

H-Diplo | ISSF Roundtable XII-11

issforum.org

Jason Lyall. *Divided Armies: Inequality and Battlefield Performance in Modern War.* Princeton: Princeton University Press, 2020. ISBN: 9780691192437 (hardcover, \$99.95); 9780691192444 (paperback, \$35.00).

14 May 2021 | <https://issforum.org/to/ir12-11>

Editor: Diane Labrosse | Commissioning Editor and Chair: Dan Reiter | Production Editor: George Fujii

Contents

| | |
|--|----|
| Introduction by Dan Reiter, Emory University..... | 2 |
| Review by Alexander B. Downes, The George Washington University..... | 7 |
| Review by Kristen A. Harkness, University of St. Andrews..... | 15 |
| Review by Michael C. Horowitz, University of Pennsylvania | 19 |
| Review by Yuri M. Zhukov, University of Michigan..... | 22 |
| Response by Jason Lyall, Dartmouth College | 26 |

INTRODUCTION BY DAN REITER, EMORY UNIVERSITY

The oldest question in the study of international relations (IR) is: what helps armies win their battles? This is the IR question the ancients struggled over more than any other. The Old Testament, for example, is replete with discussions of armies fighting and trying to win battles, such as the Israelites hoping the Ark of the Covenant would bring them victory, and successfully using an ambush feint at the Canaanite city of Ai. Over the millennia, scholars, and observers from Thucydides to Sun Tzu to Machiavelli to contemporary political scientists have been keen to move beyond exploring material answers to this question, that bigger armies with better weapons win, to developing non-material answers, that cultural, ideational, political, and social factors might help explain war and battle outcomes.¹ Are the armies of some kinds of societies more likely to win? Self-servingly, some have asked, might some noble virtue hard wired into our national identity also help us win at arms? Conversely, might some fatal flaw within our cultural or political genetic code fate us to be destroyed?

Into the twenty-first century, scholars have continued to find the determinants of battle and war outcomes to be an exciting area for research, developing and exploring new insights.² Jason Lyall explores this millennia-old question in his new and powerfully argued book, *Divided Armies: Inequality and Battlefield Performance in Modern War*. His book, a *Foreign Affairs* Best Book of 2020,³ presents the theoretical claim that societies that repress their own minorities field less effective militaries. Mistreated citizens are less willing to fight for oppressive and discriminatory regimes and more likely to defect or desert during war, and even fight their military comrades directly. Relatedly, this perceived lack of loyalty forces the military to make choices that undermine its effectiveness, including dividing military units from each other to prevent collusion and rebellion, and selecting inferior military tactics. To test his claims, Lyall presents quantitative analysis of a new data set of outcomes of conventional wars since 1800, the Project Mars data, as well as case studies of wars in Morocco, the Ottoman and Hapsburg Empires, Ethiopia, and the Eastern Front in World War II.

This book provides several exciting contributions; regrettably, there is space here to describe only a few. First, the book invigorates the study of the non-material determinants of war outcomes, in particular by moving beyond a focus on political institutions and towards the understudied areas of social cleavages and discrimination. The field has enjoyed almost thirty years of exciting, contentious, and extensive research on whether democracies are more likely to win their wars.⁴

¹ On non-material explanations of military effectiveness, see Filippo Andreatta, “Conclusion: The Complexity of Military Effectiveness,” in Dan Reiter, ed., *The Sword’s Other Edge: Trade-offs in the Pursuit of Military Effectiveness* (Cambridge: Cambridge University Press, 2017), 254-267.

² Recent political science books on military effectiveness and conventional war outcomes include Jessica L. P. Weeks, *Dictators at War and Peace* (Ithaca: Cornell University Press, 2015); Caitlin Talmadge, *The Dictator’s Army: Battlefield Effectiveness in Authoritarian Regimes* (Ithaca: Cornell University Press, 2015); Jasen J. Castillo, *Endurance and War: The National Sources of Military Cohesion* (Stanford: Stanford University Press, 2014); Reiter, ed., *The Sword’s Other Edge*; Stephen Biddle, *Nonstate Warfare: The Military Methods of Guerrillas, Warlords, and Militias* (Princeton: Princeton University Press, 2021); Jonathan D. Caverley, *Democratic Militarism: Voting, Wealth, and War* (Cambridge: Cambridge University Press, 2014); Ryan Grauer, *Commanding Military Power: Organizing for Victory and Defeat on the Battlefield* (Cambridge: Cambridge University Press, 2016); Rosella Cappella Zielinski, *How States Pay for Wars* (Ithaca: Cornell University Press, 2016); Kenneth N. Pollack, *Armies of Sand: The Past, Present, and Future of Arab Military Effectiveness* (Oxford: Oxford University Press, 2019); Patricia A. Weitsman, *Waging War: Alliances, Coalitions, and Institutions of Interstate Violence* (Stanford: Stanford University Press, 2014). There have also been several recent works on counterinsurgency effectiveness and outcomes, including Eli Berman, Joseph H. Felter, and Jacob N. Shapiro, *Small Wars, Big Data: The Information Revolution in Modern Conflict* (Princeton: Princeton University Press, 2018).

³ <https://www.foreignaffairs.com/lists/2020-12-04/best-books-2020>.

⁴ Prominent works on democracy and war outcomes include David A. Lake, “Powerful Pacifists: Democratic States and War,” *American Political Science Review* 86 (March 1992): 24-37; Allan C. Stam III, *Win, Lose, or Draw: Domestic Politics and the Crucible of War* (Ann Arbor: University of Michigan Press, 1996); Dan Reiter and Allan C. Stam, *Democracies at War* (Princeton: Princeton

Surprisingly, there has been much less work on the relationships between social cleavages, military effectiveness, and war outcomes, and certainly nothing nearly as thorough as *Divided Armies*.

From a policy point of view, it is critical in the twenty-first century to know how the presence or absence of cleavages and minority repression might affect the fighting power of states such as Iraq, Russia, Ukraine, China, North Korea, Syria, Saudi Arabia, Turkey, Iran, and others. It also provides an additional angle on social justice debates in the United States as to whether racial inequality and tension in American society might or might not undermine American fighting power. Diversity need not hobble American fighting power; some of the most effective fighting units in the U.S. military in World War II, such as the African-American Tuskegee Airmen and the Japanese-American 442nd Regimental Combat Team, contained repressed minorities. But America has not always been so fortunate. Black-white tensions during the Vietnam War were sometimes so severe they culminated in refusals to fight and even the murder of officers,

Second, the conceptual structure of the book pushes the general study of war in an important direction by rearranging the categories of warfare. The long-standing approach in scholarly IR has been to separate interstate wars (conflicts between nation-states) from civil wars (conflicts usually involving a nation-state government and one or more substate actors), and to assume that belligerents fight conventionally in interstate wars (regular armies, fixed lines of battle, seeking territorial gain, and so on), and unconventionally in civil wars (weaker insurgents generally avoiding stronger government forces, using guerrilla tactics, etc.). This was to some degree an outgrowth of the Correlates of War (COW) data project's decades-old distinction between and categorization of interstate and civil wars.⁵ It also was accepted during the Cold War, accompanying the greater scholarly interest in interstate war and great power war in particular and neglect of the study of civil war.

However, scholars are beginning to understand that it is factually incorrect as well as in some cases intellectually undesirable to equate conventional strategy with interstate war on one hand and unconventional strategy with civil war on the other. Scholars in recent years have begun to relax this assumption, noting in particular that some civil wars are fought with conventional strategy, and some interstate wars are fought with unconventional strategy.⁶ In *Divided Armies*, Lyall builds a theory predicting to battle performance in wars characterized by conventional strategy, inclusive of interstate wars and civil wars. And, he builds a data set of all conventional wars, some of which are interstate and some of which are civil.

Other scholars are also fruitfully ignoring a clean conceptual divide between interstate and civil wars. Caitlin Talmadge's landmark work on coup-proofing and battle performance predicts to interstate and civil war performance, and her empirics include both kinds of conflicts.⁷ The most ambitious work to address this issue head-on is Stephen Biddle's 2021 book, which takes as its leading task the goal of casting aside the interstate vs. civil war distinction, instead developing a conceptual structure of warfare describing a spectrum of conflict from conventional to guerrilla.⁸

University Press, 2002); Michael C. Desch, *Power and Military Effectiveness: The Fallacy of Democratic Triumphalism* (Baltimore, MD: Johns Hopkins University Press, 2008); Michael E. Brown, Owen Coté Jr., Sean M. Lynn-Jones, and Stephen E. Miller, *Do Democracies Win Their Wars?* (Cambridge: MIT Press, 2011); Caverley, *Democratic Militarism*.

⁵ <https://correlatesofwar.org/>.

⁶ Stam, *Win, Lose, or Draw*; Stathis N. Kalyvas and Laia Balcells, "International System and Technologies of Rebellion: How the End of the Cold War Shaped Internal Conflict," *American Political Science Review* 104(August 2010): 415-429.

⁷ Talmadge, *Dictator's Army*.

⁸ Biddle, *Nonstate Warfare*.

For informing policy, understanding many military conflicts in the twenty-first century is going to require relaxing the assumption that rebels always fight unconventionally and state armies always fight conventionally. Israel found this out the hard way in its 2006 Lebanon War when it faced Hezbollah forces that employed strategy that looked more like conventional warfare than guerrilla actions. Warfare against the Islamic State in Iraq in the 2010s frequently looked quite conventional. Russia has been inserting its forces into Ukraine since 2014, and those forces have engaged in unconventional and conventional military tactics.

Third, the depth, detail, and care of the empirical work in *Divided Armies* is highly impressive, and sets a high mark for other scholars of war, on both the quantitative and qualitative side. On the quantitative side, scholars these days usually focus on statistical methods, debating the most appropriate technique to analyze the data at hand, and to be sure *Divided Armies* chooses statistical techniques and research design with great care. But this focus on statistical technique sometimes looks past a priority that is at least as important: data quality. Lyall's theory demanded the creation of a list of conventional wars, and then the coding of the necessary independent and dependent variables for each war. In pursuing this task, Lyall bravely left the familiar by completely abandoning not only the COW list of wars, but also the COW list of members of the nation-state system, starting completely from scratch to build a new list of conventional wars. He sought a larger set of wars than COW, including conflicts that killed at least 500 in battle, whereas COW was limited to wars that killed at least 1000. He hired dozens of undergraduate and graduate students to code his data, using primary and secondary sources in a variety of languages. Then he hired another group of students whose sole job was to 'red team' the original codings, to look for errors and serve as devil's advocates. This level of care represents a steady raising of the bar of standards of quantitative data quality in IR over the last few decades, and other scholars should consider themselves challenged to match or exceed it in their own work.

Another admirable aspect of the quantitative data collection effort is its transparency. This helps scholars understand exactly what was coded and how it was coded, facilitating replication.⁹ It also recognizes that the collection of quantitative observational data is never 'finished,' and that part of the scholarly dialogue over time is ongoing efforts to improve data quality. At the same time that Lyall was collecting his data, two coauthors and I were collecting our own new data set on wars.¹⁰ Lyall helpfully provided us with his data, which we used to improve the coverage of our data set. Our data set includes some adjustments to Lyall's data, and these differences may shed light on how the Project Mars data might be revised in the future.

The data on the qualitative side are also impressive. One trend that IR scholars do not often recognize is the gigantic improvement in the rigor of qualitative IR research over the past few decades. Scholarship recognized as being landmark, qualitative works in security studies in the 1980s and 1990s sometimes lacked primary sources or even rigorous historiographical surveys of secondary sources. Over time, the caliber of qualitative empirical research has steadily improved. Works like Jack Snyder's 1984 book began to make wider use of non-English language primary and secondary sources.¹¹ Jeremy Weinstein's 2007 work was a landmark, in that its study of insurgency was one of the first to build on field work in

⁹ On the importance of replication in quantitative and qualitative work, see Dan Reiter, "Measurement Replication in Qualitative and Quantitative Studies," in Colin Elman, John Gerring, and James Mahoney, eds., *The Production of Knowledge: Enhancing Progress in Social Science* (Cambridge: Cambridge University Press, 2020), 284-300.

¹⁰ The Interstate War Data. See Dan Reiter, Allan C. Stam, and Michael C. Horowitz, "A Deeper Look at Interstate War Data: Interstate War Data version 1.1," *Research and Politics* (October-November 2016): 1-3, and Dan Reiter, Allan C. Stam, and Michael C. Horowitz, "A Revised Look At Interstate Wars, 1816-2007," *Journal of Conflict Resolution* 60 (August 2016): 956-976.

¹¹ Jack Snyder, *Ideology of the Offensive: Military Decision Making and the Disasters of 1914* (Ithaca: Cornell University Press, 1984).

active war zones, and interviews with rebel leaders.¹² And a growing number of studies, marked notably by Stathis Kalyvas's careful study of internal conflict in 1940s Greece in his widely read book on violence during civil war, began to present careful, micro-level quantitative evidence on individual conflicts.¹³

Realistically, whatever language skills or research budgets one has, it is not feasible for all case studies to undertake a deep dive into source material since the universe of information is at times unavoidably thin. IR scholars of ancient conflicts in regions such as China and Greece have faced this problem.¹⁴ But scholars should be challenged to use important materials, including primary materials, when they exist. For example, in his case study of the 1941 Battle of Moscow, Lyall makes extensive use of Russian-language primary materials to provide an extremely detailed account of Soviet forces, both with regards to their internal divisions and their levels of combat performance.

Last, the book's question and theory provide an opportunity to bring the phenomena of racial and social inequality more squarely into IR research. Most discussions of racial and social inequality in IR have tended to focus on repression and diversity as potential causes of internal conflict. Beyond the small body of research on inequality and military effectiveness, there has been very little consideration by political scientists of how racial and social injustice or even race more broadly might affect more traditional IR outcomes, or how IR outcomes affect race or racial and social injustice.¹⁵ *Divided Armies* may encourage additional scholars to follow in its path and delve into the many possible angles connecting race and IR.

