CHAPTER 1

Introduction

WHEN MIKHAIL GORBACHEV came to Washington D.C. in December 1987 for a summit meeting with Ronald Reagan, the U.S. President took the opportunity to repeat for the cameras one of his favorite Russian proverbs. The phrase, *doveryai no proveryai* (trust but verify), became indelibly associated with the two men and the end of the Cold War.[1] The phrase nicely captured the mistrust that plagued the superpower relationship while at the same time suggesting that trust could be rebuilt if words were accompanied by deeds that could be verified. As if inspired by the proverb, the Intermediate-range Nuclear Forces (INF) treaty signed at the summit contained verification provisions that were unprecedented in U.S.-Soviet arms control.

This book is about the role of trust and mistrust in international relations and the Cold War. I define trust as a belief that the other side is trustworthy, that is, willing to reciprocate cooperation, and mistrust as a belief that the other side is untrustworthy, or prefers to exploit one's cooperation. The topic is important because trust and mistrust can make the difference between peace and war. States that trust each other sufficiently can cooperate; states that do not may end up in conflict. As a result, states constantly make inferences about each other's motivations. In the Cold War, for instance, from George Kennan's famous 1947 article on the sources of Soviet conduct to the debates over Gorbachev's policies of *glasnost* and *perestroika*, the United States was obsessed with the question of whether the Soviet Union was innately expansionist and whether and over what it could be trusted (Kennan 1947; Allison 1988).

Indeed, trust is central to our understanding of the Cold War. Two of the most important questions asked about the Cold War are why it began and why it ended when it did. Another key question is how the European states managed to cooperate with each other and eventually with Germany so soon after a devastating war that sowed deep fears and hatreds. These three questions are all related to international trust. With respect to the origins of the Cold War, many authors in the "post-revisionist" school of Cold War historiography have traced the origins of the Cold War to mistrust (Gaddis 1983; Leffler 1992). These authors argue that

---

[1] David K. Shipler, "The Summit: Reagan and Gorbachev Sign Missile Treaty and Vow to Work For Greater Reductions" *New York Times,* December 8, 1987, A1.

the Soviet Union and the United States were both animated by a search for security—a defensive goal—but that their desire for security propelled them into conflict. Thus, the Cold War takes on a tragic cast, because if only the two sides could have trusted each other, the conflict could have been avoided (Collins 1997). For instance, Deborah Larson argues that there were "missed opportunities" to end the conflict when both sides' interests supported a cooperative deal but mistrust prevented them from realizing it (Larson 1997: 5). Some argue that the United States should have pursued a policy of reassurance, to overcome this mistrust (Lebow and Stein 1994: 375–76).

Ranged against this interpretation are both "traditionalist" and "revisionist" accounts. Traditionalists in the United States believe that the Cold War was driven by the expansionist goals of the Soviet Union. The Soviets are seen as genuinely aggressive, not reacting to the West in a defensive manner. Hence, the West had to firmly oppose the Soviet Union through the policy of containment (Feis 1970). The West mistrusted the Soviets, it is true, but this mistrust was fully justified because the Soviets were untrustworthy. The mirror-image revisionist thesis argues that the Soviet Union was primarily defensively motivated while the capitalist West was the imperialistic and aggressive party. The Soviet Union, devastated by the war and fearing that Germany would eventually rise again from the ashes, had legitimate security interests in controlling its periphery. The United States, driven by the quest for markets for goods and investment, sought to roll back the advance of socialism and make the world safe for international capital (Kolko and Kolko 1972). For the revisionists, the United States is the untrustworthy actor, and Soviet mistrust is justified.

The question of European cooperation and German rehabilitation is also a matter of trust. During the 1930s, the European states had moved from cooperation to conflict. Germany had proven itself extremely untrustworthy and attacked its neighbors with genocidal fury. The task of overcoming this mistrust was formidable. Yet the European nations gradually raised their level of cooperation to heights never before achieved. From the alliance Britain and France signed at Dunkirk in 1947 through the founding of NATO to the rearmament of Germany, the Europeans and Americans cooperated and built institutions to cement their cooperative relationships. U.S. hegemony is often credited with fostering this cooperation, but the mechanism by which hegemony can foster cooperation in the face of mistrust is poorly understood.

