

1

Introduction

The ideas of economists and political philosophers, both when they are right and when they are wrong, are more powerful than is commonly understood. Indeed the world is ruled by little else. Practical men, who believe themselves to be quite exempt from any intellectual influence, are usually the slaves of some defunct economist.

John M. Keynes (1936)

WHAT causes war? Why do states sometimes trade freely and other times protect their domestic industries? Why are some environmental treaties successful and others fail? Wars, international trade, and environmental treaties shape the lives of people around the world. Whether people live or die, are prosperous or poor, have a clean environment or a polluted one, all are affected by international relations. Mistaken beliefs about how the world works can lead to flawed policies, which can cause unnecessary harm to millions.

This book presents an approach to international relations that yields at least some tentative answers to questions such as these. The approach analyzes international relations through the lens of game theory, the mathematical study of strategic interaction. In this introduction, I discuss international relations theory, why game theory is useful for studying it, how the approach fits into the overall international relations theory landscape, and why an acquaintance with the approach may be of use even to those who do not pursue it in depth.

1.1 International relations theory

Why would we want a theoretical approach to international relations when an empirical one would seem more practical and useful? Theory helps guide our thoughts when we approach the world for empirical answers. If our theoretical ideas are confused or inconsistent, we are unlikely to find solid answers to questions we might pose about

2 INTRODUCTION

world events (Mearsheimer and Walt 2013). Theory helps us formulate models or mechanisms of international processes that we can then compare with reality to see if they seem to capture what is going on. For this reason, theory has been important to all scientific disciplines, from physics to economics, as well as in international relations. As Keynes famously argued, even those who consider themselves immune to abstract speculation are usually guided by unconsciously held theoretical perspectives, which appear to them merely as principles, or rules of thumb. We might as well get these perspectives as clear and coherent as possible.

Theories are deductive logical frameworks that imply potential laws, or more modestly, hypotheses, about how the world works (Waltz 1979, 5–7). As such, they must start with a set of core assumptions, or postulates. What counts as a core assumption and what as an auxiliary (or even unstated) assumption is somewhat arbitrary, but there is a fair amount of convergence in the rationalist international relations theory world about what the central ideas are.¹ I focus on three in particular.

1. *States* are the most important actors. State actions determine war and peace, and set the conditions under which economic activity takes place. Other actors are clearly important, but in the models of this book they will matter through their influence on state preferences or behavior. This makes the theory “state centric,” which has obvious limitations (Ashley 1984, 238–242). Some rationalist scholars interpret the actors in the models as state leaders rather than states per se, and hence adhere to an individualist ontology (Bueno de Mesquita *et al.* 2003). States are still primary, however, because the individuals that matter are the ones who occupy important decision-making roles in the state.
2. States interact in a context of *anarchy*. This implies that states may fight each other if they wish, but it also means that they cannot make *commitments* that they do not wish to keep in the future. In well-ordered states, the legal system enables private agents to make contracts that bind them to behave in certain ways, even when they would prefer not to. Between states, there is no such option – states will only agree to things that they want to do anyway, or can enforce upon each other through threatened punishments.
3. States are *rational*. What this means exactly will be spelled out in Chapter 2. For now, let it be understood as an assumption that states make decisions based on their evaluations of the consequences of their actions, rather than via an evaluation of the appropriateness of the behavior (March and Olsen 1998).

While the first two assumptions are fairly accurate, in my view, the third one is clearly false, as a descriptive matter. However, it may be close enough for certain contexts, and

¹ For the realists, see Waltz (1979, Chapter 5), Mearsheimer (2001, 30), and Glaser (2010, 28); for neo-liberal institutionalism, see Keohane (1984); and for the constructivist approach, Wendt (1999, 1).

is useful to produce coherent theories. The status of the assumption is open to debate, in the social sciences in general as well as in international relations in particular, and the conditions under which it is more or less accurate are subject to investigation (Kahler 1998, MacDonald 2003, Camerer and Fehr 2006). I omit from the above list the common realist assumption that states are *unitary* actors. It will hold for the majority of the book, but will be relaxed in Chapter 11, where I examine the influence of domestic actors. I also omit from the above a specific assumption about the content of state preferences, or what it is that they seek to maximize. I will discuss this issue in the next chapter; the models in the book will accommodate multiple interpretations on this point.