Divided Armies is a massive and fertile contribution to scholarly debates on war, battlefield performance, and the implications of racial and social justice in IR. It will no doubt encourage waves of research going forward, work that unpacks, extends, and critiques the main theoretical and empirical arguments of the book. This H-Diplo/ISSF roundtable on *Divided Armies* contains outstanding contributions that provide a deeper understanding of the book and a number of ideas for advancing the research agenda while demonstrating their deep appreciation for the book's many contributions. Here I offer only one or two points from the individual contributions, which will hopefully whet the appetite of the reader to examine the entirety of each essay. Alexander Downes underscores the importance of understanding why militaries exhibit inequality, developing the idea that leaders' threat environment may affect their decisions to import exclusionary policies into their militaries; Kristen Harkness notes the need for very careful attention to the measurement of identity; Michael Horowitz pushes us to think harder about combat power and military strategy in order to build on Lyall's arguments a more sophisticated understanding of how these two phenomena interact with inequality; and Yuri Zhukov makes a point made by some other contributors of the need for building a dyadic theory and empirical test, recognizing that battle and war outcomes are driven by diversity among the militaries and societies of all belligerents.

¹² Jeremy Weinstein, *Inside Rebellion: The Politics of Insurgent Violence* (Cambridge: Cambridge University Press), 2007.

¹³ Stathis N. Kalyvas, *The Logic of Violence in Civil War* (Cambridge: Cambridge University Press, 2006).

¹⁴ Victoria Tin-Bor Hui, *War and State Formation in Ancient China and Early Modern Europe* (Cambridge: Cambridge University Press, 2005), Dan Reiter, "Gulliver Unleashed? Restraint, Hierarchy, and the Case of Ancient Athens," unpublished manuscript, Emory University, 2020.

¹⁵ Among the exceptions are Christopher S. Parker, *Fighting for Democracy: Black Veterans and the Struggle against White Supremacy in the Postwar South* (Princeton: Princeton University Press, 2009); Menna-Weiwot A. Demaïssie and Errol Henderson, "Race and Ethnicity in Black Congressional Representation: The Case of U.S. Foreign Policy Towards Africa," *Congress and the Presidency* (forthcoming); Scott Sigmund Gartner and Gary Segura, "Race, Casualties, and Opinion in the Vietnam War," *Journal of Politics* 62(February): 115-146.

Participants:

Jason Lyall is the inaugural James Wright Chair of Transnational Studies and Associate Professor of Government at Dartmouth College, where he also directs the Political Violence FieldLab.

Dan Reiter (Ph.D., University of Michigan, 1994) is Samuel Candler Dobbs Professor of Political Science at Emory University. He is the author, editor, or coauthor of several books, including *Crucible of Beliefs: Learning, Alliances, and World Wars* (Cornell University Press, 1996), *Democracies at War* (with Allan C. Stam, Princeton University Press, 2002), the award-winning *How Wars End* (Princeton University Press, 2009), and *The Sword's Other Edge: Tradeoffs in the Pursuit of Military Effectiveness* (Cambridge University Press, 2017), as well as dozens of articles on war outcomes, military effectiveness, the causes of war, alliances, military strategy, terrorism, nuclear proliferation, and other topics.

Alexander B. Downes (Ph.D., University of Chicago, 2004) is Associate Professor of Political Science and International Affairs at the George Washington University, where he is also co-director of the Institute for Security and Conflict Studies. His first book, *Targeting Civilians in War* (Cornell University Press), won the Joseph Leggold Prize awarded by Georgetown University for best book in international relations published in 2008. His second book, *Catastrophic Success: Assessing the Consequences of Foreign-Imposed Regime Change*, is forthcoming from Cornell in 2021.

Kristen A. Harkness is Senior Lecturer in the School of International Relations at the University of St. Andrews. Her research focuses on understanding how ethnicity shapes the loyalty and behavior of military institutions in Africa and has been funded by the British Academy. She is the author of *When Soldiers Rebel: Ethnic Armies and Political Instability in Africa* (Cambridge University Press, 2018) and her work has been published in journals such as *Democratization*, the *European Journal of International Security*, the *Journal of Conflict Resolution*, the *Journal of Peace Research*, and the *Journal of Strategic Studies*.

Michael C. Horowitz is Richard Perry Professor and Director of Perry World House at the University of Pennsylvania.

Yuri M. Zhukov is an Associate Professor of Political Science at the University of Michigan, Ann Arbor, and a Faculty Associate with the Center for Political Studies at the Institute for Social Research.

REVIEW BY ALEXANDER B. DOWNES, THE GEORGE WASHINGTON UNIVERSITY

The product of ten years of hard labor, Jason Lyall's excellent new book, *Divided Armies*, conveys a simple, compelling idea: inequality undermines battlefield performance. States that discriminate against or violently persecute segments of their populations field armies that at a minimum underperform—and sometimes collapse disastrously—on the battlefield. The problem for such states is that they cannot rely solely on the privileged (“core”) societal group to staff their armies; they must depend to a greater or lesser extent on those groups they oppress (“non-core” groups) for military power. This creates a fundamental dilemma because the willingness of these groups to fight on behalf of states that sometimes violently abuse them is highly questionable: soldiers from such groups are liable to flee for home at the first opportunity, defect to the enemy, or surrender in droves. These armies are thus likely to lack both skill and cohesion. Their casualty rates will be high and their tactical and operational sophistication will be low. They will suffer debilitating rates of desertion and defection, and will often deploy violence against their own troops to force them to fight and prevent them from running away.

Lyall assembles a mountain of evidence to support this simple claim. At the base of this mountain lies Project Mars, a new dataset of conventional wars that more than doubles the number of such conflicts that were available in the previous best available source, the Correlates of War (COW) Interstate War Data. Project Mars includes all conventional wars fought from 1800 to 2011—whether they were fought within or between political entities.¹ Quantitative analysis of these data yields strong support for the argument that as military inequality increases, combat performance worsens. Lyall also employs a sophisticated case selection strategy, analyzing four pairs of cases selected from Project Mars using a matching algorithm in addition to four Soviet rifle divisions in the Battle of Moscow in World War II. The result is a tour de force of social science: persuasive logic and compelling evidence that is simultaneously methodologically refined, approachable, and beautifully written. It is an incredible achievement of which its author should be justifiably proud.

This review proceeds in two sections. In the first section, I place Lyall's argument in the prevailing literature on military effectiveness. I argue that the book contains the most comprehensive conception of battlefield performance to date, combining measures of both skill and will in gauging an army's ability to fight. It offers a new independent variable—military inequality—but continues the recent trend in this literature of emphasizing the domestic political determinants of battlefield effectiveness.

In the second section, I critically engage with the argument and evidence presented in the book, raising four questions that may be worthy of further investigation. First, I ask a question that Lyall largely sidesteps at the beginning of his argument: why do leaders implement visions of society that exclude large swathes of the population? I suggest that the answer may lie in the threat environment they face: leaders who face challenges from other ethnic groups when they take power may be more likely to rely on strategies of violent exclusion, thereby producing higher levels of military inequality. Second, although Lyall compellingly demonstrates significant differences *between* militaries at varying levels of inequality, his theory struggles to explain variation *within* those categories. In particular, it has trouble explaining the poor performance of a substantial number of armies with low military inequality. I argue that while the absence of military inequality may be a necessary condition for good battlefield performance, it is not a sufficient one, and further contend that other theories of military effectiveness are needed to explain variation among countries at similar levels of inequality. Third, I ask why some states—such as the Confederacy in the American Civil War—seemingly defy Lyall's logic and refuse to draw on non-core groups for military power. Fourth, I explore why non-core groups sometimes believe promises from exclusionary leaders that they will

¹ In other words, Project Mars includes conflicts that fall into all three of COW's categories of wars: intrastate, interstate, and extrasystemic (i.e., imperial or colonial). Moreover, Lyall's collection does not rely on COW's outmoded rules for membership in the international system to identify political entities. For his definitions and coding rules, see 22-26. For an earlier criticism of COW's rules, see Tanisha M. Fazal, *State Death: The Politics and Geography of Conquest, Occupation, and Annexation* (Princeton: Princeton University Press, 2007), 243-258. Project Mars may be accessed through the Harvard Dataverse, DOI: <https://doi.org/10.7910/DVN/DUO7IE>. The COW data are available at <https://correlatesofwar.org/data-sets/COW-war>.

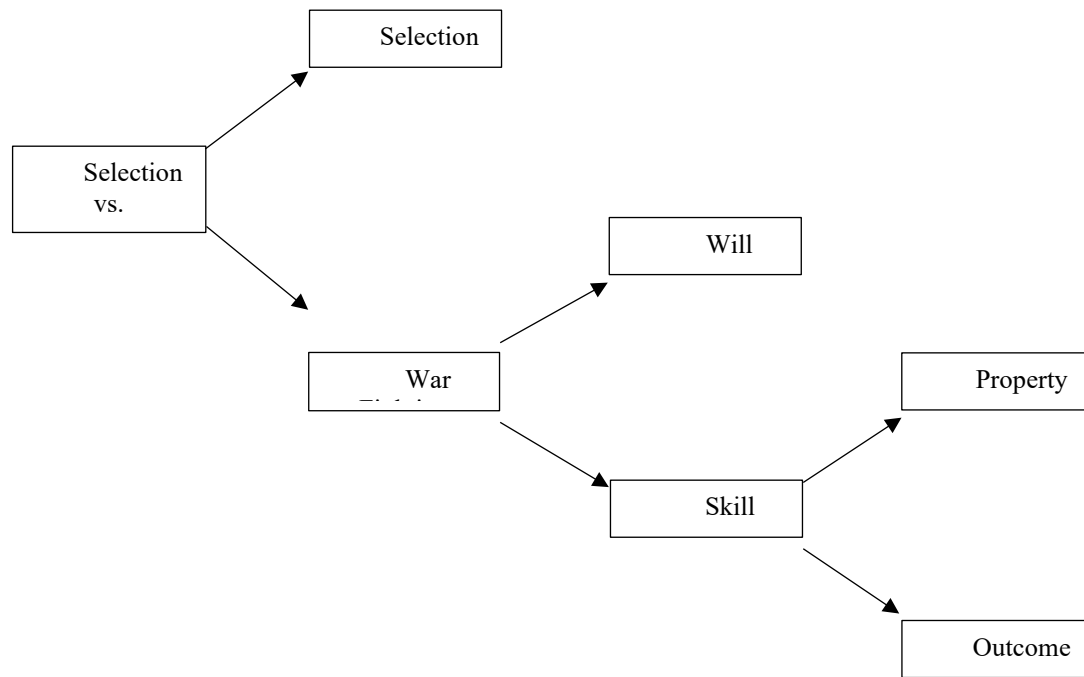
revise the social contract and treat non-core groups better if only these groups will fight hard to save the regime now—promises that obviously lack credibility.

Placing Divided Armies in the Literature

Although *Divided Armies* goes well beyond the existing military effectiveness literature in certain respects, in other ways it fits quite comfortably with recent studies. Regarding the former, the book offers the most comprehensive definition and measurement of battlefield effectiveness yet available. Regarding the latter, it fits comfortably with recent work that emphasizes the domestic sources of military power while offering a new independent variable at that level of analysis.

Existing literature is divided depending on how it answers three questions (see Figure 1). The first question is whether the causal mechanism that explains effectiveness resides at the level of selection *into* wars or how well states fight once *in* wars. Selection arguments ‘explain’ military effectiveness by focusing on how wisely civilian policymakers (with military input) choose which wars to start or enter. The assumption is that states will obtain better outcomes in war if they place their militaries in advantageous positions to fight, such as by targeting weak or diplomatically isolated adversaries. The principal mechanism that explains the quality of war selection is how accountable leaders are to mass or elite audiences that can remove them for poor performance.² Studies have argued that leaders are more vulnerable to removal in democracies and certain types of autocracies—such as single party regimes and military juntas—and thus these regime types should be more likely to prevail in war than non-accountable regimes, such as personalist dictatorships.³

Figure 1. A Typology of Theories of Military Effectiveness



² Additional factors argued to influence the quality of civilians’ choices are the marketplace of ideas, sound military advice from a meritocratic officer corps, and the military or non-military background of leaders and audiences. See Dan Reiter and Allan C. Stam, *Democracies at War* (Princeton: Princeton University Press, 2002); and Jessica L.P. Weeks, *Dictators at War and Peace* (Ithaca: Cornell University Press, 2014).

³ Reiter and Stam, *Democracies at War*; Weeks, *Dictators at War and Peace*.

