Introduction: Why War?

War and the preparations for it are among the costliest of human endeavors. About 40 million people died in World War I. Not to be outdone, over 70 million died during World War II, or about 3 percent of the world population at the time. The war was especially harsh on the Soviet Union, with 26 million dead, or about 14 percent of its population. In 2018, the world spent $1.8 trillion on military expenditures to prepare for war, with the United States alone accounting for $649 billion of that figure (Tian et al. 2019). Civil war combatants may suffer a worse fate, with incomes typically dropping by about 15 percent by the end of a conflict (Collier et al. 2003: 2). Whether to fight wars dominates both elite political conversations and news headlines. Understanding it is therefore a primary research task for political scientists.

Broadly, we can classify explanations for both interstate and intrastate war into three groups. First, war may occur due to some sort of irrational process. Brilliant, clear, level-headed thinking leaders would not fight these wars. But perhaps some leaders face cognitive limitations and cannot interpret all the information and incentives in front of them in a coherent manner. As such, they fight.¹

Second, war may occur due to frictions between leaders and their citizens. Under this framework, citizens may want their leaders to fight wars when it will benefit the state overall. Leaders share that desire to improve the state’s outlook. However, leaders also have their own incentives to stay in office and extract the rents and ego benefits that come from that (Bueno de Mesquita et al. 2005). Moreover, they may have access to the spoils of war that the citizens do not. As such, wars could start because leaders wish to exploit that asymmetry or dupe their citizens into keeping them in office for longer than they should be.²

Finally, war may occur in the absence of those issues. These are cases where the states on both sides have intelligent leaders that act in good faith toward their citizens. Nevertheless, they cannot broker an agreement. On one hand, it may seem obvious that wars can happen with these groups. For example, it may be that both sides could calculate that their expected outcome of a war is better than maintaining the peace and fight accordingly. On the other hand, it might not be obvious that such explanations exist.

¹ Of course, peace could result from irrational behavior, so it is unclear whether irrationality is an overall force for war.
² Of course, leaders may also have private incentives for peace (Chiozza and Goemans 2011).
After all, war is a costly and risky affair. Those costs and risks would seem to incentivize leaders to broker an efficient agreement instead.

Neither of these perspectives is correct. The costs of war indeed encourage states to negotiate settlements. In fact, as long as war is costly, we will later see that settlements exist in principle that leave both parties better off. However, agreements in principle may not work in practice. Unless states recognize those settlements, agree to them, and then follow through on them, they may end up at war. These pathways to bargaining breakdown are not as straightforward as they may appear. As such, this book focuses on this third class of explanations.

To do so, we develop formal models of crisis bargaining. Many have discussed the benefits of formalizing theory, but the short version is that it ensures the logical consistency of our argumentation. In essence, formal models provide “accounting standards” (Powell 1999: 29–33).\(^3\) States live in a strategic and interdependent world. What one does affects the other, and vice versa. The logic of strategic interdependence quickly grows complicated, especially when we add many additional moving parts, long-term interactions, and asymmetric information. Game theory was developed to assist in structuring logical thinking about these situations, and so it is natural to apply it here.

Formalization also provides ancillary benefits. Having explicit assumptions can highlight strategic trade-offs that might not be obvious before writing down a model. The evolution of preventive war theory provides a good example of this. The original preventive war models helped explain that it is not the fear of future war that incentivizes proactive strikes but rather concerns about how disadvantageous the future peace will be. In the process, the modeling enterprise demonstrated the importance of the assumption that states cannot constrain how power shifts. From there, the literature has tried to explain why preventive wars happen over controllable armaments. Meanwhile, formalization can help with theory destruction as well. A common research structure within international relations is the development of some causal mechanism and then an empirical test that supports the mechanism’s expectations. If formalization can show that the theory is incoherent, then we must rethink what underlies the empirical patterns observed.

Like all theorizing, formal models require some abstraction. This is not just a point about simplification, which we will return to in a moment. Rather, the models we develop do not involve conflicts between France and Germany, the United States and the United Kingdom, North Korea and South Korea, Russia and Ukraine, or any of history’s great rivalries. Rather, the states we will analyze are named A and B. Those states do not even have something in particular at stake, like territory, natural resources, foreign support of a rebel group, or influence within a buffer state’s domestic politics. Instead, they will argue over some object with a standardized value of 1.