Theories exist to explain some things, the dependent variables, with reference to other things, the independent variables. What does the theory of this book try to explain, and with reference to what? At the broadest level, game theory attempts to explain *strategic choice*, or *behavior*, with reference to *preferences* and *constraints*. This behavior could be the initiation of wars, setting tariff levels, forming alliances, joining multilateral treaties, accepting offers of compensation, etc. The rationality assumption implies that states are choosing the best option available to them, from a set of possible options, or strategies. Rational choice theorists are, therefore, interested in *optimization* or the mathematical problem of selecting the value of a variable (the strategy choice) that maximizes the value of a function (the actor's *utility* or happiness). All the actors in the models considered below will be attempting to maximize their utility, or make themselves as well off as they can, given the constraints posed by the structure of the game and the actions of the other players.

As a result of this commitment to the idea that individuals are maximizing their utility, rational choice theorists are especially puzzled by – and interested in explaining – *inefficiency* or outcomes that leave both sides worse off than they could have been. Inefficiency is viewed as unambiguously bad because at an inefficient outcome, everyone could be made better off, so no one would be harmed by a change. In international relations theory, this concern for efficiency often leads to a focus on *conflict* and *cooperation*. Conflict is assumed to be inefficient in comparison with cooperation, that is, more costly for both sides than alternative outcomes that are at least in theory available to the players in a strategic situation. Conflict, be it war, the interruption of trading relations, or economic sanctions, is assumed to impose costs on both sides. If the same outcome could be achieved without conflict, and so without the costs of conflict, both sides would seem to be better off. Note, this assumes that decision makers do not enjoy war or conflict for its own sake, which is probably true for most leaders in the current era, but it is clearly not always true. Cooperation is assumed to be efficient, in that it avoids the costs of conflict.

There are two principal contexts in international relations in which cooperation and conflict arise: *bargaining* and *implementation* or *enforcement* (Fearon 1998a). In bargaining, conflict can take the form of prolonged bargaining, rather than quick agreement, and negotiations that fail to reach an agreement at all. Negotiations that fail can impose costs to all, associated with the foregone economic benefits of cooperation. In the international security context, failed negotiations can even lead to war. In the implementation of an agreement, conflict can arise when one side fails to fulfill its obligations, leading the other side to reciprocate. If there are short-term incentives to exploit the other side, it may be hard to sustain cooperative behavior over time that honors an agreement. For instance, a national leader may be tempted to impose non-tariff barriers on foreign goods in violation of a trade treaty in order to please a domestic constituency, knowing that this will harm another state and possibly cause it to retaliate.

There are five main explanations of inefficiency in international relations. The first three derive from the bargaining theory of conflict as articulated by Fearon (1995), while the last two are associated with the literature on enforcement, usually known as cooperation theory (Oye 1986).

1. *Undervalued or non-feasible intermediate outcomes* means that if states simply do not value intermediate or compromise solutions very highly, or such solutions are not feasible or absent altogether, then they may fail to cooperate because each side prefers to take a chance on getting all or nothing, rather than settling for compromise. This is sometimes known as the “indivisible goods” issue, although the problem is more general, as we will see in the next chapter.
2. *Private information* means people operate with different beliefs, which may prevent them from coming to agreement in bargaining situations. If one state underestimates another’s resolve to prevail over a certain issue, it may be too intransigent in the bargaining process, which may then break down in conflict. Mistrust can also prevent cooperation in implementation settings if each side thinks the other side is motivated to exploit them rather than reciprocate the cooperation. If private information is combined with incentives to misrepresent knowledge, as it often is in bargaining and in the mistrust context, the parties will have difficulty overcoming their uncertainty and may fail to cooperate.²
3. *Changing power* can also cause bargaining to fail if a state fears it will lose power in the future and wishes to attack while it is strong rather than allow itself to decline and then have to make concessions from a position of weakness. This is the logic of preventive war. Changing power is often referred to as “the commitment problem”

² This corresponds to the problems of uncertainty over preferences and the state of the world, as discussed in Koremenos *et al.* (2001).

following Fearon (1995) and Powell (2006). I prefer to think of the inability to commit as a general implication of anarchy that underlies all the models considered in various ways.