Studies that privilege how well or poorly militaries fight once engaged in battle in turn are differentiated by how they answer a second question: whether they focus on the will or skill of soldiers. The will to fight and the cohesion of troops in combat has long been a major concern of students of military effectiveness; the primary axis of debate has been between those who attribute combat motivation to primary social group ties versus ideology (e.g., nationalism, liberalism, fascism, etc.).⁴ However, since the publication of Stephen Biddle's influential book *Military Power*, which popularized the view that the key to effectiveness on the modern battlefield lies in proper force employment—which in turn requires great skill to implement—studies have placed greater emphasis on explaining why some armies are sufficiently skilled to implement the modern system of force employment whereas others are not.⁵ One influential argument again stresses regime type, namely the degree to which leaders fear a military coup. Such leaders take steps to protect themselves that involve hindering the military's ability to fight, leading it to perform poorly in wartime.⁶

Finally, scholars disagree whether skill and effectiveness are attributes of an armed force or qualities that can be measured only relative to other militaries in combat. An example of the former is Caitlin Talmadge's argument that armies are skillful and effective to the extent that they are able to implement basic tactics and complex operations (i.e., combined arms) in battle—no matter whether they win or lose.⁷ An example of the latter view is Biddle's conception of effectiveness as “the ability to destroy hostile forces while preserving one's own; the ability to take and hold ground; and the time required to do so,” a definition which is inherently comparative and best measured by looking at combat outcomes such as loss-exchange ratios or ground gained or lost.⁸

⁴ Edward A. Shils and Morris Janowitz, “Cohesion and Disintegration in the Wehrmacht in World War II,” *Public Opinion Quarterly* 12:2 (Summer 1948): 280-315, DOI: <https://doi.org/10.1086/265951>; Omer Bartov, *Hitler's Army: Soldiers, Nazis, and War in the Third Reich* (New York: Oxford University Press, 1992); Stephen G. Fritz, “‘We Are Trying ... to Change the Face of the World’—Ideology and Motivation in the Wehrmacht on the Eastern Front: The View from Below,” *Journal of Military History* 60:4 (October 1996): 683-710, DOI: <https://doi.org/10.2307/2944661>; Reiter and Stam, *Democracies at War*; and Michael C. Desch, *Power and Military Effectiveness: The Fallacy of Democratic Triumphalism* (Baltimore, Md.: Johns Hopkins University Press, 2008). For an argument that emphasizes cohesion at both the national level and within the military, see Jasen J. Castillo, *Endurance and War: The National Sources of Military Cohesion* (Stanford: Stanford University Press, 2014).

⁵ Stephen Biddle, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton: Princeton University Press, 2004).

⁶ Stephen Biddle and Robert Zirkle, “Technology, Civil-Military Relations, and Warfare in the Developing World,” *Journal of Strategic Studies* 19:2 (June 1996): 171-212, DOI: <https://doi.org/10.1080/01412399608437634>; James T. Quinlivan, “Coup-proofing: Its Practice and Consequences in the Middle East,” *International Security* 24:2 (Fall 1999): 131-165, DOI: <https://doi.org/10.1162/016228899560202>; Ulrich Pilster and Tobias Böhmelt, “Coup-Proofing and Military Effectiveness in Interstate Wars, 1967-99,” *Conflict Management and Peace Science* 28:4 (2011): 331-350, DOI: <https://doi.org/10.1177/0738894211413062>; Caitlin Talmadge, *The Dictator's Army: Battlefield Effectiveness in Authoritarian Regimes* (Ithaca, N.Y.: Cornell University Press, 2015); and Vipin Narang and Caitlin Talmadge, “Civil-Military Pathologies and Defeat in War: Tests Using New Data,” *Journal of Conflict Resolution* 62:7 (August 2018): 1379-1405, DOI: <https://doi.org/10.1177/0022002716684627>. Some argue that democracies are more likely to implement the modern system. See Ryan Grauer and Michael C. Horowitz, “What Determines Military Victory? Testing the Modern System,” *Security Studies* 21:1 (February 2012): 83-112, DOI: <https://doi.org/10.1080/09636412.2012.650594>.

⁷ Talmadge, *Dictator's Army*.

⁸ Biddle, *Military Power*, 6.

Lyall's book fits solidly in the war-fighting camp, which focuses on how armies perform once engaged in combat, but breaks new ground in defining battlefield effectiveness as including elements of both will and skill—and treating skill as both a property and an outcome. In fact, and in contrast to much recent work, three of the four measures that comprise Lyall's battlefield performance index are indicators of soldiers' will to fight: mass desertion, mass defection, and the deployment of fratricidal violence (blocking detachments) by an army against its own troops. Yet Lyall does not neglect skill, which he assesses by checking belligerents' loss-exchange ratios quantitatively and the sophistication of their tactical and operational art qualitatively. The former, of course, is an outcome measure whereas the latter is a property of an armed force. Lyall thus offers the most comprehensive understanding and measurement of battlefield effectiveness available in the literature—and gathers new data on these indicators, several of which have never before been used in the study of battlefield performance.

On the independent variable side of the equation, however, Lyall's focus on military inequality meshes well with a literature that prioritizes domestic variables. The modern system of force employment, after all, is a unit-level characteristic that varies across states, and studies that seek to explain why some armies are capable of executing it have likewise focused on domestic variables such as regime type, civil-military relations, organizational culture, and now military inequality.⁹ What all of these studies share is a deep skepticism of the view prevalent in realism that states build and optimize their armies to meet external threats; that one can assume optimal force employment and thus assume it away; that analysts can measure military power by simply knowing how many troops or tanks each side has; or generally the view that armies are interchangeable.¹⁰

However, Lyall's argument pushes beyond existing unit-level explanations for variation in force employment and battlefield cohesion. He provides an original alternative account based on what is essentially a self-inflicted dilemma: leaders sometimes discriminate against or violently exclude ethnic groups yet cannot generate sufficient military power without relying on these same groups. Incorporating suspect populations into the armed forces, however, creates a problem with no good solution that leads to pathologies—including desertion, defection, and coercion against one's own troops—that are unrecognized by the current literature. Such militaries, as Lyall amply demonstrates, are prone to crumble on battlefield.

Engaging with the Argument and Evidence of Divided Armies: Questions for Further Investigation

Divided Armies is a truly exceptional book that all students of military effectiveness should read. Like all great works, it raises new questions that the next wave of scholars should address. In this section, I raise four such questions. For the first two, I suggest tentative answers; for the remaining two, I simply pose them for further research.

1. Why Do Leaders Construct Exclusionary Domestic Orders? According to Lyall, the very first choice a leader faces once assuming office is “how to construct, and then sustain, a vision of the political community that transfers the primary allegiance of the population from various subnational (“subordinate”) group identities to a collective (“superordinate”) one

⁹ In addition to the works already cited, see Risa A. Brooks, *Shaping Strategy: The Civil-Military Politics of Strategic Assessment* (Princeton: Princeton University Press, 2008); Ryan Grauer, *Commanding Military Power: Organizing for Victory and Defeat on the Battlefield* (Cambridge: Cambridge University Press, 2016); and essays by Dan Reiter, Timothy Hoyt, Deborah Avant, and Risa Brooks in *Creating Military Power: The Sources of Military Effectiveness*, edited by Risa A. Brooks and Elizabeth A. Stanley (Stanford: Stanford University Press, 2007).

¹⁰ For studies that articulate one of these views, see John J. Mearsheimer, “Assessing the Conventional Balance: The 3:1 Rule and Its Critics,” *International Security* 13:4 (Spring 1989): 54-89, DOI: <https://doi.org/10.2307/2538780>; Barry R. Posen, “Nationalism, the Mass Army, and Military Power,” *International Security* 18:2 (Autumn 1993): 80-124, DOI: <https://doi.org/10.2307/2539098>; Charles L. Glaser and Chaim Kaufmann, “What Is the Offense-Defense Balance and Can We Measure It?” *International Security* 22:4 (Spring 1998): 44-82, DOI: <https://doi.org/10.1177/0022002712438350>; and John J. Mearsheimer, *The Tragedy of Great Power Politics* (New York: W.W. Norton, 2001).

and the political organization that claims to represent it” (41). In other words, leaders need to nation build.¹¹ One question that arises right away is whether every leader actually faces this choice upon assuming office; in many states the political community will already be well-defined and thus leaders will have no need to reconstruct it out of whole cloth. Such choices would seem to be more common at founding moments—when a political unit comes into being—and transitional moments—when a new leader with different characteristics and priorities from her predecessor ascends to power in a poorly institutionalized state.

For Lyall, leaders have a choice about the type of political community they wish to build: inclusive or exclusive. Inclusive identities “draw on civic conceptions of the community that are stripped of specific group characteristics in favor of a more accessible, perhaps substantively thinner, national identification” (42). In the literature on nationalism, this broader, more inclusive definition of the community is known as “civic nationalism.”¹² Exclusive identities, by contrast, treat “targeted groups as peripheral to, or even outside, the broader community even as they remain within its political boundaries” (42). This narrower form of identification, often (as in Lyall’s theory) constructed along ethnic lines, is referred to in the literature as “ethnic nationalism.” The question is on what basis to draw the line defining who is in and who is out.

Interestingly, defining the “in-group” in ethnic terms contradicts the definition of nation-building set out by Lyall above. If nation-building is about transcending particularism and building superordinate identities tied to the broader political community, then strategies of ethnic exclusion appear to be the opposite of nation-building. Such strategies purposefully exclude some groups for any number of reasons. Perhaps the leader’s own group composes a minimum winning coalition and there is no need to assimilate others. Strategies of exclusion may also convey tacit acceptance of the fact that some boundaries are unlikely to be transcended. Whatever the reason, some leaders decide to define certain groups as second-class citizens lacking full rights, whereas others may even be defined as enemies.

Lyall argues that leaders’ choices between inclusive and exclusive identities are not a function of ethnic diversity, state strength, or regime type (44-46). But what if they are a function of perceived threats to the leader’s political survival? If leaders choose exclusive identities to block or persecute groups they perceive could threaten their hold on power, then ethnic threat is the real explanatory variable. Such threats would explain the choice to exclude other ethnic groups, thus generating high levels of military inequality, which in turn correlates with poor battlefield performance.

A good case for examining this supposition is Lyall’s theory-building case: the Mahdi State in Sudan. Upon taking power, the ‘Mahdi’ (Muhammad Ahmad bin Abdullah) sought to build an inclusive identity that would unify Sudan’s fractious tribes and sects. In this task he was no doubt aided by his claim to be the Mahdi, through which “he could access a wellspring of charismatic authority rather than narrowcast his claims to power on the basis of potentially divisive subordinate identities” (108). Unfortunately, the Mahdi died unexpectedly in 1885 without designating a successor (indeed, as the Mahdi, he was not supposed to die at all!). Of the numerous claimants to power, none of them could claim the unifying mantle of their predecessor’s exalted status. One of the Mahdi’s four khalifas, Abdallahi ibn Muhammad, a member of the Ta’isha sub-tribe of the larger Baggara tribe, held the advantage because his troops were stationed nearby, enabling him to seize power. The problem was that other groups disputed Abdallahi’s claim, most notably the Mahdi’s own tribe, the Ashraf; as Lyall puts it, “He faced a number of immediate potential challenges in claiming power” (93).

In other words, the “Khalifa,” as he came to be known, faced multiple threats to his authority but could not appeal to divine sanction to legitimate his rule. He thus turned to force, relying on those he could trust—members of his own tribe—to back him. The first targets of repression, unsurprisingly, were the tribes that “represented the strongest counterclaim to his

¹¹ Although Lyall uses this term once on page 42, the term is more noticeable for its absence, especially since nation-building is exactly what Lyall is describing.

¹² The origin of the two types of nationalism discussed in this paragraph may be found in Hans Kohn, *The Idea of Nationalism: A Study in Its Origins and Background* (New York: Macmillan, 1944). See also Michael Ignatieff, *Blood and Belonging: Journeys into the New Nationalism* (London: Farrar, Straus, and Giroux, 1993), 6-8.

authority,” including the Ashraf (125). Although perhaps not inevitable, it seems plausible that the Khalifa’s turn to violence was a rational (and predictable) response to the threats he faced to consolidating his rule. Leaders who emerge in environments in which multiple ethnic networks are competing for power might thus be (1) more likely to rely on their own ethnic network for support, and hence (2) engage in the kind of exclusive practices that generate high levels of military inequality.

2. What Explains Variation Within Categories of Military Inequality? Lyall convincingly demonstrates that there are real and significant differences in battlefield performance between armies that vary in military inequality. Militaries in the lowest quartile of inequality are the least likely to suffer greater casualties than they inflict, experience mass desertion or defection, or deploy blocking detachments to force their own troops to fight (166). As military inequality increases, armies are more likely to suffer worse casualty ratios, lose large numbers of troops to desertion and defection; and turn their guns against their own soldiers. Notably, however, one quarter of all armies in the lowest quartile of military inequality—119 in all—kill fewer of the enemy than the enemy kills of them, and more than one-fifth of these armies (99 cases) suffer mass desertion on the battlefield. These proportions are lower than for armies with higher levels of inequality but still represent a substantial number. Why do these militaries, which face no impediments to battlefield effectiveness according to the argument in *Divided Armies*, nevertheless perform so poorly?

Lyall’s answer to this question is that even militaries with the lowest levels of inequality may still be insufficiently unified to perform well on the battlefield. An example of such a case from the book is Morocco in the Hispano-Moroccan War of 1859-1860. The armies of both Morocco and Spain were characterized by extremely low levels of military inequality—the former’s score was essentially zero and the latter’s also barely registered (0.03). Yet, as Lyall recognizes, “despite possessing low prewar military inequality, Morocco still underperformed in its battlefield performance, particularly in its loss-exchange ratio” (246).¹³ Lyall describes Morocco’s prewar army as a model of inclusion stitched together by a regime that had “successfully articulated an inclusionary ‘mosaic’ vision of the political community that avoided feeding group-based distinctions and tensions” (205). Yet Lyall later explains Morocco’s battlefield underachievement by pointing to the “weakness of the overarching political community” that left Moroccan forces “unable to match the motivation and organization of [their] Spanish foe” (247). “Persistent tribalism and incomplete affinity for the Makhzan [the Moroccan state],” Lyall continues, “introduced substantial heterogeneity of motive and resolve of Moroccan forces that imposed a ceiling on its performance despite the Makhzan’s inclusionary vision” (247).

