1992; but comparison with the price of light as a physicist might calculate it in cents per 1,000 lumen-hours indicated that the true price had plummeted by "a factor of between nine hundred and sixteen hundred."[39] Presumably, the appearance since 1992 of light-emitting diode (LED) bulbs has reduced the true cost of light even further.

This example illustrates the difficulties national accounts and GDP have in measuring changes in the quality of the goods that are increasingly being consumed—both in terms of innovative products such as smartphones and greater choice in the marketplace. Again, GNP was originally intended to measure an economy's capacity to add value—predominantly through production—from one year to the next. Nevertheless, this one number has become a "lodestar for policies to set taxes, fix unemployment and manage inflation," as well as a measure of "material well-being, even though it is a deeply flawed gauge of prosperity, and getting worse all the time."[40]

The problems with GNP are many. For one reason or another, a country's GDP in a given year may be substantially revised as time goes on. A recent example is Nigeria. In 2014 the country's statistician-general announced that Nigeria's GDP for 2013 had been revised from 42.4 trillion naira to 80.2 trillion naira ($510 billion), an 89 percent increase.[41] The reason was not dodgy. A country's GDP is measured by reference to its economy's "shape" in a "base" year, each sector's weight depending on its importance to the economy. Nigeria had last "rebased" its economy in 1990, before the rise of mobile telecoms and film-making as significant economic sectors. The 89 percent increase in Nigeria's 2013 GDP resulted from rebasing the country's sector weights to 2010.

GDP's bias toward manufacturing is another distortion. Since 1950 the services sector of the U.S. economy—the other two sectors being industry and agriculture—has grown from a little over half of GDP to more than 75 percent. Nevertheless, "The output of factories is still measured more closely than that of services."[42] Financial services are a case in point. As the financial services industry grew throughout the 1980s in countries such as the United States

38 Ibid., p. 38. 1992 saw the introduction of first-generation compact fluorescent bulbs.

39 Nordhaus, "Do Real-Output and Real-Wage Measures Capture Reality?" p. 51; and Coyle, *GDP: A Brief but Affectionate History*, p. 86.

40 "How to Measure Prosperity," *The Economist*, April 30, 2016, p. 7.

41 "Step Change," *The Economist*, April 12, 2014, available at http://www.economist.com/news/finance-and-economics/21600734-revised-figures-show-nigeria-africas-largest-economy-step-change.

42 "The Trouble with GDP," *The Economist*, p. 22.

and Britain, statisticians sought to find a way of measuring earnings from financial intermediation, which for many years had been seen as adding little value to the economy or even subtracting value.[43] To address the problem *SNA1993* introduced the concept of "financial intermediation services indirectly measured" (FISIM). FISIM is used to describe the services that banks provide to their customers that are not invoiced. For bank depositors these services typically include managing current accounts, sending out bank statements, and making fund transfers between accounts. FISIM is based on a reference rate that is an estimate of a pure interest rate involving no element of risk. FISIM is calculated by adding the difference between the bank's loan and reference rates multiplied by the amount of loans, then adding that to the difference between the reference and deposit rates multiplied by the amount of deposits.[44] In effect, FISIM is "a way of measuring the service provided by banks in taking on risk."[45]

Since the financial crisis of 2008–2009 the role of banks in the economy and the measurement of their output in national accounts has drawn increased attention, with some commentators going so far as to say that banks should not have any positive output in the national accounts. FISIM, after all, records increased risk-taking as real growth in financial services. But as Andrew Haldane and his co-authors suggested in 2010, it is neither clear that bearing risk is, in itself, a productive activity nor that risk-based income flows should represent bank output.[46] Nonetheless, Lequiller and Blades defended FISIM in 2014, noting that from 2007 to 2012 the output of the financial sector in U.S. national accounts decreased both in current prices and in volume.[47] Diane Coyle concluded her discussion of FISIM with two comments. First, "The financial sector in recent years has been overstated by at least one-fifth, maybe even by as much as one-half."[48] Second, "The absurdity of recording big increases in the contributions made by financial services to GDP as the biggest financial crisis in a generation or two got underway indicates that the statistical approach [to FISIM] is mistaken."[49]

Suffice it to say, there are many, all-to-real problems with trying to reduce the output of entire national economies to a single number measured in a single currency while dealing with inflation and disagreements about what should count as productive as opposed to unproductive activity. Efforts have been made to come up with alternatives. In 1972 Nordhaus and James Tobin proposed a measure of economic welfare that counted some bits of government

43 Coyle, *GDP: A Brief but Affectionate History*, p. 100.

44 Lequiller and Blades, *Understanding National Accounts: Second Edition*, p. 124.

45 Coyle, *GDP: A Brief but Affectionate History*, p. 100.

46 Andrew Haldane, "The Contribution of the Financial Sector—Miracle or Mirage?" *Bank for International Settlements Review*, no. 95, 2010, p. 4, available at http://www.bis.org/review/r100716g.pdf. This article was taken from a co-authored chapter by Andrew Haldane, Simon Brennan, and Vasileios Madouros in *The Future of Finance: The LSE Report* (London: The London School of Economics and Political Science, 2010).

47 Lequiller and Blades, *Understanding National Accounts: Second Edition*, p. 125.

48 Coyle, *GDP: A Brief but Affectionate History*, p. 101.

49 Ibid., pp. 103–104.

spending, such as defense and education, not as outputs but as costs to GDP and also adjusted for wear-and-tear to capital and the "disamenities" of urban life, such as congestion.[50] Their paper was in part a response to environmentalist concerns that GNP treated the plunder of the planet not as a cost, but as something that added to income.[51] And in 2009 Joseph Stiglitz, Amartya Sen, and Jean-Paul Fitoussi published a report, commissioned by the French president Nicolas Sarkozy, which suggested a "dashboard" of measures would better capture human welfare than combining them into a single number as GDP does.[52] Yet, GDP, due to its enduring appeal as a summary statistic that tells people how well various economies are doing, has not gone by the wayside. Nordhaus and Paul A. Samuelson famously wrote that "While the GDP and the rest of the national income accounts may seem to be arcane concepts, they are truly among the great inventions of the twentieth century." Coyle pretty much agreed at the end of her 2014 book *GDP: A Brief but Affectionate History*. Granted, the global financial crisis of 2008–2009 has given extra urgency to the need to rethink our conceptions of economic value, and the gap between GDP and welfare or prosperity is growing bigger. But, Coyle concluded, GDP, for all its flaws, is still "a bright light shining through the mist" of the statistical fog surrounding our efforts to quantify economic activity.[53]

Most of the analytic criteria discussed so far have been far simpler and less highly aggregated than GDP. GDP is perhaps the epitome of a statistically complex metric that combines all manner of things in a single number while excluding others. To emphasize the fuzziness of the boundary between productive and unproductive economic activities, in 2013 a European Union agreement on GDP standards decided to include income from selling recreational drugs and paid sex work, yet housework and caring for elderly relatives remain excluded from GDP to this day, even though such unpaid services have considerable value.[54] There seems to be no simple solution to such inconsistencies. Not only is the selection of appropriate analytic criteria a matter of judgment, but so is the construction of metrics as highly aggregated as GDP.

Dueling Criteria

A recurring form of argument is to dispute or reject a conclusion that has been based on one set of metrics by invoking an alternative set of analytic criteria. As an example, in its December 5, 2015, edition *The Economist* addressed this question: Is the pace of business really getting quicker? The magazine's cover, which depicted a white-knuckled chief executive officer (CEO) trying to hold on as the world of business rushed ahead, certainly

50 William Nordhaus and James Tobin, "Is Growth Obsolete?" in *Economic Research: Retrospect and Prospect*, Vol. 5, *Economic Growth* (Washington, DC: National Bureau of Economic Research, 1972), pp. 7, 12–13, 16–17, available at http://www.nber.org/chapters/c7620.pdf.

51 "The Trouble with GDP," *The Economist*, p. 22.

52 Coyle, *GDP: A Brief but Affectionate History*, p. 118.

53 Ibid., p. 140.

54 "The Trouble with GDP," *The Economist*, p. 23.

captured the perception of many corporate bosses that, "21st-century business is pushing its pedals ever harder to the metal."[55]

It is easy to point to criteria suggesting that the pace of business is speeding ahead if not accelerating. For the second week in December 2015, *The Economist* estimated that,

> 10 billion shares of America's 500 largest listed firms will have changed hands in frenzied trading. Their bosses will have been swamped by 750,000 incoming e-mails and a torrent of instant data about customers. In five days these firms will have bought $11 billion of their own shares, not far off what they invested in their businesses. With one eye on their smartphones and the other on their share prices, bosses seem to be the bug-eyed captains of a hyperactive capitalism.

To further underscore current perceptions of the accelerating pace of business, *The Economist* added:

> When managers are not striving to satisfy investors whose allegiance to firms is measured in weeks, they are pumping up share prices in order to maximise their own pay. Executives feel harried, too. Competition is becoming ever more ferocious: if Google or Apple are not plotting your downfall, a startup surely is.[56]

The perception of most American CEOs regarding the pace of business, then, is very much that suggested in David Parkin's cover illustration.[57]

Unquestionably there are sectors in which the pace of American business today is frenetic. Take Apple: "A customer downloads an app from Apple every millisecond. The firm sells 1,000 iPhones, iPads or Macs every couple of minutes. It whips through its inventories in four days and launches a new product every four weeks. Manic trading by computers and speculators means the average Apple share changes hands every five months." Moreover, high-flying startups can win billion-dollar valuations within a year or two of coming into existence and in a few years erode the profits of industries that took decades to build, hence the understandable claim that firms are born and die faster than ever before.[58]

Yet as plausible as the perception may be that the pace of business in America is accelerating, *The Economist*'s judgment at the end of 2015 was that "hard evidence of a great acceleration is hard to come by."[59] In support of this conclusion, *The Economist* highlighted criteria for measuring the speed of American business in days and then compared today's values with those of five and ten years earlier. The alternative criteria were as follows (with the percentage change between ten years ago and the latest value shown in parentheses):

55 "The Creed of Speed," *The Economist*, December 5, 2015, p. 22.

56 Ibid., p. 13.

57 "Hyperactive, Yet Passive," *The Economist*, p. 13. David Parkins' drawing also adorned the cover of *The Economist*'s December 5, 2015, issue.

58 "The Creed of Speed," *The Economist*, p. 22.

59 Ibid., p. 23.

- The duration of new corporate bonds (+98.8%)
- The tenure of departing CEOs (+6.5%)
- The duration of corporate bonds (+18.2%)
- Current CEO tenure (+46.3%)[60]
- The job tenure of people over 25 years of age (+12.2%)
- Private-sector job tenure (+17.1%)
- Mutual-fund holding period (+34.7%)
- The holding period of Standard & Poor's (S&P) 500 shares (–6.3%)
- Manufacturers' inventory days (+43.8%)
- S&P 500 inventory days (+1.7%)

In only one of these ten cases—the holding period of S&P 500 shares—was the latest value in days greater than it had been ten years ago, and then by only 6.3 percent.[61] Nine out of ten of *The Economist*'s criteria for quantifying change in the pace of American business over the last decade do not show any speeding up.

Obviously, these ten criteria are not the only metrics that could be put forward to assess the pace of twenty-first-century business in America. They were chosen because data existed to quantify them in a common measure: days. Certainly, there are other aspects of business today that go faster than they did a decade ago. The speed at which ideas zip around the globe has increased, and the time lag between the introduction of a new technology in an advanced economy and its adoption by others has shortened.