4. *Monitoring problems* can also cause inefficiency if states cannot immediately detect efforts to exploit them and so fail to cooperate because it makes them vulnerable to exploitation. This is particularly problematic in implementation and enforcement questions and is what makes cooperation risky in the Prisoner's Dilemma game, analyzed in Chapters 3 and 8. If states were able to perfectly monitor each other's behavior, there would be no such thing as surprise, and in particular no surprise attacks or surprise defections. States would, therefore, be able to mirror each other's actions as closely as they like, and thereby eliminate the fear of being suddenly exploited. This problem is especially salient in the fields of peacekeeping and arms control verification (Bailey 1995, Lindley 2007, Fortna 2008, Debs and Monteiro 2014).³
5. *Impatience* can also cause inefficiency in the enforcement context. When there are short-term temptations to exploit the other side, states can still cooperate if they value future payoffs sufficiently and fear that exploiting the other side will lead to mutual defection in the future. This "shadow of the future" can keep states cooperating, but only if they care about the future (Oye 1986). If they are too focused on the short term, because of personal or institutional characteristics, they may not care about future punishment and seize the short-term gains from exploiting the other side.

These explanations of inefficiency may seem rather incomplete at present, but they will be developed at much greater length in the succeeding chapters, in a variety of settings, and will hopefully become clear.

A special form of conflict is *war*. Empirical scholars define war as sustained combat involving at least 1,000 fatalities (Sarkees 2000). I assume that war, like conflict more generally, is inefficient because it imposes costs on both players, so that both participants could be made better off by something other than fighting (Fearon 1995). War can be divided into two types, *non-decisive* and *potentially decisive*. Non-decisive wars are competitions in the infliction and absorption of costs, and the key feature is that no side can lose except by voluntarily making concessions or giving up the object in contention. A state can always keep fighting if it wants to. Such wars are sometimes called wars of attrition, because they involve the competitive destruction of value.⁴ The same kind of analysis that applies to such wars can be applied to non-lethal contests of will,

³ This assumes that states cannot change their policies discontinuously and instantaneously. This problem is called uncertainty over behavior in Koremenos *et al.* (2001).

⁴ Real wars of attrition can be potentially decisive; however, if each side has a stock of resources to fight with, once exhausted this leaves them incapable of further fighting.

6 INTRODUCTION

such as trade wars or lengthy bargaining more generally. A key question, however, is why states in a bargaining session would choose to impose on each other the radically greater costs involved in a real war rather than continue to bargain.

Potentially decisive wars embody a mechanism that may eliminate one side from the game by disarming it. In a potentially decisive war, a state can lose even if it wishes to keep fighting, and its preferences will no longer matter because it loses the ability to affect the other side. As we will see in more detail in Chapter 4, the simplest model of a potentially decisive war is a costly lottery between victory and defeat. The costs reflect the losses involved in fighting, and the lottery reflects the fact that if a state's troops are victorious, the enemy can no longer offer resistance. Potentially decisive wars are even more puzzling events because not only do they greatly increase the costs of bargaining, they dramatically increase the variance, since they introduce the possibility of total loss and total victory, which is absent in non-decisive wars where states must lose voluntarily. Why states would choose a high cost, high variance method of resolving their disputes is a deep puzzle, since we usually assume that states dislike both cost and risk.⁵

1.2 Game theory and international relations theory

The particular theoretical approach of this book is based on formal, rational choice theory, in particular, game theory.⁶ Formal theory just means theory that is expressed in mathematical terms. The advantages of doing so are the increased logical rigor that results from harnessing the power of mathematical language. Mathematicians have established many helpful tools that can be applied to thinking about processes of all kinds, including international relations. Rational choice theory is based on a particular conception of the processes being modeled. The premise is that social processes such as international relations can be best thought of in terms of the choices made by actors that have goals and are trying to achieve them rationally. Game theory is the subset of rational choice theory that deals with strategic interaction, that is, situations in which what each player wants to do depends in part on what it thinks others will do.⁷

The assumptions of game theory line up particularly well with the assumptions of international relations theory. As a branch of rational choice theory, game theory

⁵ Non-decisive wars may become potentially decisive if the combatants escalate and start fighting potentially decisive battles. Most low intensity warfare is non-decisive.

⁶ The founding book is Von Neumann and Morgenstern (1944). A good text that could be read with this book is Osborne (2004). Gill (2006) and Moore and Siegel (2013) provide overviews of mathematics in political science.

⁷ For the debate over the merits of rational choice in security studies, see Brown *et al.* (2000b). For a survey of early game theoretic applications to international relations, see O'Neill (1994).