\(^3\) Readers looking for a longer explanation should consult Wagner (2010).
1.1 The Rationality Assumption

Deeper reading of history is important for empirical evaluation of theory and development of richer models, but this type of abstraction is a good thing. It is a requirement for generalizable social science. The circumstances that caused war between, say, Vietnam and China will never be exactly duplicated. Thus, if we want to apply anything we learn from specific examples, it must be the abstract portions of those cases. Calling the actors A and B is a nod in that direction.

Meanwhile, describing the good as something valued simply at 1 has an additional benefit. The types of things states and civil war combatants fight over are numerous. In contrast, the mechanisms that cause bargaining breakdown are not. Indeed, we will observe just two state-level explanations in this book. If the long-term goal of researching war is to stop it from happening, then we can focus our efforts on those two mechanisms.

But before developing those points, it is first worth providing more detail on why this type of explanation has dominated existing work on war.

The models we investigate assume that the states are rational. Rationality has a number of definitions, so will be helpful to clarify what we mean here. As a starting point, rational actors can rank all possible outcomes from best to worst. They can also report the relative intensity of those preferences and weigh them consistently when outcomes arise probabilistically. However, as is standard for game-theoretic models, we go beyond that. The players can implement strategies to maximize their welfare given their strategic circumstances. They are also fully capable of processing information and updating their beliefs in a logically consistent manner.

These assumptions are all vital first steps toward any explanation for war. For those who want to explore rationalist explanations, the reason to impose them is straightforward. But even someone who wants to study irrational causes of war still must start here. After all, irrationality is defined as a lack of rationality. One cannot argue a conflict is irrational until we have a benchmark for when wars arise for rational reasons.

Still, if any sort of situation would be conducive to well-contemplated decision-making, it may be war. There are few, if any, strategic scenarios that have higher stakes. If a decision might cost someone $10, it would not be surprising if they did not spend the mental energy to work through

---

4 That is, they have complete and transitive preferences – all outcomes are comparable, and preferring \( A \) to \( B \) and \( B \) to \( C \) implies preferring \( A \) to \( C \). This ordering allows for indifference. Arrow (1950) presents an immediate problem with this assumption. However, virtually all crisis bargaining models feature a single-dimensional good at stake. Majority rule with single-peaked preferences creates an environment that avoids those problems.

5 That is, they follow all the general principles of expected utility theory.

6 That is, they use Bayes’ rule.
1 Introduction: Why War?

the incentives. In contrast, wars can cost trillions of dollars. This promotes a level-headed approach behind the scenes. Whether leaders display that level-headedness publicly is a different story. In fact, there are strategic reasons why leaders may wish to act like madmen (Acharya and Grillo 2015). But given all that, rationality seems like a reasonable default assumption.

Related to that, the evolutionary pressures that come with state-building and war also incentivize rational decision-making. Holding a state's decision-making apparatus constant, a state producing irrational behavior will find itself making more strategic blunders. Over time, again given the stakes, this would seem to remove irrationally behaving states from the system. Thus, irrational wars may be ephemeral, and rational explanations provide the best empirical leverage on average. Similar pressures exist within countries. It is difficult to become a country's chief executive if one consistently makes irrational mistakes. Some may get lucky and find themselves in a position of power, but they would have a hard time staying there.

Perhaps for these reasons, the literature gives a privileged position to these explanations (Fearon 1995b: 379–380).

1.2 The Unitary Actor Assumption

Speaking of abstraction, the other major running modeling choice we will make is to adopt the unitary actor assumption – that states are singular strategic entities, free from tensions of domestic politics. This is a theoretical tool researchers use to simplify much of the analysis. Employing it means that the states have well-defined preferences. There are no domestic political cleavages that a leader can use war to exploit.