> But many other measures suggest sloth, not celerity. The rate of new consumer-product launches is probably slowing or in decline. Factories do not seem to be making things faster. A crude gauge of production speed can be gained by looking at the inventories of industrial firms, which mainly comprise half-finished goods, or "work-in-progress." The ratio of work-in-progress to sales points to a slowdown over the past decade (though if you exclude Boeing, an aircraft-maker, it is merely flat). And there is no obvious evidence that outsourced production overseas differs in this respect. At Hon Hai Precision, also known as Foxconn, which makes iPhones and other gizmos in China, things have gone the same way. [62]

60 Current CEO tenure compared the latest value with 2007.

61 "The Creed of Speed," *The Economist*, p. 23. The interactive chart that appeared in the print edition of this article from which the percentages were taken does not appear in the online version.

62 Ibid., p. 23.

Similar conclusions emerge from considering inventory levels in the age of smart supply chains, the tenure of private-sector workers, and the turnover of firms.

> If products were zipping through smart supply chains faster, you would expect the overall level of inventories to fall. But in 2014 big listed American firms held 29 days of inventory, only slightly less than in 2000. For the economy as a whole, inventory ratios improved in the 1990s but have deteriorated sharply since 2011. And just as the stuff that is sold may not be turning over any more quickly, neither are the people who make it. The median private-sector worker has held his job for 4.1 years, longer than in the 1990s. There has been a slight decline in the tenure of older men, but a slight lengthening for women.
>
> More creative destruction would seem to imply that firms are being created and destroyed at a greater rate. But the odds of a company dropping out of the S&P 500 index of big firms in any given year are about one in 20—as they have been, on average, for 50 years. About half of these exits are through takeovers. For the economy as a whole, the rates at which new firms are born are near their lowest since records began, with about 8% of firms less than a year old, compared with 13% three decades ago. Youngish firms, aged five years or less, are less important measured by their number and share of employment.[63]

The selection of analytic criteria to address a question such as whether the pace of American business is faster than before (or even accelerating) is, at the end of the day, a matter of considerable judgment. There are no mechanical rules for deciding which criteria to pay attention to and which to downplay or ignore. In the case of dueling lists of analytic criteria, there appears to be no way to conclusively weigh one criterion in relation to others. These are all conclusions encountered earlier in this report. But the case of using dueling criteria to analyze a hypothesis reflecting the conventional wisdom of American CEOs about the pace of business lends added weight to these recurring limitations.

Lanchester, Casualties, Force Ratios, and WEI/WUV

In broaching the question of whether analytic criteria can be found that appear to be recurring or predominant drivers of engagement outcomes, it is useful to note the immensely greater difficulties of designing and implementing effective strategies as opposed to effective tactics. In reflecting on the American failure in Vietnam, Harry Summers acknowledged the enormous gulf between tactical superiority and strategic success:

> On the battlefield itself, the [American] Army was unbeatable. In engagement after engagement the forces of the Viet Cong and the North Vietnamese Army were thrown back with terrible losses. Yet, in the end, it was North Vietnam, not the United States, that emerged victorious.[64]

63 Ibid., p. 23.

64 Harry G. Summers, Jr., *On Strategy: A Critical Analysis of the Vietnam* War (Novato, CA: Presidio Press, 1982), p. 1.

Of course, as the mauling of the 7th Cavalry in the Ia Drang Valley by the North Vietnamese Army in late 1965 reveals, the U.S. Army did not win every battle in Vietnam. But by Operation Desert Storm in 1991, the tactical dominance of U.S. conventional forces had become vastly more pronounced than it had been in Vietnam, and this overwhelming predominance continued in Afghanistan and Iraq after al Qaeda's attack on the World Trade Center and Pentagon on September 11, 2001.

Nevertheless, major strategic blunders can lead to strategic failure no matter how competently a military force performs at the tactical level. In the case of the U.S.-led invasion of Iraq in 2003, George W. Bush's administration set out to topple Saddam Hussein and establish a democracy in place of his Baathist dictatorship. Not only did the task turn out to be "hugely more difficult" than the administration had imagined, but the Pentagon under Donald Rumsfeld insisted on going in with too few troops to maintain security after the Iraqi army was defeated.[65] And because Operation Iraqi Freedom (OIF) opened a second front before the post-9/11 war in Afghanistan had been resolved, David Kilcullen has a point in describing the decision to invade Iraq with just over 200,000 troops as the "greatest strategic screw-up" since Hitler's decision in 1941 to invade Russia without having first defeated Great Britain.[66] OIF's initial blunders in planning were followed by others in execution, including Paul Bremer's decisions to disband the Iraqi army and dismiss most civil servants after the Iraqi army had been defeated and Baghdad occupied. "About 400,000 unemployed and angry young men, many still armed, were loosed on Iraq towns, and Iraq had no security force except the undersized coalition military."[67] Predictably things rapidly fell apart: by November 2006 the violence in Iraq had massively escalated, and U.S. public support for the occupation had plummeted.[68] President Bush then took direct charge of the war and put General David Petraeus in command. Not until this point did U.S. forces begin implementing an effective counterinsurgency strategy against al Qaeda.

This sad history illustrates the complexities of building consistent tactical level successes into overall strategic success. It reinforces Antulio Echevarria's lament in 2004 that, "The American way of war tends to shy away from thinking about the complicated process of turning military triumphs, whether on the scale of major campaigns or small-unit actions, into

65 William J. Perry, *My Journey at the Nuclear Brink* (Stanford, CA: Stanford University Press, 2015), p. 173.

66 David Kilcullen, *Blood Year: The Unraveling of Western Counterterrorism* (Oxford: University of Oxford Press, 2016), pp. 14, 16. According to Kilcullen, Rumsfeld wanted to limit the number of American troops for the invasion and post-conflict stabilization to 75,000, whereas Pentagon war planners estimated that the real requirement was around 400,000. Ibid., p. 13.

67 Perry, *My Journey at the Nuclear Brink*, p. 174.

68 Kilcullen, *Blood Year*, p. 35.

strategic successes."[69] It also suggests that there may be more regularity to tactical interactions than to overall strategic outcomes. If so, then this greater tactical regularity could point to analytic criteria that more often than not drive engagement outcomes. Whether such criteria, if they can be found, are quantifiable or offer any ability to predict likely battle outcomes remains to be seen.

In 1914 Frederick Lanchester developed a simple, mathematical model of combat driven by the rates at which the opposing forces could inflict attrition or casualties on the other using aimed fire with modern long-range firearms. Lanchester proposed that each side's losses over the course of the battle would be proportional to the number of opposing shooters times a coefficient reflecting their effectiveness. If B is the numerical strength of the Blue force and β its effectiveness coefficient, and R is the strength of the Red force and ρ its effectiveness coefficient, then the attrition process for the Blue force is:

$$dB/dt = -\rho R$$

and that for the Red force is given by: [70]

$$dR/dt = -\beta B$$

When two forces have equal fighting strength, these two equations yield a relationship known as Lanchester's "square law," so called because of the exponents on B and R:

$$\rho R^2 = \beta B^2$$

"In other words, the fighting strengths of the two forces are equal when the *square of the numerical strength multiplied by the fighting value of the individual units are equal*."[71]

Lanchester's square law rested on a number of assumptions. It assumes that sides have similar firearms and defend themselves by concentrating aimed fire on the opponent; that both forces are within weapons range of each other; that the effects of weapons rounds are independent; that fire is uniformly distributed over enemy targets; that the attrition coefficients β and ρ are known and constant; that both sides are fully committed at the battle's outset; and that neither Blue nor Red has reinforcements.[72] With these assumptions in mind, Lanchester used his model to demonstrate the importance of concentration in modern warfare. In older battles, short-range weapons like swords and battleaxes meant that the numbers of men engaged at

69 Antulio J. Echevarria II, *Toward an American Way of War* (Carlisle, PA: Strategic Studies Institute, U.S. Army War College, March 2004), p. vi. For an excellent survey of U.S. strategic errors and misperceptions under the administration of Barack Obama, see Mac Thornberry and Andrew F. Krepinevich, "Preserving Primary: A Defense Strategy for the New Administration," *Foreign Affairs*, September/October 2016, pp. 26–27.

70 F. W. Lanchester, *Aircraft in Warfare: The Dawn of the Fourth Arm* (London: Constable & Company, 1916), p. 42.

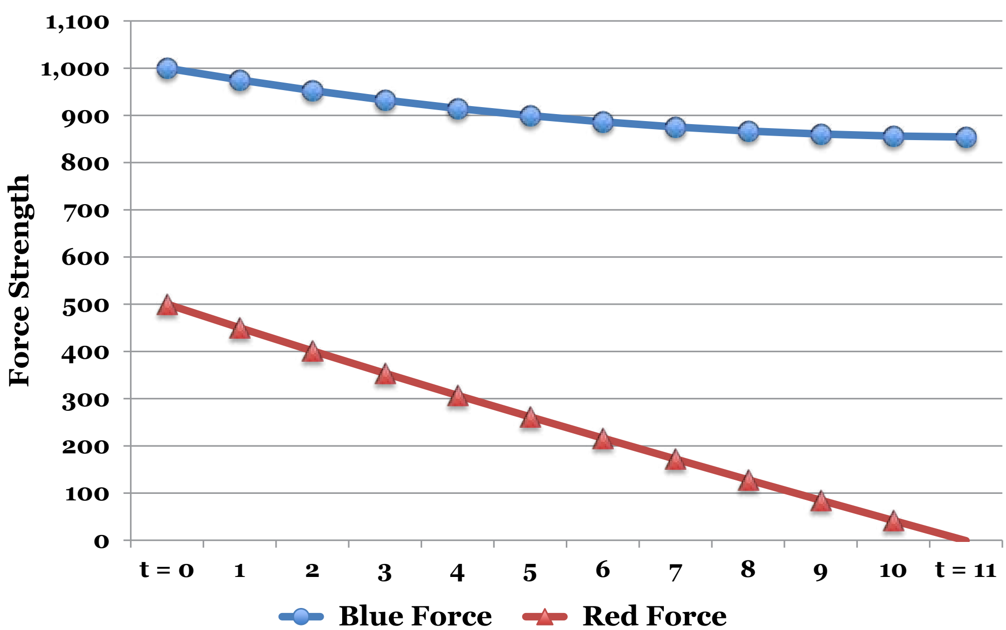
71 Ibid., p. 48. Italics in the original.

72 Ronald L. Johnson, *Lanchester's Square Law in Theory and Practice* (Fort Leavenworth, KS: School of Advanced Military Studies, U.S. Army Command and General Staff College, 1989/1990), p. 9.

any moment would be roughly the same on both sides, however unequal the two sides were overall. Assuming the fighting skills and weaponry of the opposing forces to be roughly equal, both sides would experience similar rates of attrition. Battles in which both sides experience similar rates of attrition have come to be known as following Lanchester's "linear law."

But with long-range firearms able to bring aimed fire to bear throughout the enemy's ranks, a numerically inferior force would find itself "under far heavier fire, man for man, than it is able to return."[73] Assuming $\beta = \rho = 5\%$, if Blue initially has 1,000 men and Red 500 (at t = 0), then the battle would unfold as shown in Figure 13. From t = 0 to t = 10, the Blue/Red force ratio would go from 2:1 to over 20:1, and by t = 11 the Blue force would annihilate the Red force with a loss of only 145 men. This result would leave Blue with enough men (855) to annihilate a second Red force of 500 men: hence the merits of concentration under modern conditions.