Lyall may be correct that military inequality does not always imply cohesion—although such a claim raises questions about the internal validity of the variable as a proxy for the underlying concept of cohesion—but the answer may also lie in other explanations for battlefield effectiveness. Just because an army is built from groups of citizens that suffer little discrimination or violence at the hands of their government does not mean that it will be free from the numerous other problems that existing studies of military effectiveness have highlighted. For example, Egypt and Israel in 1967 fielded militaries that were roughly equal in their levels of inequality—the Israeli Defense Forces (IDF) actually had a slightly higher military inequality score (0.145) than did the Egyptian military (0.05)—yet one launched a bold surprise attack and demonstrated shocking military proficiency while the other saw its air force destroyed on the ground and its troops routed from the Sinai Peninsula.¹⁴ Although Egypt was hardly plagued by military inequality, the country was characterized by a debilitating power struggle between President Gamel Abdel Nasser and his military chief, Abdel Hakim Amer, whose preferences also diverged regarding how aggressively Egypt should confront Israel. This condition of shared power and high preference divergence shut down information sharing and coordination between the two, causing Nasser to overestimate his

¹³ The Moroccan Army managed to kill only about four-tenths of a Spanish soldier per Moroccan soldier killed.

¹⁴ Military inequality varies from zero to one. Data are taken from the replication materials for *Divided Armies*, available online at the Harvard Dataverse, DOI: <https://doi.org/10.7910/DVN/DUO7IE>.

military's preparedness for war and correspondingly Egypt's chances of victory. He thus stood firm when he should have backed down and Egypt suffered a shattering defeat.¹⁵

Similarly, the armed forces of France and Germany in 1940 also had comparably low levels of military inequality (0.04 and 0.075, respectively), yet one executed a daring plan with courage and panache while the other collapsed ignominiously. The French Army was not built out of excluded groups, yet it was crushed by the German *Wehrmacht* in a six-week campaign during which two million of its soldiers surrendered. Although numerous factors help explain poor French performance in this case, a compelling case has been made that France's Left-Right schism resulted in an army that lacked cohesion and the will to fight.¹⁶ Despite numerous opportunities to counterattack and cut off the German armored spearheads, French forces repeatedly stumbled and failed to seize the initiative, with some officers even faking orders to retreat.¹⁷

These examples suggest that although low military inequality may be necessary for good battlefield performance, it is not sufficient, and indeed is no guarantee of success. Armies with low inequality have performed both brilliantly and disastrously. To explain the differences among armies at similar levels of inequality likely requires other theories that emphasize different factors. Indeed, one reason why scholars prior to Lyall may have overlooked inequality is that many of the canonic cases that form the evidentiary base for the military effectiveness literature—World Wars I and II on the Western Front, World War II in the Pacific, Korea, and Vietnam—were fought between combatants at similar levels of military inequality. Lyall, by compiling a comprehensive dataset of conventional wars and plumbing cases in which military inequality varies, is able to uncover the effects of this variable where previous studies—by virtue of their research designs—could not. The field is indebted to him for expanding the case universe and exploring this variation.

3. *Why Does Recruitment from Non-Core Groups Vary?* Lyall correctly argues that most leaders who propagate an exclusive identity cannot afford to recruit solely from the favored core group in building their militaries. These states thus end up with mixed armies composed of a combination of core and non-core groups—and thus face the dilemmas of how to design units to minimize the vulnerabilities created by the inclusion of potentially disloyal and disobedient troops. Yet some leaders and states do choose to close the ranks of their armies to excluded groups.

As Lyall notes, one example of this phenomenon was the Confederate States of America during the American Civil War. Despite facing a numerically superior foe in the Union—and one which began recruiting from its own (excluded) black population as the war progressed—the Confederacy refused to expand its base of manpower by doing the one thing that might have evened the score—bring slaves into its military. Some Confederate generals understood that the South would be swamped by the Union's numbers unless it exploited the last remaining source of manpower available to it. Lyall quotes an assessment by Patrick Cleburne, commander of the Confederate Army of Tennessee, who argued that enlisting blacks would “enable us to have armies numerically superior to those of the North, and a reserve of any size we might think necessary...[and] enable us to take the offensive, move forward, and forage on the enemy” (86). But top Confederate political leaders vetoed the idea; as the *Richmond Examiner* put it, “The existence of a negro soldier is totally inconsistent with our political aim and with our social as well as political system” (86). The Union, by contrast, which also discriminated against its black population, eventually enrolled nearly 180,000 black soldiers over the course of the war.

¹⁵ For this interpretation of the 1967 case, see Brooks, *Shaping Strategy*, 63-101.

¹⁶ Castillo, *Endurance and War*, 94-127. For other explanations of French defeat, see Don W. Alexander, “Repercussions of the Breda Variant,” *French Historical Studies* 8:3 (Spring 1974): 459-488, DOI: <https://doi.org/10.2307/286023>; Ernest R. May, *Strange Victory: Hitler's Conquest of France* (New York: Hill and Wang, 2000); and Williamson Murray and Allan R. Millett, *A War to Be Won: Fighting the Second World War* (Cambridge: Belknap Press, 2001), 63-90.

¹⁷ Castillo, *Endurance and War*, 116.

In Lyall's story, even states that violently exclude portions of their populations end up relying on those same populations for military power. Why would a state in dire need of military manpower eschew a ready supply of soldiers that could help it stave off defeat and—in this case—potentially win independence? What explains variation in the willingness of leaders to exploit these populations?

4. Do Wartime Commitment Problems Hinder Recruitment from Non-Core Groups? Finally, one question that arises in Lyall's theory is why leaders are unable to revise the social order in order to grant rights to previously excluded groups to get them to fight harder for the state in the current war. As Lyall puts it, when faced with difficult conditions in wartime, "why don't leaders simply shift prevailing norms of citizenship in a more inclusive direction to unlock additional military power" (84)? The problem, as he points out, is that such commitments are not credible: any regime that grants concessions during an emergency has incentives to renege once the emergency passes.¹⁸ Given this commitment problem, it is unclear why any excluded group would trust a regime's wartime promises to reverse its past discriminatory or violent practices.

Lyall's cases, however, reveal significant variation in the willingness of non-core groups to believe leaders' promises despite their lack of credibility. In Kokand, for example, Lyall writes that 'Alimqul offered "official forgiveness and an 'amnesty for their past transgressions' to all non-core tribes that increased their supply of soldiers to his army" (248). Unfortunately, "[h]e found few takers" among the tribes, which recognized his appeal for the "pragmatic, if not openly cynical, gesture" that it was (248). Similarly, in Austria-Hungary during World War I, the Dual Monarchy was "compelled by wartime setbacks to offer at least piecemeal concessions" to various non-core groups in the empire, but these proposals "were hamstrung...by the entirely understandable belief among nationalist leaders that imperial authorities might simply claw back any wartime concessions once the danger had passed" (313).

These examples are counterbalanced by a number of instances in which incredible appeals to non-core groups by embattled regimes appear to have worked. In Ethiopia on the eve of its war with Eritrea (1998-2000), for example, the officer corps was dominated by core groups (Tigrayans and Amhara) while non-core groups (Oromo, Afar, Somali) were almost completely shut out. As battlefield setbacks mounted, however, "recruitment was thrown open to all ethnic groups" and the regime issued promises that the "social contract would be revised after the war" (331). Apparently, this worked: Lyall writes that non-core groups' "[r]esentment over their second-class status was tempered with the realization that fighting well could lead to a revision of the social contract, as promised by the regime" (347). Why would non-core groups in Ethiopia find such promises credible, however, when those in Kokand and Austria-Hungary did not? Similarly, in the Soviet case, Lyall notes that Joseph Stalin, who had been "[p]ushed to the wall" by the Nazi invasion, "modified his appeals toward national minorities, particularly Central Asians," and "cleverly encouraged hopes that the status of these groups would be elevated in the postwar order" (398). These policies, although "tactical modifications...did manage to increase recruitment among Central Asians" (398). A few pages later, though, Lyall argues that "Stalin was largely entrapped by a credible commitment problem, having lashed his regime (and himself) to the mast of a Soviet vision that foreclosed a more egalitarian political community" (403). If Stalin was so boxed in by his ideology, however, and non-core groups understood that his promises of political inclusion were merely expedients driven by military desperation, why did these policies generate non-core recruits?

In closing, let me reiterate what a landmark achievement *Divided Armies* is. In making his argument for—and offering compelling evidence of—the effects of military inequality on battlefield effectiveness, Lyall has not just advanced the ball in the study of this subject, he has run through the end zone and out of the stadium with it. I look forward to teaching it to new crops of students and to the debates it will surely spawn.

¹⁸ Similarly, excluded groups that gain additional rights during wartime have incentives to keep pushing for additional concessions, which threatens the regime.

REVIEW BY KRISTEN A. HARKNESS, UNIVERSITY OF ST. ANDREWS

Divided Armies places identity and how leaders imagine the political community at the very heart of our understanding of military power. To discriminate and commit violence against ethnic groups undermines the state's ability to fight and win wars. Either these groups must be excluded from the armed forces, limiting manpower, or their inclusion results in fears (and acts) of disloyalty that subvert tactics and operations. This is a most welcome correction to the literature on military power and combat effectiveness that often ignores identity.¹

Additionally, *Divided Armies* deeply enhances our understanding of the role of ethnicity within security forces. Its underlying dataset, Project Mars, represents an extraordinary effort to collect and code global data on ethnic recruitment and representation within military institutions—albeit only those that fought conventional battles—and the pre-war treatment of those groups by the state. The geographic and temporal range of the data, spanning over two centuries, is impressive, as are the variety and depth of primary sources consulted, from archival documents to regimental histories to war journalism. This is a valuable contribution to broader comparative, constructivist efforts to move beyond ethnic demography and analyse how the ethnic practices of leaders and states shape societies and their politics.

Without detracting from these achievements, I would argue that Lyall's book could have more critically engaged with: (1) the role of the adversary in battle outcomes and (2) the fluidity of ethnic identity during the era of conquest and colonialism. Delving deeper into these issues, moreover, elucidates opportunities for furthering *Divided Armies* already rich research agenda.

First, battle and war outcomes are inherently relational. The adversary matters, from its motivation and investment in war objectives, to its doctrine and ability to adapt in the field, to the organization and quality of the forces it deploys.

While the book does account for some limited characteristics of the adversary, it concentrates heavily on only one side of any given conflict. The statistical analysis and matching covariates for case study selection do include initial relative power, whether the opponent is democratic or a colonizer, whether a major religious fault line divides the belligerents, and whether they have recently fought each other (97 & 160-164). Excluded, however, are important characteristics such as whether the adversary is coup proofed, how far it must project power to fight, and the population and economic might it has to draw upon, not to mention harder to quantify criteria such as motivation and limited war aims. This also leads the paired comparisons to be less controlled than perhaps they could be, as the opponents within the matched cases may vary quite substantially.

Indeed, relative neglect of the adversary could make a compelling difference in the comparison of the First and Second Mahdist wars. On the one hand, this is a very cleverly matched pair that carefully controls for possible contextual variation—on the Mahdist side. Not much changed within this state between wars except for the sudden and random death of the Mahdi and the choice by his successor to transform an inclusive political community into a stringent, exclusive ethnic hierarchy (90-103). The evidence convincingly demonstrates how ethnic groups that experienced violence at the hands of the new leader fought particularly poorly in the later war, deserting and defecting en masse.

Yet, the adversary changed substantially between the Mahdist conflicts. In the first war, the Mahdist army faced a predominantly Egyptian force (under the auspices of British rule). This was a war of independence, for the creation of a new

¹ See, for example, Stephen Biddle, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton: Princeton University Press, 2004); Risa A. Brooks and Elizabeth A. Stanley, *Creating Military Power: The Sources of Military Effectiveness* (Palo Alto: Stanford University Press, 2007); Klaus Eugene Knorr, *The War Potential of Nations* (Princeton: Princeton University Press, 1956); Jack Levy, "The Offense/Defense Balance of Military Technology: A theoretical and Historical Analysis," *International Studies Quarterly* 28:2 (1984): 219-238; Caitlin Talmadge, *The Dictator's Army: Battlefield Effectiveness in Authoritarian Regimes* (Ithaca: Cornell University Press, 2015).

state, and what the Mahdists encountered were local opponents. The British government was hesitant to intervene, with Prime Minister William Gladstone even remarking that the Sudanese were “a people struggling to be free and they are struggling rightly to be free.”² It was only after the Mahdist army slaughtered Egyptian and European civilians at Sinkat that Gladstone’s government relented to public outcry and Queen Victoria’s personal appeal, sending a contingent of 4000 British soldiers on a limited reprisal mission. It arrived in February 1884, fought and handily won two battles at El Teb and Tamai, and then withdrew. As the military situation further deteriorated, an even larger relief expedition was dispatched to rescue the Governor-General in Khartoum, but it arrived too late. These British troops lingered throughout the summer, engaging in skirmishes with the Mahdists, but were rather quickly withdrawn as more pressing interests in Afghanistan required their redeployment.³

Eleven years later, however, the British government became deeply concerned for its colonial empire and encroachments by its competitors into Sudanese territory. A determined British government invaded, sending one of its most historically renowned commanders, Lord Herbert Kitchener, and deploying a large contingent of home units (8,200 strong) with superior training, equipment, and experience to local Egyptian auxiliaries, who nevertheless comprised an important part of the overall force (123).

Did this shift, from an Egyptian to British adversary, impact battle outcomes or the behaviour of Mahdist soldiers? This is an open empirical question, and not one that can be thoroughly answered here. Yet, the limited deployment of British troops in the First Mahdist War provides some clues. At least on one measure, the first Mahdist army performed quite poorly when fighting British rather than Egyptian troops. The loss-exchange ratios at El Teb and Tamai were both well below parity—while the Mahdists typically achieved above parity ratios when fighting the Egyptians (115). The skirmishes in the summer of 1885, after the book’s case study ends, reinforce these patterns. In a second battle near Tamai, the British killed 1000 Sudanese while suffering under 200 fatalities themselves.⁴ Similarly, at Kosheh, numerically inferior British forces defeated a large contingent of Mahdists (possibly 6000 strong), inflicting 800 casualties while suffering only 37 themselves.⁵ These loss exchange ratios appear roughly similar to those suffered by the Mahdists in the second war (115). No matter their vision of the political community, or degree of inclusiveness, the Mahdists could not defeat the British army on the battlefield.