Is this a realistic assumption? Absolutely not. Theoretical work on veto players shows the importance of domestic politics (Putnam 1988) and predates modern models of crisis bargaining. More recent work has brought domestic politics into mainstream security theory (Goemans 2000; Chiozza and Goemans 2004; Bueno de Mesquita et al. 2005). Modeling unitary actors means that by assumption we rule out wars due to principal-agent problems – frictions that arise when one actor (here, the leader) must take actions on behalf of another actor with a different set of preferences (here, the broader public).

Realism aside, is it a useful assumption? Yes, for four reasons. First, it is convenient. The fewer moving parts a model has, the easier it is to solve. Treating a state as a single entity rather than an amalgamation of many different actors allows us to focus on the interstate bargaining and not exhaust endless effort on resolving the domestic issues.

Second, it makes the mechanisms we find more transparent (Paine and Tyson 2020). Imagine that war occurred in the equilibrium of a model with
1.2 The Unitary Actor Assumption

domestic politics. It might then be unclear whether the interstate strategic environment caused the conflict, or if domestic politics prevented the parties from reaching a settlement. In contrast, a model without domestic politics makes the answer clear. If war occurs, it must be due to the impossibility of the situation.7

Third, the unitary actor assumption matches how leaders describe their own motivations for war. When the United States destroyed Sudan’s Al-Shifa industrial facility in 1998, Bill Clinton did not justify the operation as a domestic political distraction from the ongoing Monica Lewinsky scandal. Instead, Clinton emphasized the facility’s connections to Al-Qaeda, which had bombed US embassies in Kenya and Tanzania thirteen days earlier. Congressional Republicans suggested otherwise. As the United States prepared for war against Iraq in 2003, George W. Bush argued that Saddam Hussein was the next major threat to the Western world. Dissenting Democrats instead grew concerned that Bush was trying to prolong the greater War on Terror to divert attention from his weaker domestic policies.

Finally, the mechanisms we explore also persist to varying degrees once domestic politics returns to the discussion (Slantchev 2011: 9–10). Regardless of whether a leader’s constituents are politically important, leaders still grapple with not knowing their opponents’ resolve, process new information coming from the battlefield, struggle to credibly promise to uphold agreements in the future, and so forth. For example, even if Bush had a personal benefit for removing Saddam, the dynamics of shifting power still concerned Americans more broadly. Of course, the domestic political cleavages altered some of the finer details with Iraq, as they would with other cases. Nevertheless, having a unitary actor model builds a useful baseline for comparison. When this book incorporates domestic politics, it is often to investigate those differences.

These examples underlie a broader theme. When leaders motivate war, they speak about the greater good and the national interest. They never cite their own personal motivations or self-interest. Building unitary actor models helps us adjudicate the competing narratives. If smart-thinking policymakers should always avoid a particular type of war, then diversionary war arguments would hold greater weight. But imagine that it is possible for rational leaders who only follow the public interest to find themselves in wars. Then pinning any one conflict on diversionary incentives becomes more difficult. Thus, our exploration is a test of whether the mechanisms leaders publicize have coherent theoretical foundations.

7 In the context of institutional design, Myerson (1999) makes a similar point. Suppose an institution fails to realize its goal. This could happen for two reasons. First, the institution itself may be incapable of fulfilling its goal. Second, the people who implement the institution may be fallible. Modeling a strategic environment in a vacuum—for our purposes, taking domestic politics out of the equation—generates a clear answer.
1 Introduction: Why War?

1.3 Model Simplicity

To reinforce a recurring point, another theme we will see throughout this book is the utility of using simple models. There are a number of good reasons to work with simple models. For starters, simple models allow an author to communicate theories to a broader audience. One reason that “Rationalist Explanations for War” has become the canonical paper within this literature is its accessibility, which has given it a broad audience. Second, the actors playing real-world games have cognitive limitations (Gintis 2009: xv–xvi). Under such circumstances, it is not clear that a technically more impressive model has better predictive leverage than a simpler one (Tarar 2017). Third, the fact that new results are emerging from simple models in the third decade of crisis bargaining literature suggests that we as researchers still have fundamental gaps in knowledge regarding the causes of war. The best way to get a handle on the situation is to understand simple mechanisms and delve into more complicated ones later.