FIGURE 13: CONCENTRATION UNDER MODERN CONDITIONS[74]



The question to ask at this juncture is whether empirical data from historical battles support Lanchester's presumption that combat outcomes are fundamentally driven by attrition. Since the 1950s a number of operations researchers have addressed this question. In 1954, for example, J. H. Engel concluded that a Lanchesterian attrition model reasonably fit the data from the battle of Iwo Jima.[75] But most studies since Engel's have "generally suggested that

73 Lanchester, *Aircraft in Warfare*, pp. 40–41.

74 Lanchester, *Aircraft in Warfare*, p. 42.

75 Johnson, *Lanchester's Square Law in Theory and Practice*, pp. 11–12.

Lanchester equations may be poor descriptors of large battles extending over periods during which forces are not constantly engaged in combat.[76] However, the most telling critique of Lanchester's model was published by Robert McQuie in 1987.

McQuie based his critique on data from 80 defeats (49 defender defeats and thirty-one attacker defeats) that took place in World War II and the 1967 and 1973 Arab–Israeli wars. These battles were located in the Far East, the Near East, and Eastern and Western Europe; all were combined-arms engagements involving infantry and artillery and, in most cases, armor and aircraft. "The median attack in this data involved about 22,000 troops, while the median defeat involved about 12,000. In the tank battles that were examined, a median of about 360 armored vehicles was employed against about 260. The median battle resulted in 20 hours of combat over two days."[77]

In analyzing these 80 battles, McQuie first considered whether defeat had been due to the proportion or total number of troops lost.

> In one-half of the battles, an attack was recognized as a failure when casualties had reached less than four percent of the attacking force. The median defense was abandoned when casualties of the defending force reached less than eight percent. When combat was broken off, however, casualties ranged widely; some forces admitted defeat when their losses were negligible, while other fought to the last man.[78]

Next, McQuie considered the possibility that the rate at which casualties were incurred might have triggered acceptance of defeat. But here, too, the data from the 80 battles suggested otherwise. Casualties per hour of actual combat excluding lulls in the fighting of at least one hour revealed that the median attack was recognized as unsuccessful with casualty rates of less than one-fifth of 1 percent an hour, and the median defense was recognized as a failure when casualties reached approximately two-fifths of 1 percent an hour. Nor did median casualty exchange ratios appear to drive defeat: the median exchange ratio at which an attacker terminated a battle was about 2-to-1, while for defenders the median ratio was about 0.8-to-1. Like

76 Ibid., p. 21. Johnson's monograph reviews the work on verifying Lanchester's model by J.H. Engel (1954), Herbert K. Weiss (1957, 1966), Daniel A. Willard (1962), Robert L. Heimbold (1961, 1971), and Janice B. Fain (1974). For a later assessments from 1977 and 1985, see Trevor N. Dupuy, *Numbers, Prediction and War: The Use of History to Evaluate and Predict the Outcome of Armed Conflict* (Fairfax, VA: HERO Books, 1985), pp. 148–150. Wohlstetter's judgment, too, was that Lanchester's equations were neither a universal law governing all combat nor a great help in predicting the outcomes of classical wars between large armies. Wohlstetter, "Theory and Opposed-System Design," in Zarate and Sokolski, *Nuclear Heuristics*, p. 134.

77 Robert McQuie, "Battle Outcomes: Casualty Rates as a Measure of Defeat," *Army*, November 1987, p. 32. The battles McQuie examined were taken from the database Trevor Dupuy's Historical Evaluation and Research Organization (HERO) prepared under contract for the Army's Concepts Analysis Agency (CAA). See HERO, *Analysis of Factors That Have Influenced Outcomes of Battles and Wars: A Data Base of Battles and Engagements*, Vol. VI, *World War II, 1939–1945; Campaigns in France 1940, on the Eastern Front, and of the War against Japan; The 1967, 1968, and 1973 Arab-Israel Wars* (Dunn Loring, VA: HERO, June 1983). In 1985 HERO completed compilation of a land warfare database that included significant statistics, environmental circumstances, and operational facts related to 603 land and air-land battles and engagements from 1600 through 1973. Trevor N. Dupuy, "Understanding War from Historical Perspective," *Marine Corps Gazette*, June 1985, p. 53. Some inconsistencies or gaps in the data for eight battles reduced the data set to 595.

78 McQuie, "Battle Outcomes: Casualty Rates as a Measure of Defeat," p. 32.

the casualty totals and proportions, some forces gave up with less than one twentieth of their opponent's casualties, whereas others took more than ten times the casualties of the opponent before admitting defeat.[79]

TABLE 8: REASONS FOR ABANDONING AN ATTACK OR DEFENSE[80]

Envelopment, encirclement, penetration	33%	No reserves left	12%
Adjacent friendly unit withdrew	13%	Supply shortages	2%
Enemy occupied key terrain	6%	Truce or surrender	6%
Enemy achieved surprise	8%	Change in weather	2%
Enemy reinforced	4%	Orders to withdraw	2%
Maneuver by the Enemy	**64%**	**Other reasons for defeat**	**24%**

Casualties, equipment losses	10%
Heavy enemy artillery/air attacks	2%
Attrition from enemy firepower	**12%**

McQuie's conclusion was that over the last 50 years, casualties had not often been the reason battles had been lost.[81] While most defeats had multiple causes, in 52 of the 80 battles he examined, the most likely reasons could be identified. They are shown in Table 8. Maneuver by the enemy accounted for some 64 percent of the defeats, whereas casualties and equipment losses inflicted by enemy firepower only led to the acceptance of defeat in 12 percent of the battles. Consequently, McQuie was led to question whether casualties, however measured, had ever played the decisive role posited by Lanchester. The casualties, he wrote,

> incurred by losing forces appear to have been fewer than usually envisioned by those of us concerned with either command of troops or analysis of war. It appears as well that Mr. Lanchester's equations present a drastic misstatement of what drives the outcome of combat. The evidence indicates that in most cases, a force has quit when its casualties reached less than ten percent per battle. In most battles, moreover, defeat has not been caused by casualties.[82]

In 1993 McQuie published a parallel analysis of the role played by force ratios in driving battle outcomes. In this instance, he looked at 571 battles in the Land Warfare Database that Trevor Dupuy's Historical Evaluation and Research Organization (HERO) had produced for the U.S. Army's Concepts Analysis Agency (CAA). The engagements in this database initially spanned

79 Ibid.

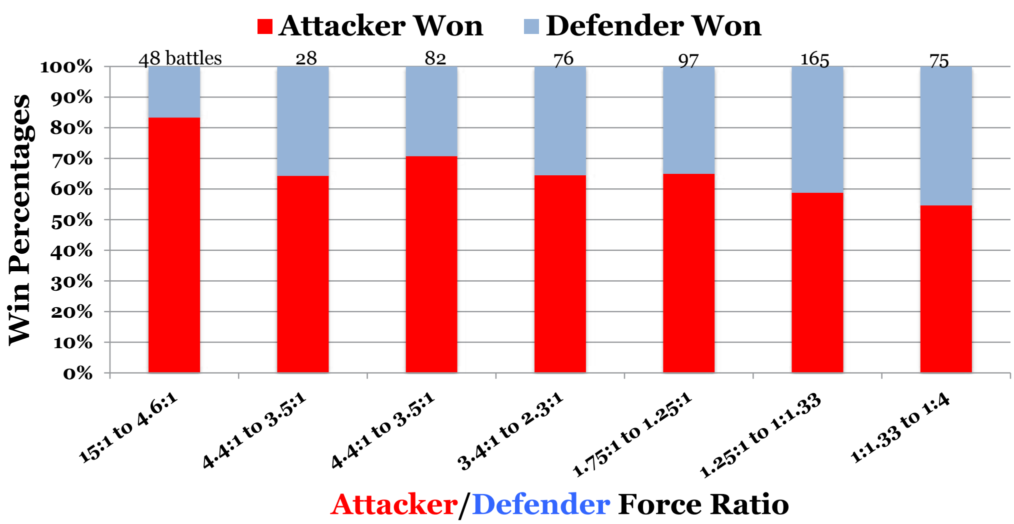
80 Ibid., p. 34. In discussing this table with McQuie in 1994, he indicated that the 64 percent of the defeats due to enemy maneuver had constituted 53 of 52 battles. Based on the percentages in Table 8, however, it appears more likely they were based on 51 of 80 battles.

81 Ibid., p. 33.

82 Ibid., p. 34.

1600 to 1973[83] and the force ratios McQuie considered appear to have been based on the ratios of the numbers of troops on each side. He did note, however, that in predominately armored battles, the ratio of tanks on the opposing sides could also be used as a baseline force ratio.

FIGURE 14: FORCE RATIOS AND COMBAT OUTCOMES[84]



As with casualty rates, the historical data did not reveal the kind of correlation between force ratios and outcomes one would expect if force ratios drove outcomes. If they did, then as the attacker/defender force ratio shifted in the defender's favor, one would expect the percentage of attacker wins to decline steadily and substantially. But it did not decline in HERO's 571 battles. For example, the percentage of successful defenses at force ratios of 4-to-3.5:1 in favor of the attacker is 36 percent, but this only increases to 45 percent with force ratios of 1:1.33-to-4 in favor of the defender. Granted, if the attacker's force ratio advantage is at least 4.6-to-1 (and as high as 15-to-1), the percentage of successful defenses drops to only 17 percent. But, as McQuie observed, attackers have lost even with highly favorable force ratios and won despite highly inferior force ratios.[85]

The implication of McQuie's 1987 and 1993 analyses, then, would appear to be that casualties and force ratios very rarely drive engagement outcomes. They are certainly not reliable predictors of likely combat outcomes. Nevertheless, there has been a strong inclination among operations analysts, especially those devising mathematical models of combat, to rely on force

83 By 1987 the database had added the 1982 war in Lebanon. Trevor N. Dupuy, *Understanding War: History and the Theory of Combat* (New York: Paragon House, 1987), pp. 237–250.

84 Robert McQuie, "Force Ratios," *Phalanx*, June 1993, p. 27. McQuie provided the data underlying this figure to the author.

85 McQuie, "Force Ratios," p. 28.

ratios. Indeed, in every combat model known to McQuie as of 1993, he had found key algorithms built around force ratios.

In 1972 the Army's deputy director for plans tasked the Concepts Analysis Agency to develop a static firepower scoring methodology—Weapon Effectiveness Indices/Weighted Unit Values (WEI/WUV)—that could provide "useful approximations of the relative value of forces when viewed in the context of theater-level analysis" and wargaming.[86] The underlying concern at the time was to improve understanding of the conventional force balance between the North Atlantic Treaty Organization (NATO) and Warsaw Pact (WP) in Central Europe. CAA produced three versions of this methodology—the final one, WEI/WUV III, appearing in November 1979.