To take the adversary into account does not necessarily undermine the book’s argument. Indeed, if the theory is correct, then perhaps the most vital factor to analyse in any opponent is its degree of military inequality. This in turn suggests an important and logical theoretical extension: how the military inequality of belligerents interact as they clash. We would expect that when a low or medium inequality actor faces an adversary with high inequality, it would likely win on the battlefield (see Figure 1). Returning to the Mahdist Wars, this interaction could further help us understand how the change in adversary mattered. The Egyptian army in 1889 had a higher military inequality score, 0.425, than the Mahdist army at 0.01. By 1896, a stark reversal occurred with the second Mahdist army witnessing much greater inequality than the forces fielded by the British, 0.67 versus 0.219.⁶ Not only did the Mahdist state become more exclusionary, its adversary became less so—reinforcing predictions that the Mahdist military would struggle deeply in the second war.

² As quoted in Edward M. Spiers, *The Victorian Soldier in Africa* (Manchester: Manchester University Press, 2004), 112.

³ Spiers, *The Victorian Soldier*, 99-128.

⁴ Spiers, 124.

⁵ Spiers, 127.

⁶ I would like to thank Jason Lyall for providing transparent access to the Project Mars codebooks and replication data.

Figure 1: Interaction Effects of Military Inequality on Battle Outcomes

| | Side A | |
|-----------------|----------------|-----------------|
| Side B | Low Inequality | High Inequality |
| Low Inequality | Indeterminate | B wins |
| High Inequality | A wins | Indeterminate |

Outcomes seem less clear when both belligerents evidence similar levels of inequality. Perhaps when neither army suffers much inequality, our existing theories on technological prowess, sheer numbers, or combined arms proficiency operate. This could help explain why low inequality failed to help the first Mahdist army against the limited deployments of British home units. With, respectively, inequality scores of 0.01 and 0.219 (assuming roughly the same score for the British as eleven years later), neither military was straightjacketed by the pathologies of ethnic inequality, allowing other variables to predominate.

This interactive framework raises another important question: who prevails when both parties are highly unequal? The Iraq/Islamic State comparison in the book's conclusion highlights the utility of theorizing this particular box. Both the Iraqi government and ISIS held heavily exclusionary visions of the political community and inflicted violence on ethnic groups they viewed as being outside of that community. Both should thus predictably have experienced the types of pathologies emphasized by *Divided Armies*: poor combat performance, desertions, defections, and coercion against their own soldiers. These dysfunctions are, moreover, exploitable weaknesses by a wily adversary. Could we theorize, given mutually high levels of inequality, which army will best mitigate its own weaknesses while taking advantage of those of its adversary?

Second, I would push *Divided Armies* on its conceptualization and treatment of ethnic identity, particularly in the era of colonial conquest and resistance to colonialism. Project Mars is built on the All Minorities at Risk (AMAR) categorizations of socially relevant ethnic groups and subgroups. AMAR is one of the first datasets to map global ethnic demography without conditioning inclusion on political relevance.⁷ Yet, it provides only a single list of contemporary groups per country—it is entirely ahistorical. Project Mars thus reads these groups backwards in time, for over 200 years, into what was a far messier and more fluid past. I find this worrisome, for the reasons articulated below.

The book assumes a relatively high degree of coherence and stickiness to ascriptive identity categories, at least enough to consider ethnic groups as (1) possessing a sense of collective groupness prior to the war, (2) understanding their treatment by the state in terms of that groupness, and (3) persisting intact through the experience of war itself, especially their boundaries—that is, who can claim group membership and who cannot.

These assumptions may hold for some groups, but perhaps not others. Especially in regions like Africa, we know that ethnic identity was altered, and occasionally invented, by colonial powers during conquest and the territorial changes, state reorganization, and repression that followed.⁸ For example, precolonial identities in Rwanda/Burundi were largely politically and economically defined, with mobility between Hutu and Tutsi identities possible based on the accumulation, or loss, of wealth and status. Belgian missionaries and the colonial administration racialized these identities and, in the 1930s, rigidified their boundaries—categorizing every individual, noting their ethnicity on identity documents, and

⁷ Jóhanna L. Birnir, Jonathan Wilkenfeld, James D. Fearon, David Laitin, Ted Robert Gurr, Dawn Brancati, Stephen Saideman, Amy Pate, and Agatha S. Hultquist, "Socially Relevant Ethnic Groups, Ethnic Structure and AMAR," *Journal of Peace Research* 52:1 (2015): 110-115, DOI: <https://doi.org/10.1177%2F0022343314536915>.

⁸ Donald L. Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985), 66-74.

enforcing the paternal inheritance of identity. What being Hutu or Tutsi meant in the nineteenth century was thus quite different from in the eighteenth century.⁹

Resistance to colonialism also changed ethnic identities, particularly the complex relationships between ethnic subgroups and how they perceived themselves in relation to broader categorizations. In colonial Kenya, the Kikuyu, Embu, and Meru were considered branches of an overarching identity which shared a common language and cultural tradition. While each occupied a separate (but neighbouring) territory under the tribal homeland administrative system, they were often treated as a single group, clumped together in such important state practices as the census and military recruitment quota system. But the Mau Mau rebellion of the 1950s broke down this sense of a broader ethnic community. Wishing to escape from the harsh reprisal violence of the colonial state, the Embu and Meru petitioned to be separated from the Kikuyu and to have governance over their homelands transferred to an entirely different province.¹⁰

These illustrative anecdotes highlight that ethnic identity formation can be fluid and endogenous to conquest and war; although not always and not always in the same ways. This matters insofar as individuals could change and alter their identities, or whole group categorizations, in response to the violence and opportunities of warfare—which would alter how pre-war treatment of groups shapes individual behaviour during conflict, especially perhaps during civil wars.

Indeed, the problems of ethnic categorization and fluidity worsen the further back in time we travel. Today, ethnonationalism—the idea that the state belongs to the nation, which itself is comprised by a particular ethnic community—pervasively orders political thinking. It also shapes how individuals perceive their relationship to the state and their understanding of the consequences of group exclusion and violence.¹¹ Yet, this was an outside mode of conceptualizing identity and structuring state power that was imposed on most of the world via colonialism and its institutions.¹² In other words, colonialism did not merely alter the boundaries or even existence of particular communal groups, it changed the very idea of group identity. To apply modern ethnic categories to colonized societies risks missing the great transformations of identity that occurred. To apply them to pre-colonial societies might mischaracterize entirely how individuals conceived of their identities and thus the motivations for their behaviour.

⁹ Mahmood Mamdani, *When Victims Become Killers: Colonialism, Nativism, and the Genocide in Rwanda* (Oxford: James Currey, 2001), 41-102.

¹⁰ Daniel Branch, *Defeating Mau Mau, Creating Kenya: Counterinsurgency, Civil War, and Decolonization* (Cambridge: Cambridge University Press, 2009), 101-103.

¹¹ Lars-Erik Cederman, Andreas Wimmer, and Brian Min, “Why Do Ethnic Groups Rebel? New Data and Analysis,” *World Politics* 62:1 (2010): 87-119. DOI: <https://doi.org/10.1017/S0043887109990219>.

¹² Benedict Anderson, *Imagined Communities: Reflections on the Origin and Spread of Nationalism* (London: Verso, 1983), 163-185.

REVIEW BY MICHAEL C. HOROWITZ, UNIVERSITY OF PENNSYLVANIA

I will start my review of Jason Lyall's excellent book with the bottom line up front, in the form of a limerick.

This book contains a great lesson
(Though it's a lot of pages to express in!)
I read, transfixed,
That if your army's more mixed
The fighting you'll have more success in

Jason Lyall has written a masterful book that is a must-read for scholars of international relations, international conflict, war, and strategy. Through a number of different methodological approaches, he clearly and persuasively demonstrates the way that inequality makes armies weaker and prone to inferior strategy choices, while more inclusive militaries are more likely to succeed. Particularly given the current context in the United States, his book has obvious policy relevance in addition to his academic contribution.

Nonetheless, like any academic book, and certainly a book this detailed at the micro and macro level, there are areas that could have been more persuasive.

One question concerns the way that Lyall defines combat power. Lyall creates a battlefield performance index that features loss exchange ratio as a measure of combat power. This measures efficiency of operations, in a way, since it is the ratio of soldiers one side kills compared to its own soldiers killed (adjusted for army size). The definition of battlefield performance also includes several other measures that one could generally imagine as mechanisms that influence combat power: desertion, defection, and fratricidal violence.

There is a risk of endogeneity here because desertion, defection, and fratricidal violence all potentially influence loss-exchange ratios. There is some risk here from an inference perspective. In Biddle's *Military Power* book,¹ for example, his dependent variable is the loss exchange ratio, and questions surrounding military force employment strategies are independent variables. Lyall is classifying all of these in one index, in a way. Or, put differently, why is it better to define cohesion as part of battlefield performance, rather than as an independent variable that influences battlefield performance?

This also becomes important in chapter 3, when the reader considers Figure 3.1 on page 74. That figure presents battlefield performance as a concept separate from soldier morale. Lyall argues that soldier morale helps drive battlefield performance. But soldier morale is potentially endogenized into the definition of battlefield performance. Morale, after all, is a portion of cohesion, and cohesion is endogenized into how Lyall defines battlefield performance. Thus, there is a risk that Lyall is including morale both as an independent variable and as part of the dependent variable. To be fair, I suspect Lyall has thought this issue through, but the book does not explain it clearly.

Lyall's understanding of tactical and operational sophistication is reasonable, but does seem to bake in an assumption that more is always better, meaning more sophistication makes success more likely. However, academic research suggests that what effective militaries generally need is a match between tactical/operational sophistication and their force employment strategy.² Some less "sophisticated" battlefield strategies may be appropriate for a given war, and the relative advantage of a given combatant. Consider, for example, Union General Ulysses S. Grant's more attrition-focused tactics in the U.S. Civil War versus Confederate General Robert E. Lee's greater interest in maneuver. Given the relative attributes of both sides,

¹ Stephen Biddle *Military Power* (Princeton: Princeton University Press, 2004)

² For example, see James M. McPherson, *Battle Cry of Freedom: The Civil War Era* (New York: Oxford University Press, 2003).

and the superior numbers aiding the North, most research suggests that Grant's strategy was appropriate.³ But does Lyall's framework suggest a critique of Grant for not being more 'sophisticated' in force employment? Essentially, Lyall presents sophistication as a linear variable whereby the more sophisticated, the better, even though military history suggests the answer is more complicated.

One could argue that Lyall's notion of sophistication as a uniformly positive good is limited to a particular period in military history – the period where Biddle's modern system is operative in the twentieth and early twenty-first century. That is a reasonable response, but not something Lyall explicitly considers in the context of the book. Lyall's discussion of tactical/operational sophistication also does not really explain what is similar or different between this and something like Biddle's conception of force employment (of which the modern system is an example). Is sophistication, for Lyall, a proxy for adoption of the modern system? If not, what is the difference?

Chapter 4 presents a broad quantitative test of Lyall's theory. It is generally persuasive, but there are also a few issues here. First, Lyall's models break out 1800-1917 and 1918-2011, based on Biddle's modern system delineation. But this seems to assume the validity of the modern system rather than testing it. But why would this be important for Lyall's Battlefield Performance Index, conceptually? If inequality is an essential element that undermines military success, presumably that should be true throughout time (as long as the conflict size is at a certain level, potentially).

Second, and more important, Lyall's theory is in many ways a theory of battles and when militaries are likely to win battles. But the quantitative tests focus on wars as the unit of analysis, not battles. One could imagine reasonable arguments about how battle performance should scale to cover wars, but even so there are questions that Lyall does not cover in depth. For example, Germany's experience in the Second World War I is generally seen as a classic example of why battles don't scale linearly to cover wars.

Third, Lyall compares his data in many places to the Correlates of War interstate war data.⁴ That is fine and appropriate, since that is the standard. But there are now other new datasets that fix issues with existing data. For example, Gibler's new MID's dataset. Or Reiter/Stam/Horowitz on interstate wars.⁵ What are the differences between this and Project Mars, and how might that influence Lyall's results? I am sure that there are differences, given how much Lyall put into this data gathering effort, and it would be helpful to see those articulated more clearly.

Finally, what is the relationship between regime type and military inequality? One of the interesting things about Lyall's results is the way they contribute, indirectly, to some older debates about regime type and battlefield performance. They provide more evidence that democracies do not always do relatively better on the battlefield, as he points out in the conclusion (page 404). However, what if lower levels of military inequality is a key mechanism that explains some prior findings about the success of democracies because they are relatively more likely to have more inclusive militaries? That would be very interesting. I would be curious to hear more from Lyall about the extent to which regime type might help predict the degree of inclusion by militaries, thus influencing combat power via the mechanisms he suggests. Essentially, there is a possibility of endogeneity here.

I would also be curious to hear how Lyall is thinking, given some of the debate about military inclusion in the Obama administration and the Trump administration, about the future of U.S. military power. The Obama administration

³ For example, see McPherson.

⁴ *Meredith Reid Sarkess and Frank Wayman, Resort to War: 1816–2007* (Washington, D.C.: CQ Press, 2010).