Relatedly, and perhaps most importantly, simple models are better for understanding mechanisms. Models, by their very nature, are not real life. Only real life is real life. But real life is complex and therefore hard to examine and understand the causal linkages between primitives and outcomes. Rather than thinking about whether a model is “right” or not, we should instead evaluate whether a model is useful (Clarke and Primo 2012).

For the bargaining and war literature, the main purpose of modeling has been to understand the causes of conflict. As Paine and Tyson (2020) describe, using an experimental setting is helpful. Imagine that we write down a model of bargaining and conflict that captures many of the central incentives that warring factions face. After solving for the game, we find that peace always transpires in equilibrium. Now suppose we take that same game and add some extra component $x$ to it. After solving the game once again, we discover that war occurs. Then we know, at least in the stylized setting, that $x$ causes war.

The “stylized” caveat highlights the limitations to this approach, but it is important to think about the alternative. In this book, we will see that power shifts and uncertainty about the outcome of war are common features of militarized conflict. Undoubtedly, both problems exist in many historical cases. But adding both to a model at the same time creates a problem with drawing an inference. Suppose that the baseline model resulted in peace. Then we add shifting power and uncertainty to the model. In equilibrium, we observe war occur. Did shifting power cause the fighting? Or was it the uncertainty? Do both? Does one incentivize conflict while the other actively discourages it? The answer would not be obvious. But handling one extension at a time gives us direct answers.\[8\]

This is not to suggest that combining multiple mechanisms into one model is a bad idea. New theories can emerge from exactly this type of work (Wolford, Reiter, and Carrubba 2011; Tarar 2013). It is just that those theories need a basis of comparison first.
None of this is an indictment of complicated models. Undoubtedly, some of them contain fruitful results worth analyzing. And as later chapters will attest to, even some seemingly simple strategic environments require an unexpectedly heavy proof load. Rather, the goal for this book is to get a better working understanding of some core theoretical trade-offs.

### 1.4 Roadmap

This textbook has three parts. Chapter 2 stands on its own to present the central research question of the crisis bargaining literature. One might suspect that wars happen because there may be no agreement that both sides prefer to war. In other words, sometimes war is inevitable. Fearon (1995b) shows that this reasoning fails if states are unitary actors and war is costly. The chapter unravels this logic, leaving us to wonder why states do not reach a mutually agreeable settlement.

The next sections of the book provide some answers, divided into two central mechanisms. The first is the commitment problem. Alternatives more efficient than war may exist, but one or both parties may be unable to credibly commit to implementing them. Chapters 3 provides the introduction to this idea, with applications to preemptive war, fait accompli, bargaining over objects that influence future bargaining power, and stalling. Chapter 4 takes a deep dive into preventive war theory, a type of commitment problem that the literature has paid special attention to. Chapter 5 expands the horizons of commitment problems, applying it to war finance, compliance issues, and wars of economic independence. It also explains how issue indivisibility and risk acceptance – mechanisms that often stand on their own – are forms of commitment problems (Powell 2006: 176–180).

The final section is on uncertainty. Even if mutually preferable peaceful settlements exist, asymmetric information may prevent the states from reaching one of them. Chapters 6 and 7 introduce this idea, using costs and power as the canonical sources of uncertainty. Chapter 8 asks how states may try to learn about one another and looks into the effectiveness of cheap talk communication. Chapter 9 expands the options to costly signals. Chapter 10 highlights the importance of peace premiums – the difference between two types’ war payoffs – in avoiding war and the counterintuitive implications that follow from it. Chapter 11 introduces mechanism design as a way to explore general results in crisis bargaining games with incomplete information. Finally, Chapter 12 examines models that treat war as a learning process.
Suppose a disputed border sits between two countries. One day, prospectors discover vast oil deposits in the region. Suddenly, one country declares that the area – and thus the oil – is under its unalienable sovereign territory. Not to be outdone, the other country makes a similar public pronouncement. A serious diplomatic crisis ensues.

Military planners from both sides begin meetings with their respective heads of state. The first side’s planners believe that fighting a war to secure the region is likely to be successful. They estimate that their country will prevail in a war 70 percent of the time. Compensating for casualties and resources diverted from the economy, they expect the war to cost $100 billion. But the first side’s prospectors believe this is a small price to pay – they estimate $1 trillion in oil lies within the disputed region.