The hope was that WUV aggregations at the theater level, corps level, or lower levels in Central Europe would be more comprehensive approximations of the relative combat potential of opposing forces than would ratios of the numbers of opposing troops or tanks. But the WEI/WUV methodology, like GDP, turned out to involve considerable complexity. The first step in developing the WEI/WUV methodology was to devise weapons effectiveness scores (WEIs) for the various categories of weapons in NATO and Warsaw Pact ground forces. The methodology identified nine weapons categories: (1) tanks, (2) attack helicopters, (3) infantry fighting vehicles (IFVs), (4) armored personnel carriers (APCs), (5) air defense weapons, (6) anti-tank weapons, (7) artillery and rockets, (8) mortars, and (9) small arms. Next, in each category, a reference weapon had to be selected to which all other WEI scores in the category would be normalized. In the case of small arms (point-and-fire weapons below 20 millimeters in caliber), the standard was the U.S. M-16A1 rifle.

With the weapons categories defined and reference weapons selected, CAA then developed formulas for calculating WEIs in each category. For example, the WEI for APCs was calculated using the basic formula:

$$0.3F + 0.3M + 0.15S + 0.25T$$

86 CAA, *Weapon Effectiveness Indices/Weighted Unit Values (WEI/WUV)*, Vol II, *Basic Report*, CAA-SR-73-18 (Fort Belvoir, VA: U.S. Army, April 1974), p. XVI-1.

where *F* was a firepower index, *M* a mobility index, *S* a survivability index, and *T* a troop-carrying index. While the troop-carrying index was simply the number of troops (including crew) that an APC could carry divided by that of the reference APC (the M113A1 with an M2 machine gun), the indices *F*, *M*, and *S* each were defined by other formulas involving various characteristics. For instance,

$$F \text{ (firepower index)} = 0.85\ (\Sigma FW/\Sigma FW_s) + 0.15\ (P/P_s)$$

where *FW* is the firepower index of an automatic weapon on the APC being scored; ΣFW and ΣFW_s are the sums of the firepower indices for all the automatic weapons on the APC being scored and the APC standard (the M113A1); and *P* and P_s are the respective porthole factors. *FW*, in turn, is given by another formula involving indices for the volume of fire, effective range, and dependability of the APC's automatic weapons.[87] In CAA's first version of WEI/WUV, the mobility factor, *M*, for APCs had no less than 11 additive terms, all of which were measurable in inches, degrees, kilometers, etc. Frontal armor thickness, for example, could be measured in inches. And to reiterate the complexity of all this, the underlying WEI characteristics that could not be measured were estimated by panels of military experts using a Delphi technique to achieve consensus. WEI scores, therefore, were highly aggregated measures of the combat potential of individual weapons in each of nine categories.

87 CAA, *Weapon Effectiveness Indices/Weighted Unit Values (WEI/WUV)*, Vol II, *Basic Report*, p. V-2.

TABLE 9: ILLUSTRATIVE WUV CALCULATION FOR A U.S. DIVISION[88]

Category	Quantity in Unit	WEI	Quantity x WEI	Category Weight	WUV
Tanks					
M60A3 Patton	150	1.11	167		
M1 Abrams	150	1.31	197		
Tank Total	300		363	94	34,122
Attack Helicopters					
AH-1S Cobra	21	1.00	21		
AH-64 Apache	18	1.77	32		
Helicopter Total	39		53	109	5,762
Air Defense Weapons					
20mm Vulcan M741	24	1.00	24	56	1,344
Infantry Fighting Vehicle					
M2 Bradley IFV	228	1.00	228	71	16,188
Armored Personnel Carriers					
M113 APC	500	1.00	500	30	15,000
Artillery					
155mm Howitzer	72	1.02	73		
8-inch Howitzer	12	0.98	12		
Multiple Launch Rocket System	9	1.16	10		
Artillery Total	93		96	99	9,468
Antitank Weapons					
TOW Missile Launcher	150	0.79	119		
Dragon Launcher	249	0.69	172		
Light Antitank Weapon (LAW)	300	0.20	60		
Antitank Total	699		350	73	25,573
Mortars					
81mm	45	0.97	44		
107mm	50	1.00	50		
Mortar Total	95		94	55	5,151
Small Arms					
M16 Rifle	2,000	1.00	2,000		
Machine Guns	295	1.77	522		
Small Arms Total	2,295		2,522	4	10,089
Division Total					**122,696**

88 Congressional Budget Office (CBO), *U.S. Ground Forces and the Conventional Balance in Europe* (Washington, DC: GPO, 1988), p. 15. TOW stands for Tube-launched, Optically tracked, Wire-guided.

To be able to compare the combat potential of opposing ground forces in Europe, the next step was to multiply the numbers of various weapons in NATO or Warsaw Pact units by their corresponding WEI scores. However, since individual weapons categories might have different values depending on whether a unit or force was attacking or defending, CAA again used a Delphi technique to develop offensive, defensive, and average category weights.[89] Thus, the "Quantity x WEI" values in Table 9 have to be multiplied by their corresponding category weights to get a WUV total in each category. Then, by summing WUV values across weapons categories, a total WUV value could be calculated for a unit or force. Table 9 shows the results of applying WEI/WUV III to a notional U.S. division in the late 1970s with a mixture of older and newer weaponry.

To be stressed is that the Delphi methodology CAA used to develop WEI/WUV category weights was subjective. It amounted to structured consensus-building among experts in land warfare as to the relative contributions of the nine weapons categories in various situations and circumstances. In historical cases such as the Battle of the Atlantic or the rise of the West, quantitative category weights simply do not exist. Hence the impossibility of weighing the contributions of Ultra in explaining the U-boats' defeat in May 1943 or of determining the role cultural factors played in the West's escape from poverty is revealed.

Despite the considerable effort invested in WEI/WUV during the 1970s, the insights into the conventional force balance in Central Europe that the methodology offered were limited. To compare the WUV total in Table 9 with Soviet tank or motorized rifle divisions, one also needed to develop WEIs and category weights for Soviet equipment and then apply them to the tables of organization and equipment (TO&E) for Soviet units. Doing so showed the Warsaw Pact gaining some ground in Central Europe from 1965 to 1977. During that period the WUV ratio of Warsaw Pact to NATO ground forces deployed in Central Europe grew from 1.5-to-1 to 1.85-to-1.[90] But if force ratios at any level are not reliable predictors of combat outcomes, then the most that can be inferred from the growth in the Warsaw Pact's force ratio advantage from 1965 to 1977 is that the trend in NATO and Warsaw Pact ground forces was adverse to NATO. It is far from obvious, however, what this trend might mean for, say, the adequacy of conventional deterrence. Among other reasons, both sides had come to believe by the late 1970s that NATO enjoyed a substantial edge in tactical air power. Thus, the WUV ratio of Warsaw Pact and NATO ground forces omitted a major component of the overall conventional force balance in Central Europe.

89 CAA, *Weapon Effectiveness Indices/Weighted Unit Values (WEI/WUV)*, Vol. II, *Basic Report*, pp. XIII-1, XIII-5.

90 The Office of Net Assessment (ONA), Office of the Secretary of Defense, *The Military Balance in Europe: A Net Assessment* (Washington, DC: DoD, March 1978), p. 52. Central European NATO and Warsaw Pact ground forces included those deployed in Belgium, the Netherlands, Luxembourg, West and East Germany, Poland, and Czechoslovakia. Army Lieutenant Colonel Peter Bankson wrote this assessment.

By the mid-1980s other static scoring systems had been developed. One was the Analytic Sciences Corporation's TASCFORM methodology (Technique for Assessing Force Modernization), which had initially focused on scoring aircraft rather than ground forces. Another was Trevor Dupuy's Quantified Judgment Model (QJM). To explore the differences between these methodologies, Andrew Marshall's Office of Net Assessment initiated an effort in 1986 with the Central Intelligence Agency to apply these methodologies to actual NATO and Warsaw Pact force data over the period 1965–1985. The goal was to see if an agreement could be reached on the adequacy and validity of these various measures of effectiveness (MOEs). Before the project was even completed, however, Marshall offered this assessment:

> [I]t is very clear from past experience in comparing forces that no MOE will solve the net assessment problem. We have found them useful in understanding long-term trends, but they clearly fail as predictors of outcomes of military operations between opposing forces because they reflect only a few dimensions. In particular, historical cases tend to show that other variables often dominate.[91]

In 1992, Ed Vandiver, the director of CAA, offered an even harsher judgment of WEI/WUV and other popular American static scoring methodologies:

> Both the weapon scores and the unit/force scores made from them are undefined. Soviet Combat Potential Scores at least had the virtue of being precisely defined (we think) as equivalences. The original Army Firepower Scores of the 1950's had a similar definition with the addition of some cautions about the bounds within which the substitutions could be made. This was lost with the development of Firepower Potentials in the late 1960's and with WEI/WUV's in the early 1970's. I claim that lacking a precise definition, no one knows the meaning of any of the current crop of static scores. This leaves the user free to impute any meaning desired, which is the equivalent of the Red Queen saying that words meant whatever she wished them to mean.[92]

Starting with Lanchester's laws, this discussion set out to look for regularities or patterns in tactical engagements that might point to drivers of outcomes in at least many, if not most, cases. The conclusion that emerges is not simply that care and judgment must be used in choosing analytic measures, but that one can easily be misled by making poor choices, however attractive and apparently objective metrics like casualty rates and force ratios may seem at first blush. Nevertheless, just as being outmaneuvered led to the acceptance of

91 Andrew W. Marshall, "Joint MOE (Measures of Effectiveness) Project," memorandum for Thomas Armour, CIA/ASG, February 20, 1986, p. 2.

92 E. B. Vandiver, letter to Andrew W. Marshall, April 9, 1992, p. 1. Regarding Soviet combat potential scores, Vandiver wrote: "Soviet analysts were willing to make equivalences at which most U.S. analysts with operational experience balked. For example, "If an artillery piece was to be taken as the initial unit of measurement, a tank would be worth 4 to 5 units, a helicopter from 6 to 9 units, a fixed-wing attack aircraft from 10 to 12 units." See Viktor Tatarnikov, "TASS news conference, Moscow, March 29, 1989," FBIS-SOV-89-060, March 30, 1989, p. 2. Similarly, if the combat potential of a T-72 is equal to one, then the contribution of an "F-16 aircraft to a battle" is "equal to 18." Vitaly Tsygichko, "Combat Potential Method for Vienna Talks," Военный вестник [*Military Herald*], March 1989, p. 8. For examples of Soviet combat potential scores and commensurability coefficients see Steven Zaloga, "Soviets Denigrate Their Own Capabilities," *Armed Forces Journal International*, July 1991, pp. 18, 20.

defeat in the majority of the 80 battles McQuie examined in 1987, air-to-air engagements going back to the Vietnam War suggest an even more dominant causal regularity: situation (or situational) awareness (SA).