⁵ Dan Reiter, Allan C. Stam, and Michael C. Horowitz, "A Deeper Look at Interstate War Data: Interstate War Data Version 1.1" *Research & Politics*, October-December 2016, 1-3; Douglas M. Gibler, Steven V Miller, and Erin K Little. "An Analysis of the Militarized Interstate Dispute (Mid) Dataset, 1816–2001." *International Studies Quarterly* 60:4 (2016): 719-30.

expanded the military franchise, opening up many previously-closed military occupations to women, and also made the U.S. military more accepting of transgender and non-binary individuals. The Trump Administration, while not reversing the Obama Administration's decision to lift the gender-based combat exclusion rule, did enact restrictions on military service by transgendered individuals. Moreover, the Trump Administration undertook a number of actions to make military service by immigrants more difficult, and narrowing the ability of immigrants serving in the United States military to have a fast track to citizenship. Commentary throughout the Trump Administration suggested that this might undermine efforts to recruit for the US military, and even impact combat effectiveness.⁶ What does Lyall think about this? I presume the logic of his theory would suggest similar concerns on the basis of military power (putting aside the human and ethical issues), but how important will this be? And once the Biden Administration enters office, if these Trump-era policies are reversed, will that immediately fix the issue, or is there a time lag between inclusion and improved combat effectiveness?

⁶ Rebecca Best, Kyleanne Hunter, Theresa Schroeder and Jeremy Teigen, "Military Service Was once a Fast Track to U.S. Citizenship. The Trump Administration Keeps Narrowing that Possibility." *The Washington Post Monkey Cage Blog*, September 6, 2019, <https://www.washingtonpost.com/politics/2019/09/06/military-service-was-once-fast-track-us-citizenship-trump-administration-keeps-narrowing-that-possibility/>.

REVIEW BY YURI M. ZHUKOV, UNIVERSITY OF MICHIGAN

Jason Lyall's *Divided Armies* is an ambitious and important book about the relationship between army and society, which challenges our understanding of the sources of military effectiveness in conventional war. The empirical foundations of this volume rest on a new dataset, Project Mars, spanning 250 conventional wars from 1800 to 2011, with information on belligerents' battlefield performance (e.g. loss-exchange ratios, desertion, defection) and levels of ethnic discrimination (Military Inequality Coefficient, or MIC).¹ Through a comprehensive array of data analyses and matched case comparisons, Lyall shows that states whose ethnic minorities experienced greater prewar and wartime discrimination fared more poorly in battle. The book's findings and methods are likely to be of interest to a diverse audience well beyond scholars of international conflict and security. *Divided Armies* also leaves some important questions unanswered, charting a course for future research.

If one were to imagine how future scholars might cite and engage with this book's theoretical contributions, among the most immediate takeaways is that the army is an extension of society. This by itself is not a novel observation,² but international conflict literature has too often treated the military as a seemingly direct, cohesive agent of the state. There is an implicit assumption that state leaders purposefully initiate all disputes, while the military faithfully carries out their orders.³ Even some literature that relaxes the unitary-state assumption, like research on civil-military relations, often treats the military as a unified entity within that state.⁴ Lyall's book rejects these assumptions, and shows that militaries replicate and reinforce the divisions, inequalities, and patterns of discrimination that exist within broader society. *Divided Armies* marshals some of the most comprehensive cross-national evidence on this topic to date, showing that these inherited inequalities can hamper battlefield resolve and alter the dynamics and outcomes of wars.

In building this argument, Lyall enters into dialogue with an unusually diverse array of social science research. The point of departure for this narrative is the classic military sociology literature on unit cohesion, which expects socialization and primary-group bonding to have a greater influence on soldiers' actions than pre-existing ethnic identities and intergroup relations.⁵ To understand why ethnic inequalities can nonetheless persist within the ranks, Lyall draws on intergroup

¹ <https://doi.org/10.7910/DVN/DUO7IE>

² Stephen Peter Rosen, *Societies and Military Power: India and its Armies* (Ithaca: Cornell University Press, 1996).

³ For example, research using the militarized interstate disputes (MID) or Blechman-Kaplan datasets on uses of force often make this assumption. See Daniel M. Jones, Stuart A. Bremer, and J. David Singer, "Militarized Interstate Disputes, 1816-1992: Rationale, Coding Rules, and Empirical Patterns." *Conflict Management and Peace Science* 15:2 (1996): 163-213; Barry M. Blechman and Stephen S. Kaplan, *Force without War: US Armed Forces as a Political Instrument* (Washington, D.C.: Brookings Institution Press, 1978).

⁴ See, for example, Peter D. Feaver, *Armed Servants: Agency, Oversight, and Civil-Military Relations* (Cambridge: Harvard University Press, 2009).

⁵ Edward A. Shils and Morris Janowitz, "Cohesion and Disintegration in the Wehrmacht in World War II," *Public Opinion Quarterly* 12:2 (1948): 280-315; Samuel Lyman Atwood Marshall, *Men against Fire: The Problem of Battle Command* (Norman, OK: University of Oklahoma Press, 2000); S.A. Stouffer, E.A. Suchman, L.C. DeVinney, S.A. Star, and R.M. Williams, *The American Soldier: Adjustment during Army Life* (Princeton: Princeton University Press, 1949).

contact theory from social psychology,⁶ and recent American politics scholarship on race.⁷ Lyall traces the origins of these inequalities to status hierarchies within political communities, as documented by historical sociology research on ethnic exclusion and state formation.⁸ Each of these scholarly traditions is a load-bearing pillar in Lyall's argument, and each broadens the book's appeal across disciplines.

Another contribution for which *Divided Armies* deserves recognition is its conceptualization and measurement of conventional war. The quantitative study of modern armed conflict has traditionally revolved around the universe of cases in J. David Singer and Melvin Small's canonical Correlates of War (COW) database.⁹ COW divides armed conflicts since 1816 into the somewhat dubious categories of "inter-state" (wars between two or more states), "intra-state" (civil wars), "extra-state" (wars between states and non-state actors) and "non-state" (wars between non-state actors). Beyond the fact that more than one of these labels can simultaneously apply to one war (e.g. Vietnam, Congo, Ukraine), the "extra-state" category has notoriously included virtually all instances where a European colonial power fought a non-Western opponent (e.g. Ashanti Empire, Princely States of India), even if both belligerents controlled territory, collected taxes, distributed public goods, fielded conventional armies, and had all the markings of a state. This categorization has inadvertently limited the scope of empirical inquiries. For example, studies of conventional war have typically included only "inter-state" wars. The Project Mars dataset dispenses with these antiquated and theoretically incoherent categories, and expands the scope of inquiry to all conflicts involving belligerents who fought conventionally and exercised a degree of sovereign control over territory.

Scholars writing multi-method dissertations and books may also look to *Divided Armies* as a model for how to integrate quantitative and qualitative empirical strategies, using the former to inform the latter. Particularly notable is Lyall's use of statistical matching for case selection in a 'most similar' comparative design. This approach offers a transparent and replicable way to select pairs of observations with similar observable background characteristics, while guarding against accusations of 'cherry picking' cases in ways that might bias the findings. While this is not a novel innovation,¹⁰ Lyall offers one of the most formidable applications of this case-selection approach in a project on this scale. The full potential of this strategy is most apparent in Lyall's analysis of Red Army units during the Battle of Moscow in 1941. Some of the more macro-level matched comparisons of countries and wars are less convincing, as I note below. Yet even beyond the qualitative comparative case studies in the book, this approach should be of interest to any scholars who combine cross-national and sub-national quantitative evidence in their work. One could imagine using a cross-national dataset to identify country case studies for which micro-level data are available, and then running a series of disaggregated statistical analysis within those matched country pairs.¹¹

⁶ Thomas F. Pettigrew, and Linda R. Tropp, "A Meta-Analytic Test of Intergroup Contact Theory," *Journal of Personality and Social Psychology* 90:5 (2006): 751-783.

⁷ Ryan D. Enos and Noam Gidron, "Exclusion and Cooperation in Diverse Societies: Experimental Evidence from Israel," *American Political Science Review* 112:4 (2018): 742-757.

⁸ Andreas Wimmer, *Waves of war: Nationalism, State Formation, and Ethnic Exclusion in the Modern World* (New York: Cambridge University Press, 2012).

⁹ Joel David Singer and Melvin Small, *The Wages of War, 1816-1965: A Statistical Handbook* (New York: John Wiley & Sons, 1972).

¹⁰ Richard A. Nielsen, "Case Selection via Matching," *Sociological Methods & Research* 45:3 (2016): 569-597.

¹¹ Yuri M. Zhukov, Christian Davenport, and Nadiya Kostyuk, "Introducing xSub: A New Portal for Cross-National Data on Subnational Violence," *Journal of Peace Research* 56:4 (2019): 604-614.

For all its merits, *Divided Armies* is not without flaws, and there are several places where it is open to potential points of criticism. Perhaps most consequentially, Lyall advances a monadic theory and employs monadic empirical tests to explain what is ultimately a strategic interaction between two or more players. While the book does mention that ethnic inequalities represent critical vulnerabilities that adversaries may attempt to exploit, it largely glosses over the fact that inequalities surely exist on both sides. Lyall doesn't make any claims or offer any predictions as to what we should expect when a high-inequality belligerent fights a similarly high-inequality opponent, when two low-inequality belligerents fight each other, or what the outcomes might be for mixed low/high-inequality dyads. Of course, this story need not take the form of a 2-by-2 table, but simply holding the opponent's inequality level 'constant' is not theoretically satisfying.

Empirically, this monadic orientation leads to some strange paired comparisons. In chapter 6, Lyall compares Turkey's experience in the Italo-Turkish War of 1911 to Austria-Hungary's experience against Russia on the eastern front of World War I. He shows that Ottoman Turkey and Austria-Hungary had different levels of military inequality (MIC), but were otherwise similar on important dimensions like relative power, regime type, and military doctrine. Left unsaid is the fact that the other two belligerents in these matched cases – Italy and Tsarist Russia – were polities with vastly different size and political-military institutions. Most crucially, Tsarist Russia was a territorially contiguous multiethnic empire, which had spent the better part of a century suppressing revolts and otherwise repressing potentially 'disloyal' religious and ethnic groups. Some minorities faced discriminatory conscription quotas and were systematically denied commission as officers. Although Lyall does not report these statistics in the text, it is a safe bet that Russia had a significantly higher (i.e. worse) MIC than Italy. It is therefore unclear whether we can attribute Austria-Hungary's more favorable loss-exchange ratio (relative to Turkey) to the fact that Austria-Hungary had a lower MIC than Turkey, or to the fact that Russia had a higher MIC than Italy. Empirically, the solution to this problem appears simple: the set of matching covariates should include the MIC (and other relevant characteristics) of the opponent. Indeed, Lyall does match the pairs on the regime type and great power status of the opponent, just not MIC. The logic behind this omission is difficult to comprehend.

A related problem is that some of the book's macro-level case studies compare two-party (dyadic) conflicts to multi-party conflicts. To take an example from chapter 7, Ethiopia versus Eritrea was a two-party conflict, but the Second Congo War involved dozens of state and non-state actors on the ground that interacted in complex ways. In some respects, the government of the Democratic Republic of Congo was almost a junior partner in its coalition. Similarly, in chapter 6, the Italo-Turkish War was (for the most part) dyadic. Yet the Eastern Front of World War I was certainly not dyadic, and Austria-Hungary played, at most, a secondary role in the disintegration of the Russian imperial army. Austria-Hungary's main ally, Germany, had an army with vastly different ethnic composition, and an almost certainly lower MIC. Given the complex nature of this conflict, why should we attribute Austria-Hungary's battlefield performance to its own MIC, and Germany's (or Bulgaria's, Turkey's, Russia's, Serbia's, Romania's, etc.)? The empirical solution here is also straightforward: the number of belligerents to a conflict should be among the attributes used in matching.

Finally, it is possible to question the temporal scope of Lyall's theory and analysis. The book focuses on recent and ongoing exposure to state discrimination and repression, which is operationalized as the five years preceding a war. Yet recent literature on the long-term legacies of violence has shown that repression casts a much longer shadow.¹² A natural extension of this project would be to see how the results might change as one extends the time window, and considers the effects of ethnic discrimination and repression on a multi-year, multi-decade and inter-generational timescale. It is quite possible that Lyall's current MIC metric underestimates the amount of discrimination that societies face, and more importantly, what they remember. We may not know the true 'half-life' of historical discrimination, but it is surely longer than five years.

Divided Armies is an extremely engaging and provocative book. Like a good war movie, its cold touch lingers, and stays with you long after you put it down. For all of the reasons outlined above (including the more critical points), this book should be required reading in undergraduate classes on security and conflict, graduate field seminars on international relations, and

¹² Christian Davenport, Håvard Mokleiv Nygård, Hanne Fjelde, and David Armstrong, "The Consequences of Contention: Understanding the Aftereffects of Political Conflict and Violence," *Annual Review of Political Science* 22 (2019): 361-377.

applied methods courses on the empirical study of political violence. This book offers many touchstones for future scholarship on military effectiveness, and any serious student of the topic would be wise to give it a close and thorough read.

RESPONSE BY JASON LYALL, DARTMOUTH COLLEGE

I count myself extremely fortunate to have had such a wonderful group of scholars engage with my book, *Divided Armies*, and all the more so during a global pandemic. I thank Alexander Downes, Kristen Harkness, Michael Horowitz, and Yuri Zhukov for their thoughtful reviews. I also thank Dan Reiter for organizing the roundtable and for writing the introductory essay.⁶⁴ Taken together, their generous comments raise important questions about the book's argument and evidence. Perhaps most importantly, their interventions highlight the promise of a research program that is centered on inequality and its effects on political violence, including battlefield performance in past and future wars.