7 1.3 PARADIGM WARS AND PROBLEM ORIENTED RESEARCH

assumes that actors are rational and choose the strategy that will make them best off in terms of the consequences of their behavior. Game theory also assumes no special abilities to make commitments between the actors, and so is appropriate for studying anarchical contexts.⁸ Also, there are relatively few important actors in many international interactions. Many international events are bilateral, and the two national leaders are the most important decision makers. Models with two or three actors are much more tractable than models with more players, although large numbers of players can be studied in simplified settings where the strategies available are not too complex, as in the models of Chapter 10.

Game theory alone, pursued in the abstract, can only provide limited insights into any specific empirical domain such as international relations. The complexity of the world ensures that there is enormous variation in strategic contexts across different disciplines, such as economics, sociology, and political science, and within political science between domestic and international politics, and within international relations. To really generate useful ideas or testable hypotheses, we must build models with these specific contexts in mind. Models highlight mechanisms that are thought to be important in producing the results that we observe.

Models are usually more complicated than the games pure game theorists study, but of course are radically simpler than reality. The additional complexity, motivated by substantive knowledge of a particular domain, makes the model useful for generating insights for that domain, and possibly even testable hypotheses. The simplification is equally necessary, in order to abstract away from complications and study problems in their simplest form, before building up more complex models to deal with the possibly confounding details.

Formal models are mathematical, and we often need to solve equations or prove propositions to draw out the implications of a model. However, the key to insight in modeling is choosing what the structure of the model will be, which determines what equations are set up in the first place. What is the underlying structure in a given situation, what is going on? For this reason, this book develops international relations theory together with the formal models. The game theoretic tools provide a language for thinking clearly about international relations and developing a rich, complex theoretical framework for understanding.

1.3 Paradigm wars and problem oriented research

Theory has traditionally been conceived of in international relations in terms of “paradigms,” or schools of thought. E. H. Carr drew a distinction between realism

⁸ Technically, this is true of *non-cooperative* game theory. The alternative, *cooperative* game theory, does allow for commitments.

and idealism (Carr 1946), where realism was held to constitute a tradition of thought stretching back to Thucydides (1954), and including luminaries such as Machiavelli (2003) and Hobbes (1968 (1651)). Hans Morgenthau's *Politics among Nations* expounded the realist view at length in the early Cold War period (Morgenthau 1948). The most prominent late Cold War exponent was generally acknowledged to be Kenneth Waltz who was said to have turned realism into a proper social scientific theory (Waltz 1954, 1979). Waltz's theory was based on the assumptions that the world was anarchic and states wished to survive, and led to the conclusion that cooperation, except against common enemies, was rare, war frequent, and that states formed recurrent balances of power. After the Cold War, fissures developed within realism leading to a schism. Offensive realists, most prominently John Mearsheimer (2001), were most loyal to Waltz's assumptions and conclusions and they attempted to fill in the logic. Defensive realists, such as Charles Glaser (2010) and Stephen Van Evera (1999), held to the assumptions, but abandoned the conclusions when it became apparent that security seeking under anarchy did not necessarily lead to war, or even conflict. A third school dubbed neo-classical realism attempted to return to the roots of realism by admitting the importance of domestic politics and the variety of human motivations (Rose 1998).

Meanwhile, idealism as realism's foil was replaced first by neo-liberal institutionalism (Keohane 1984) and then by liberalism tout court (Moravcsik 1997). The former allied itself with the emerging literature on the repeated Prisoner's Dilemma game (Axelrod 1984), and argued that institutions provided information that could ameliorate the negative effects of anarchy. The latter argued that preferences and domestic politics are more important than international interaction. The more radical paradigmatic alternative was once Marxism, but, since the end of the Cold War, constructivism has largely supplanted it (Wendt 1999, Hopf 2013). Constructivism focuses on state identities and argues that anarchy can be ameliorated and eventually transcended through transformations in state identity.

By the turn of the millennium, if not before, it had become apparent to many scholars that paradigmatic warfare had run its course as a mode of research. Realism and liberalism no longer seemed to have distinct identities worthy of paradigmatic status, and had been reduced to strands of rational choice analysis applied in different contexts, e.g. political economy vs. security, that emphasized different variables, domestic politics vs. international factors. Rationalism vs. constructivism emerged as the new grand debate,⁹ but skepticism almost immediately emerged from two leaders of the respective camps (Fearon and Wendt 2002). References to the paradigms gradually dropped out of articles, and eventually even from introductions and abstracts.

⁹ See the 50th anniversary issue of *International Organization* Volume 52, No. 4.