Situation Awareness

Strong evidence for the role of situation awareness in the outcomes of around 80 percent of air-to-air engagements emerged during the final months of the Vietnam War. From December 18, 1971, to January 12, 1973, there were 112 "decisive" engagements between U.S. and North Vietnamese aircraft. Decisive engagements were those in which at least one U.S. or North Vietnamese aircraft was lost. Project Red Baron III's meticulous reconstructions of the 112 "decisive" engagements credited Air Force and Navy aircrews with downing 75 MiGs[93] while North Vietnamese MiG pilots were credited with 37 American air-to-air losses.[94] In 60 percent (67 of 112) of the engagements, the losing crews were not aware of being under attack prior to enemy ordnance hitting their aircraft; in another 21 percent (24 of 112) of the engagements, the victims became aware of the attack by hostile fighters too late to initiate adequate defensive actions.[95] Taking situation awareness to be the relative "capability of opposing aircrews to develop and sustain accurate representations of where all the friendly and enemy aircraft in or near the combat arena are, what they are doing, and where they are likely to be in the immediate future," then in 81 percent of the decisive engagements (91 of 112), the loss of SA was the primary cause of being shot down.[96]

During 1971–1973 all of the 75 American kills were by Air Force, Navy, and (one) Marine Corps F-4s. In most cases, the U.S. F-4 aircrews were within visual range of the MiGs by the time they fired their guns or missiles, which meant that they did not take advantage of the F-4's potential for beyond visual range (BVR) shots. The F-4 had originally been designed by the U.S. Navy as a long-range, all-weather, fleet interceptor. To provide air defense against attacking Soviet aircraft, the F-4 was equipped with a powerful Westinghouse radar and, as primary armament, Raytheon's AIM-7 Sparrow III, a semi-active, radar-guided air-intercept missile (AIM). By the early 1960s, there was a growing expectation among engineers and technologists that the F-4's technical potential for BVR engagements would supplant hard maneuvering dogfights within visual range (WVR). But for a number of reasons—rules of engagement, the aversion of American aircrews to risking fratricide ("Blue on Blue" kills), and the unreliability of the Vietnam-era AIM-7D, E, and E-2 Sparrows—air combat in the skies of

93 MiG is the acronym for Mikoyan-Gurevich, the design bureau that produced the MiG-17s, MiG-19s, and MiG-21s that opposed U.S. aircrews during the Vietnam War. Artiom Mikoyan and Mikhail Gurevich established their famous aircraft design bureau in 1939.

94 *Project Red Baron III: Air-to-Air Encounters in Southeast Asia,* Vol. III, Part 1, *Tactics, Command and Control, and Training*, (unclassified) (Nellis Air Force Base, NV: U.S. Air Force Tactical Fighter Weapons Center, June 1974), p. 66.

95 Ibid., p. 61.

96 S. R. Dvorchak, "On the Measurement of Fog," slide presentation to the Military Operations Research Society, Washington, DC, June 1986, slide 9.

North Vietnam witnessed only two front-aspect BVR kills from 1965 to 1973.[97] The expectation that air-to-air combat would eventually be dominated by BVR engagements with radar-guided missiles would not be tested empirically until the early 1980s.

Meanwhile, further evidence emerged in the ACEVAL (Air Combat Evaluation) test confirming the dominance of SA in WVR dogfights. ACEVAL was flown in 1977 on an instrumented air combat maneuvering range at Nellis Air Force Base, Nevada. ACEVAL's aims were to explore the influence of initial conditions (Blue ground controlled intercept, or GCI, advantage; neutral; or Red GCI advantage) and force ratios (from 1-versus-1 to 4-versus-4) on engagement outcomes.[98] ACEVAL's Blue Force consisted of F-15s and F-14s "armed" with the ability to simulate firing 20mm cannons, AIM-9Ls, and AIM-7Fs; the Red Force flew F-5Es (simulating MiG-21s) "armed" with 23mm cannons and AIM-9Ls.[99] GCI battle management resembled that provided by Red Crown over UHF radios during the Vietnam War and proved difficult to sustain as the number of participants increased. ACEVAL's most dramatic departure from earlier air combat experience was equipping both sides with the all-aspect AIM-9L Sidewinder. The result was "a point-shoot war" in which the "ability to kill someone in the face made a lot of difference" compared with operations in 1971–1973 when only U.S. F-4s had a front-aspect capability with the Sparrow and were very rarely able to exploit it.[100]

Regarding SA, ACEVAL suggested that, in the long run, numbers still counted because people goon it about as often as they do something brilliant, thereby canceling themselves out. The same appeared to be true of incremental improvements in weaponry. In the short term, however, numbers only accounted for "about 10 to 20% of the variation in results. The rest of the variation . . . [was] due to the things that people do," which again pointed to SA.[101]

ACEVAL did not explore the impact of long-range BVR missiles on SA. The Navy F-14s were not allowed to employ their long-range AIM-54 Phoenix missile, which had a maximum range in excess of 100 nautical miles. In fact, the rules of engagement for ACEVAL required visual identification of the target prior to weapons employment.[102] Thus, the prospect that long-range missiles would come to dominate air-to-air combat was not addressed in ACEVAL beyond the suspicion that it is hard to "out-technology the other guy," because people adapt.[103] The

97 James Burton, "Letting Combat Results Shape the Next Air-to-Air Missile," unclassified briefing, January 1985, Slide 3. During the years 1971–1973 U.S. F-4 crews fired 276 Sparrows. Ibid., Slide 5.

98 E. J. Griffith, Jr., "ACEVAL: Origin, Description, Results, Applicability," briefing slides, undated, Slides 2, 7. Griffith was the Blue Force commander for ACEVAL.

99 The test matrix generated 360 engagements involving 1,488 sorties. Griffith, "ACEVAL: Origin, Description, Results, Applicability," Slide 8.

100 S. R. Dvorchak, "Getting It On in the All-Aspect Arena," *Tactical Analysis Bulletin* 79, no. 2 (Special), July 25, 1979, p. 8. Dvorchak, who was one of the principal analysts of the ACEVAL data, concluded that the all-aspect air combat precipitated by the AIM-9L was a different ball game from all previous air-to-air experience. Ibid., p 3.

101 Dvorchak, "Getting It On in the All-Aspect Arena," p. 3.

102 Griffith, "ACEVAL: Origin, Description, Results, Applicability," Slide 3.

103 Dvorchak, "Getting It On in the All-Aspect Arena," pp. 3, 18.

question of whether BVR missiles would increasingly dominate engagement outcomes was finally addressed in the 1983 AMRAAM (Advanced Medium-Range Air-to-Air Missile) OUE (Operational Utility Evaluation).

FIGURE 15: AMRAAM OUE LESSONS LEARNED[104]

- IN MULTI-SHIP ENGAGEMENTS, MANNED INTERACTIONS OUTWEIGH WEAPON AND AIRCRAFT INTERACTIONS
  - ENGAGEMENTS ARE MEN VERSUS MEN
- SITUATION AWARENESS IS THE SINGLE MOST IMPORTANT FACTOR IN DETERMINING ENGAGEMENT OUTCOMES
  - REGARDLESS OF AIRCRAFT, AVONICS, WEAPON ENVIRONMENT, OR OTHER TEST VARIABLES
- ANY IMPROVEMENT IN SITUATION AWARENESS SIGNIFICANTLY IMPROVED ALL MEASURES OF PERFORMANCE

Going into the AMRAAM OUE, the general expectation was that the new missile would unambiguously dominate engagement outcomes and the importance of human factors and interactions would be correspondingly diminished. The evaluation was "flown" in the McDonnell-Douglas simulator facility in St. Louis, Missouri. The Blue Force F-15s and F-16s were armed with AIM-9Ms, AIM-7Ms and AIM-120 AMRAAMs, the Red Force MiGs with AIM-9Ms. Including excursions, the test matrix generated around 1,200 valid trials and more than 10,000 sorties.[105]

When all was said and done, the AMRAAM OUE confirmed the dominance of situation awareness in engagement outcomes. The principal analysts of the test, S. R. "Shad" Dvorchak and Billy R. Sparks, summarized the bottom lines in a single slide (Figure 15). Their conclusion that, statistically, SA was the dominant driver by far in engagement outcomes during the AMRAAM OUE was based on the Statistical Package for Social Science (SPSS). Using SPSS to test this conclusion being wrong yielded a 0.00001 probability of it not being true.[106]

The principal reason for this outcome was that the Red Force pilots adapted. They were able to get enough information on the AMRAAM's performance from unclassified sources to be able to know what the new missile's envelope looked like, and then use of Radar

104 S. R. Dvorchak, "Man in the Loop Lessons Learned," March 26, 1985, Slide 1.

105 Barry D. Watts, *Six Decades of Guided Munitions and Battle Networks: Progress and Prospects* (Washington, DC: Center for Strategic and Budgetary Assessments, March 2007), p. 51.

106 Barry Watts, notes from a conversation with Billy Sparks, July 12, 1988.

Homing and Warning gear to enable them to know when Blue air-to-air radars were tracking them. Armed with this knowledge, the Red pilots then exploited hard maneuvering to slip outside AMRAAM envelopes when they were being targeted by F-15 or F-16 radars. This made the distances at which hard maneuvering occurred far greater than they had been even in ACEVAL, where the AIM-9L had expanded the maneuvering arena beyond what it had been in 1971–1973.

There is, of course, more than one way of achieving superior SA. Realistic combat training such as that pioneered by the Navy's TOPGUN Fighter Weapons School in 1968 and subsequently embraced by the Air Force's Red Flag exercises is one.[107] Realistic training that provided high levels of SA was unquestionably a major reason why by February 28, 1991, F-15C pilots achieved 28 Iraqi fighter kills without a single loss.[108] But one can also use technology to enhance situation awareness. During Desert Storm, the SA of U.S. pilots was enhanced by air controllers on E-3A Airborne Warning and Control System (AWACS) aircraft, whose pulse-Doppler radar could survey airspace to ranges of over 250 nautical miles. In addition to realistic training and the GCI advantage provided by the AWACS, one could also design a fighter with built-in capabilities to enhance SA. This is exactly what the Air Force did in designing the F-22. From the outset, the Air Force put "heavy emphasis on providing what the pilot required to build and maintain SA."[109] Making the plane stealthy was one component of achieving this goal. Another was equipping the F-22 with the AN/APG-77 active electronically scanned array (AESA) radar, which had a very low radar cross-section due to spread spectrum beams. The AN/APG-77 and associated onboard software provided the F-22 with automatic cueing, non-cooperative target recognition via inverse synthetic aperture processing, and agile beam steering that enabled the radar to automatically search throughout its field of view, track targets, and engage targets simultaneously. All these features gave the F-22 a "first look, first kill" capability. The result was a plane that, properly employed, gave the pilot a huge advantage in SA over previous fighters. For example, in the joint exercise Northern Edge 2006, flown in Alaska in early June, a deployment of 12 F-22s that provided the backbone of the "Blue" air-to-air force reportedly achieved a 108-to-0 exchange ratio in simulated air battles against as many as 40 opposing "enemy aircraft" (F-15s, F-16s, and F/A-18s simulating Sukhoi Su-27s and Su-30s).[110]

107 For the origins of TOPGUN, see Frank W. Ault, "The Ault Report Revisited," *The Hook*, Spring 1989, pp. 35–39. For an overview of realistic combat training see Ralph Chatham and Joe Braddock, *Training Superiority and Training Surprise: Final Report*, report briefing (Washington, DC: Defense Science Board, January 2001), available at http://www.comw.org/qdr/fulltext/dsbtraining.pdf.

108 Lewis D. Hill, Doris Cook, and Aaron Pinker, "A Statistical Compendium," Part I in *Gulf War Air Power Survey*, Vol. V, *A Statistical Compendium and Chronology* (Washington, DC: GPO, 1993), pp. 653–654.

109 Robert R. Kendall, "Re: SA Stuff," email to Barry Watts, July 27, 2007.

110 C. Todd Lopez, "F-22 Excels at Establishing Air Dominance," *Air Force News*, June 23, 2006, available at http://www.af.mil/News/ArticleDisplay/tabid/223/Article/130616/f-22-excels-at-establishing-air-dominance.aspx. Later reporting put the Blue Force air-to-air exchange ratio at 80-to-1 with the single loss apparently having been a Blue Force F-15.