Trust also plays a prominent role in debates about the end of the Cold War. Some argue that the key factor in the end of the Cold War is Soviet economic decline. Because the Soviet economy was stagnant while the West continued to grow, the Soviets were simply forced to concede defeat in the forty-year struggle (Brooks and Wohlforth 2000/01; Wohlforth 1994/95). For these analysts, the end of the Cold War is characterized

by capitulation, not reassurance. Others argue that trust building was central to the end of the Cold War. They claim that the Soviet Union changed fundamentally with Gorbachev's accession to power. The Soviets became less expansionist and more defensive in their international orientation (Risse-Kappen 1994; Checkel 1993; Evangelista 1999; Mendelson 1993, 1998; English 2000). This change led the Soviets to favor a more cooperative relationship with the West; in effect, it made them trustworthy. However, because preferences are not directly observable, the Soviets needed to take significant visible steps to reassure the West. Most important among Gorbachev's trust building initiatives were the INF treaty of 1987, the withdrawal from Afghanistan, and the eventual noninterference in the Eastern European revolutions of 1989 (Larson 1997: 221–34; Kydd 2000b: 340–51).

Thus, trust plays an important role in the debates about the beginning and end of the Cold War, and about European cooperation. The fact that many of these debates remain unresolved highlights the need for a better theoretical understanding of trust and cooperation in international relations. Toward this end, this book develops a theory of how trust affects cooperation between two actors as well as in larger groups, how it is eroded through aggressive behavior, and how it is enhanced through cooperative gestures designed to reassure.

There are four main implications of the theory of trust developed here. First, cooperation requires a certain degree of trust between states. The threshold of trust required for cooperation depends on a set of variables including a state's relative power and costs of conflict. Second, though conflict between trustworthy states is possible, when we see conflict it is a sign that one or both of the states are likely to be untrustworthy. Thus, we, as external observers, should become less trusting of the parties involved in a conflict, just as they themselves do. Third, in multilateral settings, hegemony—the presence of a very powerful state—can promote cooperation, but only if the hegemon is relatively trustworthy. Untrustworthy hegemons will actually make cooperation less likely. Fourth, if two parties are genuinely trustworthy, they will usually be able to reassure each other of this fact and eventually cooperate with each other. The key mechanism that makes reassurance possible is "costly signaling," that is, making small but significant gestures that serve to prove that one is trustworthy.

With respect to the Cold War, these implications support three arguments. First, the Cold War was most likely a product of expansionist drives on the part of the Soviet Union, not a mutual desire for security accompanied by mistrust.[2] Soviet expansionist behavior increased the suspicions

[2] I merely state the claims here; evidence for and against will be considered in the historical chapters that follow.

of contemporaries, and it should also increase our own, given the lack of contrary evidence that the Soviets were benignly motivated. Second, the European states were able to cooperate with each other, the United States, and Germany after World War II because the United States, as a trustworthy hegemon, enabled them to overcome serious mistrust problems. Contrary to prevalent explanations, the United States neither provided a free ride to the Europeans nor coerced them into accepting an American-preferred order. Finally, the Cold War was ended through a process of costly signaling. Gorbachev made a number of dramatic gestures that increased Western trust and dispelled the suspicions that underlay the forty-year conflict. Soviet economic decline, while important, does not by itself explain this process.

In this chapter I will first define what I mean by trust and distinguish this meaning from related ones. Second, I will discuss the role of trust in existing theories of international relations and lay out the essentials of my alternative approach. Finally, I will briefly discuss the methodological approach of the book.