The other side’s planners agree to some extent. They also think that their opponent will win 70 percent of the time, leaving their own probability of victory at 30 percent. They also think that the war will be costly, but manageable – they expect to suffer $200 billion in damage by fighting. And their prospectors expect the region to produce $1 trillion of oil as well.

At first blush, this situation seems hopeless for diplomacy. Politicians in the first country can calculate their expected profit from fighting. With $1 trillion in potential revenue and a 70 percent chance of obtaining it, war will produce a gain of $700 billion in expectation. That far exceeds the $100 billion in costs. Meanwhile, the politicians in the second country also find fighting profitable, even if the margins are a bit tighter. $1 trillion in oil 30 percent of the time still means $300 billion in expectation, again above the $200 billion in costs.

Fortunately, on the eve of war, a cunning diplomat from the first country steps forward with a plan. “Let us create a new oil company,” the diplomat proposes. “My country will own 75 percent of the company’s stock. Our friends to the north will own 25 percent. The company will drill in the region and divide the profits between its shareholders.”

The diplomat’s leader does some quick calculations and approves of the plan. After factoring in the costs of war, the net profit of fighting is only $600 billion. The joint venture, in contrast, will bring in $750 billion. The opposing country’s leader also loves the plan. War for it will result in a net profit of $100 billion, far lower than the $250 billion that the joint venture produces. The leaders then sign the agreement and live happily ever after.

Although this story may seem abstract, it roughly matches what happened between Saudi Arabia and Kuwait in the twentieth century.
Following the fall of the Ottoman Empire, British diplomats hastily redrew the map along the Arabian Peninsula and established a neutral zone between those two countries. Few cared about the details at the time. But less than two decades later, the discovery of oil fields in the surrounding areas suggested that possession of the neutral zone could be valuable. Rather than fighting over it, though, the countries divided the oil between private companies and a joint venture. War over the region would have been expensive; negotiating a solution was rather cheap. More explicitly, they recognized that war outcomes are Pareto inefficient – alternative outcomes exist that make at least one party better off without making anyone worse off.¹

That framing of the problem leads us to this chapter’s main question: if war is inefficient, can states negotiate a solution that would leave them both better off?² The example suggests that such solutions may be possible. And if the answer is yes, it may reflect a broader incentive to secure the peace. But we should not make hasty generalizations. Perhaps a numerical quirk of the example makes peace work here, but the central lesson does not apply elsewhere. What we need to adjudicate the generality of the result is a formal description of the problem and a mathematical proof of its solution.

Using that pathway, this chapter generates a clear answer. It is also a foundational – if not the foundational – result within the conflict literature. Under loose conditions, settlements always exist that are mutually preferable to war. The logic is clear once one goes through the proof but is not obvious without the benefit of the modeling exercise. Fighting is costly and ultimately produces some sort of resolution for the parties. If the states agreed to that expected resolution ahead of time, then both save on the costs of war. Fighting cannot improve their outcome on average.

Of course, this finding produces a deeper problem for later chapters. Empirically, we observe wars occur. They do not happen regularly – most states are not fighting most other states most of the time – but they are not exceptional either. War’s inefficiency puzzle asks why states choose a costly means to resolve their differences when a more efficient solution is available. By the end of the chapter, we will see that “because no settlement will make them both happy” is not a good answer to the puzzle. Put differently, greed does not explain war.

We undertake a four-step process here. The next section begins the journey by offering a straightforward proof that mutually preferable settlements generally exist. The proof is a bit abstract, so the following section illustrates the logic using the geometric interpretation of bargaining ranges that Fearon (1995b) popularized. We then incorporate the central incentives of crisis bargaining into a game-theoretic model. This ultimatum game

¹ This is the notion of efficiency we will use throughout the remainder of the textbook.
² For most of this text, I will refer to the two actors as states. However, the logic of these models also applies to civil wars between governments and rebel groups. In fact, civil wars are a primary motivation for some of the models developed in later chapters (e.g., Sections 3.4, 5.1, 5.2, 5.3, 10.3, and 12.3).