Situation awareness, then, appears to offer a dominant causal criterion for understanding engagement outcomes in air combat. Whereas seemingly objective measures such as casualty rates and force ratios (however calculated) do not provide such understanding or insight, SA does. Granted, it is not always possible, much less easy, to estimate in advance how much of an SA advantage one side may have over another in a given set of circumstances even at the tactical level. Indeed, unlike casualty rates and force ratios, SA does not appear to be quantifiable. But while situation awareness may not be measurable in the sense that temperature and casualty rates are, it does appear to provide a convincing analytic criterion for understanding engagement outcomes in air combat.

CHAPTER 5

Analytic Criteria in Net Assessment and the Persistence of the Criterion Problem

Le doute n'est pas une condition agréable, mais la certitude est absurd [Doubt is not an agreeable condition, but certainty is absurd].

Voltaire, 1767[1]

. . . no theory is "total" in the sense that it deals with all possible traits of any given subject matter, and the notion of "generality" is an ambiguous one.

Albert Wohlstetter, 1967[2]

[Diagnostic net assessment] . . . is a careful comparison of U.S. weapon systems, forces, and policies in relation to those of other countries. It is comprehensive, including description of the forces, operational doctrines and practices, training regime[s], logistics, known or conjectured effectiveness in various environments, design practices and their effect on equipment costs and performance, and procurement practices and their influence on cost and lead times. The use of net assessment is intended to be diagnostic. It will highlight efficiency and inefficiency in the way we and others do things, and areas of comparative advantage with respect to our rivals. It is not intended to provide recommendations as to force levels or force structures as an output.

Andrew Marshall, 1972[3]

1 François-Marie Arouet (pen name Voltaire), Letter to Frederick II of Prussia, April 6, 1767.

2 Wohlstetter, "Theory and Opposed-System Design," in Zarate and Sokolski, *Nuclear Heuristics*, p. 153.

3 Andrew W. Marshall, "The Nature and Scope of Net Assessments," National Security Council memorandum, August 16, 1972, p. 1. Marshall served as the Pentagon's Director of Net Assessment from October 1973 to January 2015.

This report has explored a wide range of analytic criteria for judging past developments such as the rise of the West or making decisions about everything from the design and numbers of future weapons to strategic choices in a military competition. Before drawing any conclusions from these examples of what RAND economists termed the criterion problem, there is one further example that warrants discussion: diagnostic net assessment as developed and practiced by Andrew Marshall in the Pentagon from 1973 to 2015. Net assessments, it turns out, are especially sensitive to the choice of analytic criteria.[4]

Selecting Key Issues and Opportunities

In August 1972 Marshall formulated the concept of diagnostic net assessment cited at the beginning of this chapter. At the time he was heading a nascent Net Assessment Group (NAG) on Henry Kissinger's National Security Council (NSC). The NAG's aims were twofold: to monitor the reorganization of the U.S. foreign intelligence community President Richard Nixon had directed in November 1971 and to oversee national net assessments.[5] Although an initial national net assessment of U.S. and Soviet ground forces was started in September 1973, in October Marshall, his small staff, and the net assessment function were transferred to the Department of Defense (DoD) under defense secretary James R. Schlesinger.[6] Schlesinger had urged Marshall to leave the NSC and establish a net assessment capability for him in DoD. Kissinger acquiesced, and on October 13, 1973, Schlesinger, appointed Marshall, his former RAND colleague and close friend, the Pentagon's Director of Net Assessment.

In the context of the U.S.–Soviet Cold War, the aim of Marshall's Pentagon Office of Net Assessment (ONA) was to assess where the military capabilities of the United States and its allies stood relative to those of the Soviet Union and its allies in the main areas of military competition (strategic nuclear forces, the military balance in Europe, etc.). The conception of net assessment he formulated in 1972 was that of a comprehensive description of the two sides' objectives, forces, operational doctrines and practices, training, logistics, likely effectiveness in various situations, design practices (including their effects on equipment costs and quality), and the availability of resources for each side's military programs. This basic conception of the enterprise persisted without substantial revision through the end of Marshall's tenure in the Pentagon in January 2015.

One thing Marshall shied away from throughout his decades of Pentagon service was to try to turn this early conception or framework into a well-defined methodology that members of his staff could mechanically apply to develop net assessments that would be useful to senior

4 For a sympathetic account of Marshall's long career as the Pentagon's Director of Net Assessment, see Andrew F. Krepinevich and Barry D. Watts, *The Last Warrior: Andrew Marshall and the Shaping of Modern American Defense Strategy* (New York: Basic Books, 2015), especially chapters 5, 6, and 7.

5 Richard M. Nixon, "Organization and Management of the U.S. Foreign Intelligence Community," memorandum, November 5, 1971, p. 6 (declassified October 31, 2002).

6 "Study Outline for National Net Assessment of US and Soviet Ground Forces (NSSM-186)," undated, pp. 1–3.

defense officials, starting with the secretary of defense. Instead, he often compared producing net assessments with writing a doctoral dissertation. As with a Ph.D. dissertation, there was no cookie-cutter formula whose routine application could be relied upon to produce a good net assessment. ONA staff members undertaking assessments were expected to break new ground in the area being examined—to provide some new and original insights that made a significant contribution to the existing body of knowledge about a given military balance.[7] As a result, Marshall was consistently (and frustratingly to many who worked for him) reticent about trying to tell members of his staff how to assess the balance areas he had assigned them. His view was that it was far preferable for those writing net assessments to work out for themselves what the endeavor was all about and how best to proceed.

This is not to say that Marshall provided no guidance whatsoever regarding the structure and content of net assessments. In late 1976 he circulated within ONA a memorandum outlining two preferred formats for the principal Cold War balances. The key assessments at this stage were four: (1) the balance of U.S.–Soviet strategic nuclear forces, (2) the military balance in Central Europe, (3) the U.S.–Soviet maritime balance (which was surrogate for overseas power projection), and (4) the U.S.–Soviet military investment balance (which, starting around 1979, focused increasing attention on both the size of the USSR's GNP compared to that of the United States and the burden that Soviet military programs imposed on the USSR's economy). Building on the attachment to Marshall's November 1976 memo and other ONA balances and documents, the general format he recommended consisted of the following:

I. The Basic Assessment (given U.S. and Soviet objectives)

II. Long-Term Trends and Key Asymmetries

III. Key Uncertainties

IV. Emerging Strategic Problems or Opportunities[8]

7 Krepinevich and Watts, *The Last Warrior*, p. 110.

8 Marshall's 1976 guidance on structure and content did not raise trends to a top-level structural component. Instead, trends were discussed under section II. Key Asymmetries. Andrew W. Marshall, "Memo for [ONA] Staff," November 17, 1976, pp. 1, 3 of the attached format recommendations. The ambiguity is understandable in that the development of the lengthy net assessments that ONA began publishing in 1978 was still in a formative stage at this juncture.

Marshall's suggestions in each of these areas provide considerable insight into his development of diagnostic net assessment during Donald Rumsfeld's first tour as defense secretary in the 1970s

I. The Basic Assessment. A recurring challenge in the conduct of net assessments was to clarify both sides' objectives, especially those of the United States.[9] In the case of the strategic nuclear balance, the overriding American goal was to deter nuclear war. But beyond that, what did U.S. policymakers and strategists hope to achieve in the nuclear competition? What was the state of the nuclear competition and how was it being affected over time by modernization and force postures on both sides? For example, could modernization of the U.S. nuclear bomber force with the B-1 lead the Soviets to continue investing disproportionately in territorial air defenses as had long been their wont?

By the early 1970s, the Soviet Union had achieved rough parity with the United States in intercontinental nuclear arms. From 1956 to 1970 the ratio of U.S. to Soviet operational strategic launchers—deployed heavy bombers, intercontinental ballistic missiles (ICBMs), and submarine-launched ballistic missiles (SLBMs)—went from 35-to-1 to 1-to-1.1.[10] Granted, in 1970 the United States still had 2.7 times as many warheads deployed on its strategic delivery vehicles as the Soviets had on theirs. Yet even as late as 1976 American policymakers had paid little attention to the long-term implications of parity in the U.S.–Soviet strategic nuclear competition.[11] Did American nuclear strategy need to be changed or rethought in light of nuclear parity? Given the huge size of the two sides' thermonuclear arsenals by the mid-1970s, further expansion by either the United States or the USSR appeared to be wasteful overkill (at least from a U.S. perspective).[12] But since public perceptions of the strategic balance based on crude numerical comparisons of the opposing forces affected the behavior of allies and adversaries

9 Wohlstetter argued that objectives should always be open to revision. "A government's ends cannot be accepted as the final deliverances of authority or intuition. They are subject to revision as the result of an analysis which frequently displays incompatibilities with other ends of that government or means so costly that the game is not worth the candle. Moreover, even when an opposed-systems design does not set out to revise objectives, it is quite likely to end up that way." Wohlstetter, "Theory and Opposed-System Design," in Zarate and Sokolski, *Nuclear Heuristics*, p. 137.

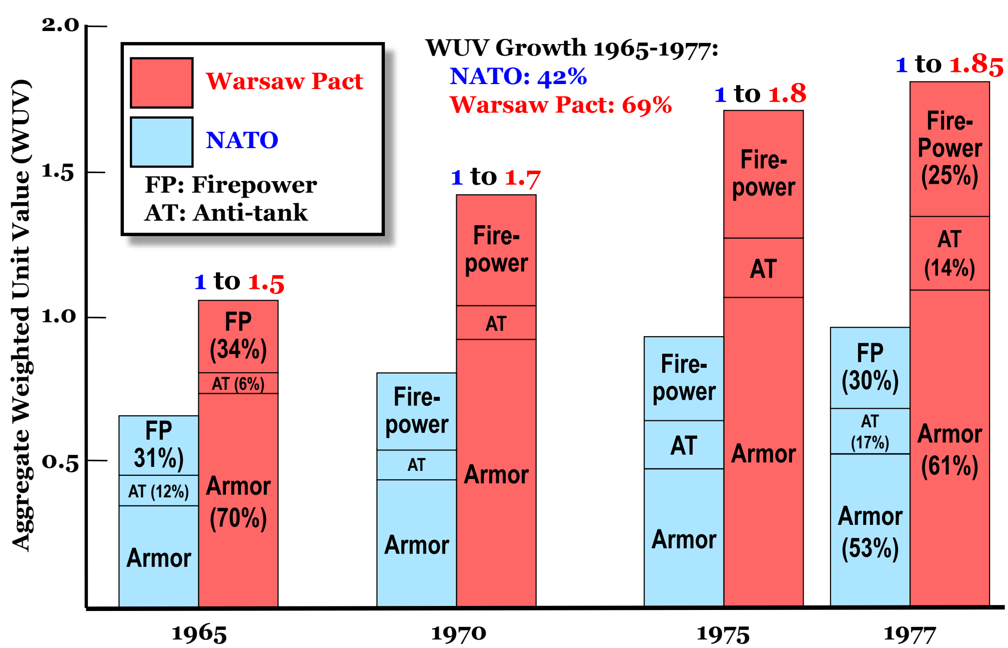
10 The U.S. data are from Natural Resources Defense Council databases that are no longer maintained online. The Russian data are from Pavel Podvig, ed., *Russian Strategic Nuclear Forces* (Cambridge, MA: MIT Press, 2001), pp. 136, 138, 248, 250, 350. While comparisons of the numbers of U.S. and Soviet strategic launchers and warheads are, at best, very crude measures of the Cold War nuclear balance, they were widely cited and used, especially in public debates over the U.S.–Soviet strategic balance.