9 1.3 PARADIGM WARS AND PROBLEM ORIENTED RESEARCH

A survey of academics revealed that the discipline was still perceived to be dominated by paradigms, but this was increasingly not the case in the actual journal articles being published (Maliniak *et al.* 2011). Graduate students were taught to scorn paradigmatic thought where they had once been encouraged to choose sides. In a recent expression of this train of thought, David Lake characterized paradigmatic debate as “evil” and a hindrance to progress (Lake 2011).

The modern alternative to paradigmatic warfare is “problem oriented research.” The eager graduate student now is supposed to pick a problem of interest and importance and consider what the best approach would be to tackle it. Theories are brought in eclectically as needed (Sil and Katzenstein 2010) and empirical methods, likewise, determined by the problem at hand. The goal is to explain the phenomenon rather than to validate a particular theory, in crude statistical terms to maximize the R^2 rather than the t statistic of your favorite variable.

The game theoretic approach stands athwart this trend, not exactly shouting stop, but certainly not offering any encouragement. The rational choice approach to politics is subject to many of the same critiques levied by the critics of paradigmatic debate (Green and Shapiro 1994, Brown *et al.* 2000b). It is theory driven. It seeks to expand and develop a particular perspective rather than understand a particular problem. Its practitioners do not wonder what would be the best way to tackle a new research question, they wonder what would be the best way to model it within the traditional framework. The boundaries of the approach are flexible, and can be expanded to include phenomena such as bounded rationality, other regarding preferences, and a concern for fairness. However, the bulk of applied work is in the traditional mode, focusing on national actors with self-regarding preferences.

One aspect of the modeling enterprise sets it apart from its predecessor paradigms, namely the rigor that goes along with mathematical formulation. This rigor allows it to progress and cumulate in a way that was difficult to discern, although not entirely absent, in the verbal theoretical debate. For example, when prominent rational choice scholars published an article in a leading journal of political science and a graduate student found a mathematical error that vitiated their results, their response was to acknowledge the error – an exchange that would be unthinkable in verbal paradigmatic debate (Bueno de Mesquita *et al.* 1997, Molinari 2000).¹⁰ This ability of the theory to correct itself also enables it to cumulate. Once a model is well formulated and solved correctly, it is done, and there is no need to wonder what it really means or write an exegesis of it. It can be built upon and extended by future scholars in a way that more closely resembles a body of work rather than an endless debate.

¹⁰ The fact that the correction itself contained an error either strengthens or undermines the point, depending on one’s perspective.

1.4 The utility of a partial view

This book is necessarily only a partial view of international relations and the theory thereof. It would be difficult to write a truly synoptic theory of international relations at present, given the diversity of approaches and the relatively fluid state of the field.¹¹ The usefulness of such an introduction will be most apparent for those already committed to, or at least interested in, the rationalist approach to international relations: liberals, realists, political economists, and the like.

I would argue that it may still have utility for those indifferent to or even hostile to the rationalist view. Problem oriented scholars, with no commitment to any particular theoretical perspective, nonetheless benefit from the availability of coherent theories and associated empirical hypotheses for testing. Game theoretic models often reveal strategic interdependencies or selection issues that are not apparent at first blush, and require careful consideration of the threats to inference (Signorino 1999). Scholars working in the psychological approach benefit from the identification, at least at the theoretical level, of a “rationalist baseline” for behavior. That is, when we know what behavior can be rationalized under certain conditions, we can begin to assess more carefully the role of psychological biases in cognition and decision making. Even constructivists, who might seem to have the least to gain from a rationalist textbook, may benefit by having a clear and coherent exposition of the position they find themselves in opposition to. The interchange between Wendt and Fearon on the rationalist–constructivist debate illustrates how greater communication between the two camps can dispel mistaken conceptions of what divides them (Fearon and Wendt, 2002).

This book should be read alongside other approaches, both theoretical and empirical. If it can contribute to strengthening the understanding of the game theoretic approach, both by adherents and opponents, it will have achieved an important aim.

1.5 Conclusion

The goal of this book is to provide a clear, structured understanding of rational choice theory as it is currently applied to international relations. The approach will be to develop families of models on the main topics that have been investigated with game theory in as systematic and understandable a way as possible. It cannot hope to be comprehensive – there will be many deserving models and whole topics that are not covered or even cited here, but I hope it will serve as a useful doorway into the field.

¹¹ See Carlsnaes *et al.* (2002), Reuss-Smit and Snidal (2008), and Dunne *et al.* (2013) for some broad surveys.