Let me first take a moment to summarize *Divided Armies'* core claims and contributions. The book was designed around a simple insight: our theories of military effectiveness have largely overlooked how inequality *within* armies ("military inequality") can help explain their battlefield performance in wars since 1800. Much as a Gini coefficient measures income inequality, I view military inequality as an index that captures the ethnic and racial hierarchies that exist within armies. More specifically, the index has two components. First, I measure the ethnic makeup of a given army, tracking the share of fielded forces that each ethnic group represents. Second, I note the state's prewar treatment of each ethnic group, ranging from full inclusion in the political community to targeted discrimination to outright repression. When summed, these values create an index that runs from 0 (perfect equality) to 1 (perfect inequality), giving us a grammar for expressing inequality as well as behavioral expectations about an army's performance.

Drawing on new data and cases, the book demonstrates how rising levels of inequality are associated with decreased performance across a wide set of wartime behaviors. As inequality increases, so too does the risk of mass desertion and defection, lopsided casualties, and the use of coercion to force one's soldiers to fight. The poison of inequality wends its way through armies via three mechanisms. Greater exposure to state-orchestrated discrimination or (especially) violence creates grievances among victimized soldiers that saps combat motivation. Interethnic trust also craters within the army, reducing its ability to execute coordinated actions while stifling the flow of information, lowering reaction times. And prior exposure to state harm strengthens intra-ethnic ties, reinforcing group solidarity and facilitating collective action against military authorities. Though not blind to these issues, military commanders often seek refuge in costly monitoring and sanctioning efforts that typically backfire, compounding the army's self-imposed inefficiencies. In the end, armies are political constructs, not hyper-efficient machines, and are often divided by design, sealing their battlefield fate well before the shooting begins.

Divided Armies makes four contributions to our study of war. Conceptually, its view of battlefield performance incorporates a range of wartime behavior, including desertion, defection, and fratricidal violence, that is usually ignored by existing studies. Theoretically, it stakes a claim for inequality as an important structural variable that can account for battlefield performance across and within armies. Empirically, it introduces a new dataset, Project Mars, that includes dozens of new wars and belligerents previously excluded from our studies, pushing us away from the same tired set of Western cases and toward a more global military history.⁶⁵ Methodologically, the book embraces a mixed-method framework that uses matching for case selection, helping identify more representative belligerents and wars for close-range process-tracing.

⁶⁴ My thanks also to Laia Balcells and Tanisha Fazal for participating in the virtual APSA 2020 author-meets-critics panel that launched this roundtable.

⁶⁵ The Project Mars dataset, replication materials, and supporting documentation can be found at: <https://doi.org/10.7910/DVN/DUO7IE>.

There's no doubt that the book is ambitious and, at times, provocative. To open the conversation with my critics, I've bundled their concerns and questions into four issue areas in the passages below. I conclude with some brief thoughts about future research avenues and policy implications.

Inequality as an independent variable

Given its absence in prior studies, it is perhaps inevitable that questions will cluster around military inequality as an independent variable. Downes, for example, notes that I spend little time on the origins of inequality and largely sidestep the question of why leaders embrace exclusionary practices. This was decision born partly of pragmatism. Chronicling the ebb and flow of inequality across hundreds of belligerents would be a massive, if amazing, endeavor, one which is worthy of its own book. I felt that my analysis was on safe theoretical and empirical ground with the assumption that status hierarchies and inequalities are durable features of human societies and that they likely persist within militaries.⁶⁶ Moreover, I believe that societal inequalities are created and maintained through multiple channels. Leaders might impose their own biases on their societies at formative moments. Societies might evolve gradually toward more egalitarian norms or swiftly fall prey to rising populism and nationalism. Wars won and lost might shape conversations about who belongs; so too might tides of immigration. What matters is less the specific content of a belligerent's inequality than the structural nature of its persistence within and across belligerents over time.

For these reasons, I disagree with Downes's contention that military inequality is merely a reflection of ethnically-motivated challenges to political leaders. If true, these threats, not inequality, would be the real causal factor shaping battlefield performance. To be sure, some decisions to marginalize or repress non-coethnics are driven by security concerns (real or imagined). Yet it is often the state's imposition of a racial or ethnic hierarchy that actually creates these threats to the regime. Political leaders have also turned to power-sharing, not exclusion, to defuse threats. And, in other cases, rulers have excluded their own coethnics, not rival ones, to consolidate power.⁶⁷ The status of ethnic groups can also shift over time while their demographic weight remains static. In short, while some inequalities might stem from perceived threats from non-coethnics, it is just as likely that the imposition of inequality from above creates threats from below.

Taking a slightly different tack, Horowitz asks whether military inequality is actually regime type in disguise. Briefly put, no. Democracies are compatible with military inequality; we need look no further than the United States, which has fought the majority of its wars with segregated forces. Autocratic armies also vary widely in their levels of inequality. Using Project Mars data, we find only a weak correlation between military inequality and Polity2 measures of regime type (-0.22).⁶⁸ More generally, there is a fascinating dialogue now unfolding in comparative politics over whether political institutions create or

⁶⁶ See, for example, Kingsley Davis and Wilbert Moore, "Some Principles of Stratification." *American Sociological Review* 10:2 (1944):242-49 and Walter Scheidel, *The Great Leveler: Violence and the History of Inequality from the Stone Age to the Twenty-First Century* (Princeton: Princeton University Press, 2017). There is a small but important literature on "ethnic" armies in political science and sociology. See Ibn Khaldun, *The Muqaddimah: An Introduction to History* (Princeton: Princeton University Press, 1980); Stanislav Andreski, *Military Organization and Society* (Berkeley: University of California Press, 1954); Cynthia Enloe, *Ethnic Soldiers: State Security in Divided Societies* (New York: Penguin Books, 1980); and Donald Horowitz, *Ethnic Groups in Conflict* (Berkeley: University of California Press, 1985). The debt to Stephen Rosen's notions of "social divisiveness" should also be clear. See Stephen Rosen, "Military Effectiveness: Why Society Matters." *International Security* 19 (1995): 5-31 and Stephen Rosen, *Societies and Military Power: India and its Armies* (Ithaca: Cornell University Press, 1996).

⁶⁷ Philip Roessler, *Ethnic Politics and State Power in Africa: The Logic of the Coup-Civil War Trap* (Cambridge: Cambridge University Press, 2016).

⁶⁸ Consolidated democracies (≥ 7 Polity score) do have lower average MIC values than non-democracies, though even here the differences are surprisingly modest.

reflect societal inequalities.⁶⁹ In one reading of this evidence, it is plausible that our current measures of political institutions in the military effectiveness literature are masking the causal weight of inequality, not the other way around.

Downes also asks for greater detail on why leaders sometimes exclude certain ethnic groups entirely from military service. This does indeed happen; the Confederate Army, though desperate for soldiers, refused to enlist black soldiers (emancipated or not) within the ranks. Certain ethnic groups might also be safely excluded if they represent a small share of the overall population. Short wars might also have different mobilizational dynamics than total wars; longer wars would allow states to tap into previously excluded manpower reserves to fill out their ranks. Yet while some groups are excluded, the broader point remains: monoethnic armies are rare. Since 1800, belligerents have fielded armies with an average of five ethnic groups. This is a conservative estimate that does not reflect new soldiers who were incorporated once the war began (i.e. from defeated foes or neighboring allies) and omits ethnic groups with less than one percent share of the army's forces.⁷⁰

The explanatory power of inequality hinges partly on the claim that these prewar decisions about an army's composition are difficult to revise or reverse once war begins. Leaders, in other words, face a commitment problem. Why would non-core groups believe a leader's promises of better future treatment given a prior record of discrimination, if not violence? As Downes notes, though, we do observe variation in the historical cases of (some) non-core groups who proved willing to answer the regime's recruitment calls.⁷¹

So what explains this variance? Two motives are likely present. First, we do see some genuine enthusiasm for military service in the lower-inequality states (for example, Ethiopia), where soldiers might credibly believe that promised reforms will materialize if they fight hard. This is especially the case if the war is seen as opening pathways to (group) advancement that have been blocked by the existing political system.⁷² Such sentiment is generally absent in the higher inequality states outside of small, carefully-selected, token minority units. Second, in these high inequality states, service is often not a matter of choice but of coercion. Central Asians in the Red Army, to answer Downes, were subject to mandatory conscription and faced brutal wartime pressures, including fratricidal violence, if they failed to fight. Yes, Soviet leader Joseph Stalin's

⁶⁹ Rachel Kleinfeld, *A Savage Order: How the World's Deadliest Countries Can Forge a Path to Security* (New York: Pantheon, 2018); Ben Ansell and David Samuels, *Inequality and Democratization: An Elite Competition Approach* (Cambridge: Cambridge University Press, 2015); Carles Boix, *Political Order and Inequality: Their Foundations and Consequences for Human Welfare* (Cambridge: Cambridge University Press, 2015); Lars-Erik Cederman, Kristian Gleditsch and Halvard Buhaug, *Inequality, Grievances, and Civil War* (Cambridge: Cambridge University Press, 2013); Thomas Piketty, *Capital in the Twenty-First Century* (Cambridge: Belknap Press, 2013); Darren Acemoglu and James Robinson. 2012. *Why Nations Fail: The Origins of Power, Prosperity, and Poverty* (New York: Crown Business, 2012); Darren Acemoglu and James Robinson, *Economic Origins of Dictatorship and Democracy* (Cambridge: Cambridge University Press, 2006).

⁷⁰ Note, too, a possible selection effect at work. If leaders preemptively remove the most unreliable ethnic groups from their armies, then we are *underestimating* inequality's negative effects on battlefield performance.

⁷¹ For example, Some Oromo, Afar, and Somali citizens joined the Ethiopian military once prewar recruitment quotas were lifted.

⁷² The desire for recognition and equal citizenship is a clear theme in leading accounts of African-American soldiers fighting in the U.S. Army during World War I, World War II, and the Korean War. See, for example, Le'Trice Donaldson, *Duty beyond the Battlefield: African American Soldiers Fight for Racial Uplift, Citizenship, and Manhood, 1870-1920* (Carbondale, IL: Southern Illinois University Press, 2020); Brian Taylor, *Fighting for Citizenship: Black Northerners and the Debate over Military Service in the Civil War*. Chapel Hill, NC: The University of North Carolina Press, 2020); Christopher Parker, *Fighting for Democracy: Black Veterans and the Struggle Against White Supremacy in the Postwar South* (Princeton: Princeton University Press, 2009).

softened rhetoric helped on the margins, but we must be careful about viewing the presence of these groups as evidence that they always had meaningful choice about the matter.⁷³

These conceptual issues also raise additional questions about the measurement of military inequality. As Harkness notes, it was a formidable challenge to gather the data necessary to code each belligerent army's prewar military inequality. I agree with many of the points she raises. I do believe that (state) violence can shift identities; that identities change over time; and that the state itself plays a role in determining the salience of ethnic identities. As a point of clarification, Project Mars did *not* rely on the All Minorities at Risk (AMAR) dataset, for precisely the reasons that she notes: its conception of identity is largely ahistorical and fixed. Instead, we consulted AMAR's list of ethnic groups, along with many additional sources (153-57, Appendix, 14-16), to cast our net widely for which ethnic groups were present in a given society and army. But we reset each measurement of an army's demographics to the immediate prewar period to allow for changes in the composition of its recruitment base. A great book could be written about how war outcomes and violence reshape identities and inequalities, but that is the opposite of the intent of *Divided Armies*.

Finally, I totally agree with Zhukov that ethnic and racial discrimination casts a long shadow. I view my measure of MIC as capturing the direct effect of inequality. Taken on the eve of war (or five years before, depending on the exact measure), my view of MIC does not require assumptions about intergenerational transfers of collective trauma and memory. Instead, it assumes that soldiers (or their immediate families) have direct exposure to state victimization when they enter military service. But Zhukov is right: this is a conservative estimate that might truncate our estimates of inequality's negative effects by overlooking the indirect, long-term, transmission of state harm. The persistence of historical trauma via informal channels and family networks might also help explain variation in battlefield performance among now-inclusive belligerents with a history of ethnic discrimination or worse (see below).⁷⁴ Caution is warranted, though, since disentangling the direct and indirect effects of inequality becomes harder as time elapses between state policies and wartime conduct. This is especially true for belligerents that fought multiple wars over decades and centuries; inequality might become both independent and dependent variables in such analyses.

Battlefield performance as a dependent variable

Divided Armies makes the case for an expanded definition of battlefield performance, one that combines measures of combat power and cohesion in a single framework. Horowitz questions this decision by waving the dreaded red flag of endogeneity. Surely, he contends, wartime behaviors like desertion and defection influence a belligerent's loss-exchange ratio. But that is exactly the reason why I chose to use an index: to acknowledge that cohesion and combat power clearly interact on the battlefield. Yes, desertion and defection, not to mention shooting one's own soldiers, will influence relative casualties. But a belligerent's own losses also affect, if only partially, soldier decisions about side-switching or abandoning the fight entirely.

While each measure can be used in a stand-alone fashion, the combined index has the virtue of bundling together wartime behaviors that we acknowledge are at least partially correlated.⁷⁵ It also raises the inferential bar: the proposed explanation needs to explain both discrete wartime behaviors and their joint occurrence. I would note, too, that it is high-time we retired loss-exchange ratios as the sole, or most important, measure of military effectiveness. Setting aside data quality issues, casualty counts mean different things for different governments, with an important subset (especially high inequality belligerents) clearly not invested in the lives of their soldiers or scoring high marks for efficiency.

⁷³ Volunteerism from marginalized or repressed groups would be most surprising when (1) state harm is recent; (2) battlefield coercion is absent; (3) channels of influence exist outside military service; and (4) proffered concessions are modest.

⁷⁴ Anthony Marx, *Faith in Nation: Exclusionary Origins of Nationalism* (Oxford: Oxford University Press, 2003).