11 In 1976 Marshall informed Rumsfeld that there was a general lack of appreciation for the loss of overwhelming nuclear superiority, which the United States had enjoyed in the early 1960s, to the Soviet attainment of rough nuclear parity in the early 1970s. Andrew W. Marshall, "The Future of the Strategic Balance," memorandum for the Secretary of Defense, August 26, 1976, pp. 9–10.

12 In 1975 the United States had over 10,000 thermonuclear warheads deployed on bombers, ICBMs, and SLBMs while the Soviets had over 3,000. By 1980 the U.S. and Soviet strategic warhead totals were over 11,000 and 9,000, respectively. A better sense of the resources nuclear weapons consumed on both sides during this period can be gleaned from stockpile data that included tactical nuclear weapons. In 1980 the United States had over 24,000 nuclear warheads in its stockpile, and the Soviets held in excess of 30,000 in its stockpile. The source of the nuclear stockpile data is Hans M. Kristensen and Robert S. Norris, "Global Nuclear Weapons Inventories, 1945–2013," *Bulletin of the Atomic Scientists*, September/October 2013, p. 78, which is available at http://bos.sagepub.com/content/69/5/75.full.pdf+html.

alike, could the United States afford to allow the Soviets to achieve a substantial force ratio advantage in intercontinental nuclear launchers or warheads? What were the most important trends and asymmetries in the competition and what were Soviet objectives? The basic assessment of any military balance, Marshall counseled, would need to address how the reader should think about the competition and its overall state in light of answers to these sorts of first-order questions about objectives and perceptions. Was the U.S. position deteriorating, improving, or staying about the same, and was the U.S. posture adequate for deterrence?

FIGURE 16: GROUND FORCE TRENDS IN CENTRAL EUROPE, 1965–1977[13]



II. Long-term Trends and Key Asymmetries. Here Marshall's guidance was to present comparable trend data to enable the reader to understand the evolution of the balance in the past and how it might evolve in the future. Was there an emerging problem for the United States in the competition? Included would be areas in which the United States had a lead, or in which it was important to stay ahead. Trends would also include areas in which the United States was falling behind. For instance, ONA's 1978 assessment of the European balance showed that from 1965 to 1977 the WUV ratio of Warsaw Pact to NATO ground forces deployed in Central Europe grew from 1.5-to-1 to 1.85-to-1.[14] This change was perceived as a worrying trend and prompted Bill Perry, in 1977, to initiate a project he called the

13 ONA, *The Military Balance in Europe: A Net Assessment*, p. 52.

14 Ibid. Central European NATO and Warsaw Pact ground forces included those deployed in Belgium, the Netherlands, Luxembourg, West and East Germany, Poland, and Czechoslovakia.

"offset strategy" to develop stealth, precision munitions, broad-area surveillance, the Global Positioning System (GPS), and computerized battle networks "to compensate for the Soviet size advantage in conventional forces and thus re-establish general military parity and shore up deterrence."[15] The potential of exploiting these areas of U.S. technological advantage to offset Soviet quantitative superiority in conventional weaponry such as tanks and artillery had been clearly recognized in the early 1975 report of the Long Range Research and Development Planning Program (LRRDPP).[16] Figure 16 illustrates long-term trends that highlighted a growing asymmetry in the NATO–WP conventional balance and led, in turn, to the strategic opportunity that Perry's offset strategy subsequently exploited.

III. Key Uncertainties. Here Marshall raised a number of concerns about the data and intelligence on which ONA's assessments would be based. In 1970 he had attempted an overall comparison of U.S. and Soviet force postures for a special defense panel that Kissinger had instigated to back up negotiations with the Russians on strategic nuclear arms. For a variety of reasons, Marshall had ended up mostly enumerating the problems such an assessment faced. Differences between the two sides' force postures made any simple balancing by specific weapon or other categories inadequate. Intelligence data was unavailable or skimpy in areas such as the logistics, training and general readiness of Soviet forces. The American defense community lacked credible means of assessing the capabilities of U.S. forces to deal with Soviet forces in specific contingencies.[17] And very little effort had been made by the U.S. intelligence community to understand the organization of Soviet forces, their operational practices, or the basic military economics of the Soviet military establishment.[18] The one conclusion Marshall was able to reach in 1970 was that U.S. manpower and weapons costs were greater than those of the Soviets. For example, one F-4 cost about $4 million, whereas the MiG-21 only cost around $1 million.[19] This sort of cost asymmetry raised the possibility that the United States might be pricing itself out of the competition. Whether or not this was true, therefore, was a major uncertainty in the long-term competition, as was American understanding of Soviet objectives. Only after the Cold War had ended did private discussions with former high-ranking former Soviet officials reveal how seriously U.S. observers had misjudged Soviet aims, often exaggerating Soviet aggressiveness. Contrary to the views of many U.S. policymakers and analysts during the Cold War, the Soviets had not been poised for strategic nuclear

---

15 Perry, *My Journey at the Nuclear Brink*, p. 33. See also pp. 36–41.

16 D. A. Paolucci, *Summary Report of the Long Range Research and Development Planning Program* (Falls Church, VA: Lulejian and Associates, February 7, 1975). Among other things the LRRDPP concluded that "non-nuclear weapons with near zero miss" appeared to be technically feasible and militarily effective, and might even provide alternative strategic options to massive nuclear destruction. Ibid., pp. iii, 44–45. The Assault Breaker program was one manifestation of Perry's offset strategy that was clearly anticipated in the LRRDPP.

17 Andrew W. Marshall, *Net Assessment of the U.S. and Soviet Force Posture: Summary, Conclusions and Recommendations* (Washington, DC: National Security Council, September 1970), p. 1.

18 Ibid., p. 2.

19 Ibid., p. 10.

preemption, nor did they develop limited nuclear options or prepare elaborate plans to escalate a theater nuclear war to the global level.[20]

IV. Emerging Strategic Problems or Opportunities: Schlesinger's original guidance—later accepted by both Donald Rumsfeld and Harold brown—was that ONA's assessment would be written for the personal use of the Secretary of Defense (SecDef). Schlesinger recognized that the day-to-day business of the Pentagon tended to be so all-consuming that it was difficult to focus on longer-term strategic issues. This last section of Marshall's 1976 balance format sought to alert the SecDef to emerging problems in any balance area, or to point out opportunities to compete more effectively—and to do so early enough for defense secretaries or other senior officials to make timely decisions. Obviously the choice of what two or three problems or opportunities to highlight in a given assessment required judgment. Such choices boiled down to identifying the most important factors or criteria likely to affect the balance going forward.

Perhaps the most compelling Cold War example of such a choice was the importance Marshall and Schlesinger attached to the burden Soviet military programs imposed on the USSR's economy (that is, the ratio of Soviet military spending to the USSR's GNP). As Schlesinger discovered during his brief tenure in the first half of 1973 as the Director of Central Intelligence (DCI), Central Intelligence Agency (CIA) economic analysts were insistent that the USSR's military burden was, like that of the United States, around 6 or 7 percent of GNP.[21] But this estimate, as Schlesinger and Marshall realized, was inconsistent with two other observations. First, the Soviet economy was widely thought to be around half of U.S. GNP.[22] Second, in absolute terms, the USSR's military establishment was outspending that of the United States.[23] If these two observations were taken to be correct, then the USSR's military burden had to be roughly double CIA's estimate in 1973. Instead of 6 or 7 percent, it should have been at least 12 percent. Thus, one of the first tasks Schlesinger gave Marshall after he had moved to the Pentagon in October 1973 was to push CIA's economists to reexamine their estimate of the USSR's military burden.

20 John G. Hines and Daniel Calingaert, *Soviet Intentions, 1973–1985: A Preliminary Review of U.S. Interpretations*, WD-6305-NA (Santa Monica, CA: the RAND Corporation, December 1992), p. v. This study, like much of ONA's work and sponsored research on Soviet assessments, was prepared for Marshall's office. In 1995 Marshall circulated a two-volume study by Hines, Ellis M. Mishulovich, and John F. Shull on Soviet intentions from 1965 to 1985 based on a series of interviews with former Soviet and U.S. officials, to include Marshall Sergei F. Akhromeev, General Colonel Andrian A. Danilevich, General Makhmut A. Gareev, Harold Brown, and Zbigniew Brzezinski. The interviews are in the second volume. Both are available online at http://nsarchive.gwu.edu/nukevault/ebb285/.

21 Noel E. Firth and James H. Noren, *Soviet Defense Spending: A History of Estimates, 1950–1990* (College Station, TX: Texas A&M Press, 1998), pp. 54–55. In Fiscal Year (FY) 1972, U.S. spending on national defense was 6.5 percent of GDP, and it fell to 5.7 percent in FY 1973 as the United States withdrew from Vietnam. OUSD (Comptroller), *National Defense Budget Estimates for FY 2017*, p. 264.

22 CIA, Directorate of Intelligence, *A Comparison of Soviet and US Gross National Products, 1960–83*, SOV 84-10114 (n.p.: CIA, April 1984), p. iii.

23 Donald H. Rumsfeld, *Annual Defense Department Report FY 1978* (Washington, DC: GPO, January 17, 1977), p. 3.

Disagreement between CIA and ONA over how much of the USSR's economy was being consumed by military programs persisted through the end of the Cold War with Marshall's estimates being consistently higher than CIA's. Much of this longstanding controversy concerned the numerator of the burden estimate. CIA economists were inclined to limit Soviet defense spending to the men, equipment, military R&D, military space program, and nuclear weapons that were visible to U.S. intelligence sources, especially by satellite reconnaissance.[24] But by 1987 Marshall had concluded that there were at least two other tranches that contributed to the USSR's military burden: a second tranche consisted of civil defense, industrial mobilization programs, and dual-use investments such as building into Aeroflot airliners, Baltic ferries, and the USSR's commercial fishing fleet capacities to aid the Soviet military in wartime; a third tranche stemmed from the costs of the USSR's external empire.[25] Marshall's 1987 estimate was that all three tranches had consumed "somewhere between 20 and 30 percent of Soviet GNP" during the previous ten years.[26] By comparison, in 1986 the CIA and the Defense Intelligence Agency (DIA) estimated that the portion of Soviet GNP "devoted to defense at 15–17 percent in current rubles" (and in the range of 12–14 percent if adjusted for inflation).[27] While later, in 2011, the National Intelligence Council conceded that the CIA burden estimates during the Cold War had been imperfect, even at this late date there was little inclination to concede that Marshall's higher estimates had been closer to reality than official estimates.

Two other developments, however, gave further credibility to Marshall's higher burden estimates. As early as 1979, the émigré Soviet economist Igor Birman had alerted Marshall to the likelihood that the CIA was overestimating the size of the USSR's economy by as much as a factor of two. If Birman was right, then the denominator in CIA burden estimates was too large, in addition to the possibility that the numerator was too small. The second development occurred in 2001 when Marshall learned that there had been a fourth tranche, one that had been invisible to U.S. intelligence. At a Paris meeting with Colonel Vitaly Shlykov, who had served in one of the major planning sections of the Soviet General Staff, the Russian revealed that the USSR had produced and maintained "gigantic war reserve stocks" based on exaggerated estimates of likely U.S. war production should the United States mobilize as the county had during World War II.[28] Taking into account both the smaller size of Soviet GNP and Shlykov's fourth tranche, Marshall estimated that

24 Andrew W. Marshall, "Commentary," in Joint Economic Committee, Congress of the United States, *Gorbachev's Economic Plans*, Vol. I, *Study Papers* (Washington, DC: GPO, 1987), p. 482.

25 Ibid., pp. 482, 483.

26 Ibid., p. 484.

27 CIA and DIA, *The Soviet Economy Under a New Leader*, report presented to the Subcommittee on Economic Resources, Competitiveness, and Security Economics of the Joint Economic Committee (Washington, DC: CIA, March 19, 1986), p. 35; Joint Economic Committee, Congress of the United States, "Proxmire Releases CIA-DIA Report on the Soviet Economy," press release, March 30, 1986, p. 1; and Firth and Noren, *Soviet Defense Spending: A History of Estimates, 1950–1990*, pp. 128–131.

28 Andrew W. Marshall, letter to Thomas C. Reed, September 27, 2001, p. 3.

during the last decades of the Cold War the USSR's military burden had probably been somewhere "in the range of 35 to 50 percent of Soviet GNP."[29]

How significant was the USSR's military burden in the Soviet Union's 1991 collapse? The USSR's heavy military burden was not, of course, the only factor that led Mikhail Gorbachev to begin implementing policies such as *perestroika* (restructuring) that got out of his control and ultimately led to the collapse of the Soviet Union. Perhaps the Soviet military's economic burden was not the main factor in the USSR's abrupt collapse, even though the country had not been directly challenged with an existential threat. In 1994 Walter Laqueur emphasized the "spiritual crisis" that began to fatally undermine the self-confidence of the USSR's intelligentsia in the country's future.[30] But the USSR's military burden was certainly a contributor to the USSR's collapse.

Regardless of how one judges Marshall's consistently higher estimates of the USSR's military burden relative to the CIA's, there can be little doubt that he and Schlesinger were right to emphasize the burden Soviet military program imposed on the USSR's economy as a driver in the Cold War competition between the United States and the Soviet Union. Witness Gorbachev's insistence in 1987 that the United States' efforts to bankrupt the Soviet Union by developing advanced weaponry would fail.[31] Marshall and Schlesinger's emphasis on the USSR's military burden, therefore, illustrates the vital importance of diagnostic net assessment in choosing appropriate analytic measures for judging the current state and likely future of military competitions.

The Universal, Enduring Problem of Selecting Analytic Criteria

At the beginning of this report, several hypotheses were advanced about the problem of choosing appropriate analytic measures. First, such choices are a widespread, if not universal, challenge regardless of whether one is making strategic choices about the future or trying to untangle the various causes of past historical outcomes. Second, because no all-purpose, by-the-numbers method for selecting appropriate analytic criteria has yet to be discovered, there is little hope of avoiding the dependence of such choices on subjective judgment. Schlesinger was right when he insisted in 1967 that the subjective element in the selection of such measures will remain substantial. Indeed, it remains substantial in the second decade of the twenty-first century. Third, because the choices of analytic criteria depend on the many, if not countless, contextual details of the situation at issue, each and every choice is unique. This observation goes far to explain why choosing appropriate analytic criteria has rarely been seen as the universal problem it is. After all, determining the principal, necessary, or sufficient

29 Ibid., p. 2.

30 Walter Laqueur, *The Dream that Failed. Reflections on the Soviet Union* (New York and Oxford: Oxford University Press, 1994), pp. 71–72.

31 Mikhail Gorbachev, *Perestroika: New Thinking for Our Country and the World* (New York: Harper and Row, 1987), p. 220.

causes of the West's rise appears, on the surface, to have little in common with trying to decide what military forces and capabilities to buy and maintain in order to deter nuclear use or major power warfare on the scale of World War II in the decades ahead.

The approach taken in this report to build a case for these hypotheses was to examine the analytic criteria selected in a wide and diverse range of cases. The cases examined largely fell into two categories. First, there were those like the Battle of the Atlantic and the rise of the West after 1500 in which various criteria were advanced to explain or understand the actual historical outcomes. The second set of cases focused on the future rather than the past. Examples included RAND's efforts to identify the most cost-effective strategic bomber force for the Air Force, the use of GDP for cross-border comparisons of national economies in pursuit of formulating effective macroeconomic policies, and diagnostic net assessment to aid the SecDef's strategic management of the U.S. Department of Defense. Of course, the boundary between these two categories is somewhat blurred by metrics such as casualty rates, force ratios, and situation awareness that can be used either to explain past results or make informed decisions and projections (though not necessarily accurate predictions) about the future in a non-ergodic world.

The deepest insights into the criterion problem came in reflecting on the limits to what we can know about the drivers of the great divergence between the West and the Rest. Even the briefest comparison of the different causal explanations in Figures 9, 10, and 11 confirms that no widespread consensus has emerged among historians, economists, or other scholars about the *principal* determinants of the West's rise to wealth and military-political preeminence since Adam Smith's *Inquiry into the Nature and Causes of the Wealth of Nations*. Two additional observations suggest that a complete, exhaustive causal account of the West's rise is likely to remain permanently beyond the reach of human knowledge.

First, there is the seldom discussed but elementary distinction between the fixed past and our ever-changing historical explanations of what happened and why. An exhaustive reconstruction of the past that left out not even the tiniest causal details would require vastly more information that we can ever aspire to possess. The fact is that much of the requisite information, if not most of it, is lost to us and not recoverable. I speculated at the end of Chapter 3 that we may possess as little as 5 percent of all the causal information that would be required for our understanding of an outcome as protracted and complex as the rise of the West to approach completeness and certainty. What is clear, however, is that the further we try to look back in time, the less of the required information we will have. John Lukacs was right when he observed that, "All history is revisionist, in one way or another."[32]

32 Lukacs, *The Future of History*, p. 143.

Second, there is the reality of emergent phenomena whose precise causal determinants can be hidden from us by the intervention of chaotic processes. Even the slightest difference between what we think we know about the past and what the past's actual necessary and sufficient causes were can lead us dramatically astray. In Lorenz's seminal paper that marked the rediscovery of nonlinear dynamics he asked whether "two particular weather situations differing by as little of as the intermediate influence of a single butterfly will generally after sufficient time evolve into two situations differing by as much as a tornado."[33] Lorenz's eventual answer was, we now know, that such divergent outcomes are possible in the case of chaotic phenomena such as weather. Lorenz was concerned with predicting the future. But as pointed out at the end of Chapter 3, this limitation applies to our understanding of the past as well. To repeat Sautoy's conclusion, "The past even more than the future is probably something we can never truly know."[34]

This is not to deny the existence of at least some regularities that permit accurate causal accounts or predictions to be made. But in areas ranging from picking stocks to predicting how a military recruit will later perform in actual combat, or truly understanding past and future performance of the U.S. economy, the presence of randomness (or luck) argues that errors of causal explanation or prediction are inevitable because the world as a whole is non-ergodic. And, as Daniel Kahneman has rightly noted, the line that separates the possibly predictable future from the unpredictable distant future has yet to be drawn.[35]

These observations argue that the criterion problem is not going away. Whether one is trying to understand the past or make strategic choices about the future, any judgments or recommendations reached will depend on3 the choice of analytic criteria. Such choices give every indication of being universal problems of analysis. In the absence of a formula or all-purpose methodology for choosing appropriate analytic criteria, such choices will continue to have a substantial subjective element. As Voltaire recognized, certainty—whether about explaining the past or deciding about future courses of action—is absurd. Or, as Andrew Marshall emphasized in response to an early draft of this report, choosing analytic measures is going to remain "a *big* problem."[36] Hopefully this report will be a first step toward recognizing the selection of good analytic measures as the universal, enduring problem it is.

33 Edward N. Lorenz, "Predictability; Does the Flap of a Butterfly's Wings in Brazil Set off a Tornado in Texas?" American Association for the Advancement of Science, 139th Meeting, Washington, DC, December 29, 1972, p. 2.

34 Du Sautoy, *What We Cannot Know: Explorations at the Edge of Knowledge*, p. 55.

35 Kahneman, *Thinking, Fast and Slow*, p. 221.

36 Barry D. Watts, telephone conversation with Andrew Marshall, May 24, 2016.

LIST OF ACRONYMS

AAF	Army Air Forces
ACEVAL	Air Combat Evaluation
AESA	active electronically scanned array
AIM	air intercept missile
AMRAAM	Advanced Medium-Range Air-to-Air Missile
APC	armored personnel carrier
ASV	air-to-surface vessel
ASW	anti-submarine warfare
AU	Air University
AWACS	airborne warning and control system
BAMS	Broadcast to Allied Merchant Ships
BC	Royal Air Force Bomber Command
B-Dienst	Beobachtungs-Dienst
B.d.U.	Befehlshaber der Unterseeboote
BVR	beyond visual range
CAA	Concepts Analysis Agency
CBO	Combined Bomber Offensive
CEO	Chief Executive Officer
CIA	Central Intelligence Agency
DCI	Director of Central Intelligence
DF	direction finding
DIA	Defense Intelligence Agency
DoD	Department of Defense
ETO	European Theater of Operations
F	Fahrenheit
FISIM	financial intermediation services indirectly measured
FY	fiscal year
GC&CS	Government Code and Cipher School
GCI	ground controlled intercept
GDP	gross domestic product
GNP	gross national product

LIST OF ACRONYMS

GPS	Global Positioning System
GRT	gross registered ton
HERO	Historical Evaluation and Research Organization
IFV	infantry fighting vehicle
ICBM	intercontinental ballistic missile
IMF	International Monetary Fund
LED	light-emitting diode
LRRDPP	Long Range Research and Development Planning Program
NAG	Net Assessment Group
nm	nautical mile
NSC	National Security Council
OECD	Organization for Economic Co-operation and Development
OEEC	Organization for European Economic Co-operation
OIF	Operation Iraqi Freedom
ONA	Office of Net Assessment
OSRD	Office of Scientific Research and Development
OUE	Operational Utility Evaluation
MOE	measure of effectiveness
NATO	North Atlantic Treaty Organization
PPP	purchasing power parity
QJM	Quantified Judgement Model
RAND	Research ANd Development
RAF	Royal Air Force
R&D	research and development
SA	situation awareness
SAC	Strategic Air Command
SecDef	Secretary of Defense
SLBM	submarine-launched ballistic missile
SNA	System of National Accounts
SPSS	Statistical Package for Social Science
SS	Schutzstaffel

LIST OF ACRONYMS

TASCFORM	Technique for Assessing Force Modernization
THz	terahertz
TNT	trinitrotoluene
TO&E	tables of organization and equipment
UK	United Kingdom
UN	United Nations
USAAF	United States Army Air Forces
USAF	United States Air Force
USSBS	United States Strategic Bombing Survey
USSTAF	United States Strategic Air Forces in Europe
WEI/WUV	Weapon Effectiveness Indices/Weighted Unit Values
USSR	Union of Soviet Socialist Republics
WVR	within visual range
WP	Warsaw Pact

CSBA
Center for Strategic and Budgetary Assessments

1667 K Street, NW, Suite 900
Washington, DC 20006
Tel. 202-331-7990 • Fax 202-331-8019
www.csbaonline.org