⁷⁵ Paul Rosenbaum, *Design of Observational Studies* (New York: Springer, 2010).

Horowitz raises a second concern about endogeneity: soldier morale might potentially be part of the independent and dependent variable. To be clear, a belligerent's prewar military inequality, not morale, is the independent variable. The dependent variables capture wartime behavior, not the attitudes of the soldiers themselves. Instead, morale, as expressed in soldier grievances and attitudes, is a mechanism that links prewar treatment of an ethnic group to its expected wartime behavior (see especially 56). Yes, morale has both attitudinal and behavioral components, but neither are captured in the independent variable, and the book's theoretical framework (outlined in Figure 2.1, 41) views soldier attitudes as a causal mechanism, not an outcome that is captured by the book's measures of battlefield performance.

Battlefield performance cannot be captured solely by quantitative measures, however. I therefore draw on the battle-level evidence from all eight historical cases to offer qualitative assessments of each army's tactical and operational sophistication. Across these cases, we witness how inequality tears pages from the commander's playbook by removing tactics and operational art that demand high levels of trust and decentralized decision-making from non-core soldiers. Horowitz suggests that this view "bakes in" sophistication as a linear process; the higher the sophistication, the greater the odds of battlefield victory. This is in some sense true; greater sophistication might result in greater lethality, for example. But here my focus is really on the fit between battlefield problems and their proposed solutions. Since divided armies have a reduced toolkit, they will struggle to match solution to problem. Sophistication is thus an expression of the range of options available to a commander, not a belief that sophistication will always produce better battlefield outcomes.

The Argument

Undaunted by the book's (long) discussion of how inequality affects battlefield performance, these scholars pose additional questions about my argument. Does inequality become irrelevant, as Harkness and Zhukov both suggest, when belligerents with the same level of inequality fight? Claims of indeterminacy often plague structural arguments: regime type, army size, wealth, military technology, and population, to name just a few structural variables, have all been criticized for leaving us without clear predictions about outcomes when these properties are similar across belligerents.

Should we be concerned that inequality falls into a similar trap? In my view, no, for several reasons. It turns out that wars with two belligerents (or coalitions) at the same level of military inequality are quite rare. Of the 322 campaigns tracked by the Project Mars data, less than one-third (108/322) featured warring states or coalitions with identical bands of military inequality.⁷⁶ In these cases, a researcher could simply revert to the actual MIC value for each belligerent, not the band level, to measure differences. And, in the unlikely event that these MIC values were also identical, the actual MIC values of the units engaged in battle could be generated to render point predictions about specific formations.

Moreover, I wholeheartedly agree with Downes's crucial observation that we have overlooked inequality because the cases that dominate the military effectiveness literature often pit low inequality belligerents against one another. In these knife-edge cases, other factors, including material power, technology, even dumb luck, can appear to play an outsized role. Once relative inequality has shrunk to a handful of points between belligerents, it is fair to conclude that it is not sufficient for battlefield victory. Even among wars with low inequality belligerents, however, I would argue that internal schisms are still important. The absence of state-sanctioned discrimination does not mean that political leaders have found a positive message to exhort their soldiers to new heights of performance, for example.⁷⁷

⁷⁶ A reminder that *Military Bands* assigns belligerents to one of four "bands" based on their MIC value. These bands are Low (0-0.20), Medium (0.21-0.40), High (0.41-0.60), and Extreme >0.61).

⁷⁷ New states might struggle to articulate a unifying political vision than a more consolidated state, for example. See Andreas Wimmer, *Nation Building: Why Some Countries Come Together While Others Fall Apart* (Princeton: Princeton University Press, 2018).

Informal hierarchies and norms of exclusion can also drag down military performance. Other forms of inequality, including ideological, class-based, or sectarian, might also be present but are not captured by my notion of MIC.⁷⁸ Before turning to more familiar variables, we should first search for hidden inequalities and hierarchies that can persist even within seemingly low-inequality belligerents.

Perhaps what matters most, however, is the imbalance in military inequality across belligerents. Indeed, both Zhukov and Harkness advocate for a dyadic account of how inequality shapes relative battlefield performance. This is a terrific suggestion. I chose to focus on a belligerent's own (absolute) inequalities in part to demonstrate how prewar choices, rather than wartime pressures, helped condition and constrain battlefield performance. Doing so also illustrates how strategic interaction between groups in one's own army is equally as important as combat between armies, a point I worried would be lost if I jumped immediately to a dyadic argument.

But it's possible to generate dyadic hypotheses with relative ease. Most simply, the greater the difference in military inequality between two sides, the worse the expected battlefield performance of the side with the higher MIC value. This difference could manifest itself in several ways: the relatively faster appearance of mass desertion and defection from the ranks; the greater the relative loss of personnel to defection and desertion; a quicker resort to fratricidal violence and the deployment of blocking detachments; and the shorter the war duration, as severe imbalances in MIC values lead high inequality belligerents or coalitions to fold quickly. Similarly, high inequality belligerents who face low inequality ones might be especially prone to suffering defeat as well as violent regime exit.

To satisfy my own curiosity, I revisited the book's case studies to see whether the belligerent (or coalition) with the higher MIC value did indeed perform worse. In all but one instance, the pattern holds: the higher the relative MIC value, the worse the relative score on the battlefield performance index. Only Morocco (0.01) and Spain (0.03) buck the overall trend. Here, the inequality difference between the two sides is vanishingly small, creating an opportunity for superior Spanish technology to have inflicted relatively higher casualties on Moroccan forces. Losses aside, both sides fought well, with no desertion, defection, or blocking detachments, as expected by their low levels of inequality. A closer examination must await the recoding of Project Mars to handle dyadic MIC values for belligerents and coalitions. But this initial look is reassuring, suggesting that absolute MIC values help explain individual belligerent performance while relative levels of inequality capture important aspects of relative battlefield performance.

Research design

Several questions were also raised about the book's nested research design. Horowitz, for example, criticizes the decision to divide my quantitative tests by time period (1800-1917, 1918-2011) rather than testing the argument on the entire pooled Project Mars dataset. It is true that most quantitative studies, especially those relying on the Correlates of War (COW), simply estimate their models in a single-shot across the entire time period. But I wanted a set of empirical tests that accurately reflected the qualitative changes in the lethality and mobility of combined arms operations that occurred in the post-1917 era.⁷⁹

⁷⁸ This was the case in Interwar France. See Jasen Castillo, *Endurance and War: The National Sources of Military Cohesion* (Palo Alto: Stanford University Press, 2014); Elizabeth Kier, *Imaging War: French and British Military Doctrine between the Wars* (Princeton: Princeton University Press, 1997).

⁷⁹ On this periodization, see Jonathan Bailey, *The First World War and the Birth of the Modern Style of Warfare* (Camberly: British Army Strategic and Combat Studies Institute, 1996); Stephen Biddle, *Military Power: Explaining Victory and Defeat in Modern Battle* (Princeton: Princeton University Press, 2004); and Jason Lyall and Isaiah Wilson, "Rage Against the Machines: Explaining Outcomes in Counterinsurgency War," *International Organization* 63 (2009): 67-106.

Confidence in my findings about inequality are only increased if they hold across both eras. I do, however, replicate this analysis with the full Project Mars dataset in the Appendix (Tables A11-A14), where the book's findings continue to hold. Similarly, inequality's negative effects on battlefield performance persist even if we restrict the sample to COW-only wars and belligerents.⁸⁰

Turning to the historical cases, both Harkness and Zhukov suggest additional covariates that could be added to the matched analysis.⁸¹ Given their shared interest in a dyadic theory, Harkness and Zhukov call specifically for the inclusion of the military inequality coefficient value for each belligerent's enemy to help control for cross-case differences. This is a great suggestion. *Casefinder*, the program used to assemble these paired comparisons, could be extended to include this variable, tightening the comparisons even further by ensuring that our selected belligerents are facing enemies with comparable MIC values.

But how concerned should we be that the playing field isn't level for our selected belligerents in these comparisons? Harkness, for example, suggests that the Mahdi's startling reversal in battlefield fortunes is due less to its own skyrocketing inequality and more to shift from fighting mostly Egyptian forces (1885-89) to mostly British ones (1896-99). As Harkness notes, even the inclusive Mahdi army stumbled against British forces in the latter stages of the first war, recording unfavorable casualty rates that presaged the drubbing it would experience during the second war.

Two points are worth noting, however. First, once we lift our gaze from loss-exchange ratios, we note that the Mahdi army's first incarnation did not suffer the mass desertion, defection, or fratricidal violence of its bitterly divided successor even when staring down British forces. Nor was its loss-exchange ratio anywhere near as dour. Second, a closer look at the actual MIC values for each pair helps clarify how the Mahdiya's own surging inequality provides the bulk of the explanatory leverage here. The Mahdiya's enemies had MIC values of 0.338 and 0.219 in the first and second wars, respectively, a modest 0.12 difference. Yet the Mahdiya itself swung from 0.01 to nearly 0.68, an enormous shift. While there is a massive 0.78 difference in aggregate MIC values between all belligerents in these two wars, the Mahdiya's slide into the depths of inequality accounts for nearly all of it.⁸²

Second, a key feature of the empirical strategy has received little attention so far: the control observations were chosen randomly. As a result, the design helps shield against imbalances in unobserved covariates, including the nature of each belligerent's enemy. Let's return to the Project Mars data for a moment. When we compare enemies across the paired belligerents, we find very little difference in their MIC values. Morocco's and Kokand's enemies — Spain and Russia, respectively — are only separated by a 0.07 difference in military inequality. That difference shrinks to a tiny 0.02 when we compare the Ottoman Empire's foe (Italy) with that of the Austro-Hungarian Empire (Russia). Even the largest difference (0.185) found between Ethiopia's and the Democratic Republic of Congo's enemies is modest. And there's no difference in MIC values for the German formations that our four Soviet Rifle Divisions fought. In short, while the enemy's MIC values should have been explicitly matched on, the research design as a whole worked to minimize these cross-case differences to a remarkable degree.

⁸⁰ A full discussion of the differences between Project Mars and COW is provided in the book's online Appendix. Project Mars was designed to be backwards-compatible with COW and retains its country codes for belligerents. Jason Lyall, *Sources, Data, Methods: The Project Mars Codebook* Version 1.0 (2020), 4-8, 33-34.

⁸¹ Harkness suggests adding measures for coup-proofing, distance to the battlefield, and population. We are currently collecting data on civil-military relations, including coup-proofing and purges. The statistical models and matching already include measures for civil-military relations and for distance to first battle, while population is used in the robustness checks.

⁸² During the first war, the Mahdiya had a score of 0.012; its enemies had a combined 0.3375 value, for a 0.32 difference. During the second war, the Mahdiya had scored 0.675; its opponent, a 0.219, for a -0.46 difference. The combined difference between the cases is thus 0.78, of which the Mahdiya's own shift represented 0.66 (or 85%).

While *Casefinder* is flexible, we should also be wary of simply adding more variables to the formal matching process for selecting cases. In particular, we need to avoid accidentally introducing post-treatment bias into our analysis.⁸³ For example, Zhukov suggests that the matching should have explicitly included a measure for whether a belligerent was part of a coalition. This makes sense; coalitional status is likely a key determinant of battlefield performance. Yet if high military inequality is the reason a belligerent joins a coalition, then we risk inadvertently biasing (downward) our estimates of inequality's effects if we explicitly match on coalitional status. The war in the DRC, to use Zhukov's example, was multi-sided precisely because the DRC's own inequality was so high that the army essentially collapsed, forcing political leaders to seek allies among neighboring states and emergent militia. Adding measures for the war's intensity (e.g. casualty counts) or its high stakes nature also risk post-treatment bias if these factors are correlated with inequality.⁸⁴

Future directions

Taken together, the various strands of this discussion neatly illustrate the wide range of new research questions that arise when inequality is centered in our studies of war and political violence. To be honest, I would judge *Divided Armies* a success if it helped kick-start a conversation about different types of inequality — including class, gender, and ideology, alongside ethnicity and race — shape how states, insurgents, and auxiliary forces all generate and field military power. We lack studies, for example, of how different social and economic inequalities combine to affect combat power and cohesion.

Inequality also likely affects how battlefield coalitions form, endure, and collapse, especially if alliances are struck not for power maximization reasons but instead to manage internal divisions. How inequality affects leader survival and war outcomes also remains unexplored. Scholars who privilege the role of emerging technologies like artificial intelligence and drones might also consider how inequality conditions both the uptake and the effects of such systems. Finally, the causal mechanisms that translate inequality into battlefield performance can be investigated using survey and behavioral experiments. These methods promise new insights into long-standing questions of combat motivation, unit cohesion, and intergroup dynamics within the military.

While motivated by the spirit of bench science, *Divided Armies* also suggests two broad sets of applied lessons for policy-makers. First, net assessment of military power should include measures of inequality within an adversary's society and military. Unlike military capabilities, these schisms are often public and highly visible, can be collected and validated across multiple types of data (including social media), and suggest actionable programs for exploiting these internal divisions in wartime. Second, diversity alone cannot guarantee battlefield victory. Without meaningful inclusion, the advantages of diversity are squandered. Armies must therefore work to include diverse voices in their decision-making. This will include often-difficult changes to existing recruitment, promotion, and retention policies at both the junior and senior ranks. Prejudice reduction programs might also prove useful in fostering more inclusive command climates. In short, racism is a national security threat, one that jeopardizes the survival of armies on the increasingly deadly battlefields of tomorrow.

⁸³ Avidit Acharya, Matthew Blackwell and Maya Sen. 2016. "Explaining Causal Findings without Bias: Detecting and Assessing Direct Effects." *American Political Science Review* 110:3 (2016): 512-529.

⁸⁴ For this reason, I excluded war duration from the statistical models and the formal matching procedure in the case studies.