Defining Trust

Trust can be understood in many different ways.[3] The definition that I will adhere to is that trust is a belief that the other side prefers mutual cooperation to exploiting one's own cooperation, while mistrust is a belief that the other side prefers exploiting one's cooperation to returning it. In other words, to be trustworthy, with respect to a certain person in a certain context, is to prefer to return their cooperation rather than exploit them. To be untrustworthy is to have the opposite preference ordering. Cooperation between two actors will be possible if the level of trust each has for the other exceeds some threshold specific to the situation and the actors.

Some concepts from game theory will help make this understanding of trust more precise. In the single play Prisoner's Dilemma, illustrated in Figure 1.1, each side has a dominant strategy to defect, that is, it is in their interest to defect no matter what they think the other side will do. Even if one side thinks the other will cooperate, it will want to defect. This means that *actors with Prisoner's Dilemma preferences are untrustworthy* as defined above because they prefer to meet cooperation with defection.

[3]See Hardin 2002: 54, for a discussion and critique of different conceptions of trust; Coleman 1990: 91–116, for an influential discussion of trust from a rational choice perspective; and Hoffman 2002, for a discussion of the concept in international relations. See also Luhmann 1979; Seligman 1997; Bigley and Pearce 1998; Braithwaite and Levi 1998; and Ostrom and Walker 2003.

Player 2

	Cooperate	Defect
Cooperate	3, 3	1, 4
Defect	4, 1	2, 2

Figure 1.1 The Prisoner's Dilemma

As a result, in the Prisoner's Dilemma, mutual defection is usually thought to be unavoidable and it is the only Nash equilibrium in the game.[4] Each player believes that the other side would prefer to exploit cooperation rather than reciprocate it, and they are right. Two untrustworthy actors facing each other will not cooperate.

In the Assurance Game, illustrated in Figure 1.2, the player's preferences are different. As in the Prisoner's Dilemma, each side prefers to defect if it thinks the other side will defect. However, if one side thinks the other will cooperate, it prefers to cooperate as well. This means that *players with Assurance Game preferences are trustworthy*. They prefer to reciprocate cooperation rather than exploit it.[5] The fact that in the Assurance Game it makes sense to reciprocate whatever you expect the other side to do means there is a Nash equilibrium in which both sides cooperate. Cooperation is possible between trustworthy types who know each other to be trustworthy. There is also a Nash equilibrium in which the players do not cooperate, because each side prefers to meet defection with defection. However, in the Assurance Game as so far stated, this equilibrium seems unlikely given that both players prefer the equilibrium involving mutual cooperation and

---

[4]A Nash equilibrium is a set of strategies that are best responses to each other (Osborne 2004: 21).

[5]Trust is not equivalent to reciprocity, however. Reciprocity is a behavioral pattern, returning good for good and ill for ill (Keohane 1986: 8). Trust is a belief that the other side is willing to engage in reciprocity.

Player 2

	Cooperate	Defect
Cooperate	4, 4	1, 3
Defect	3, 1	2, 2

Player 1 (labels rows Cooperate / Defect)

Figure 1.2 The Assurance Game

nothing is preventing them from coordinating on that one rather than the less desirable mutual defection equilibrium.

Players with Assurance Game preferences might fail to cooperate, however, if they were not *sure* that the other side had Assurance Game preferences. For instance, if one side thought the other might have Prisoner's Dilemma preferences, it would be natural to hesitate before cooperating, because the other side would then have a dominant strategy to defect. If the other side had Prisoner's Dilemma preferences, persuading them that you plan to cooperate would not induce them to cooperate in return, because they would prefer to exploit cooperation rather than reciprocate it. The Assurance Game player might then decide to defect, not because it prefers mutual defection, but because it fears that the other side has a dominant strategy to defect and therefore cannot be persuaded to cooperate. This is the problem of mistrust. Trustworthy Assurance Game actors may fear that they face an untrustworthy Prisoner's Dilemma player, and hence decide to not cooperate.[6]

Along these lines, we can think of the *level of trust* one actor has for another as the *probability it assesses that the other actor is trustworthy* (Hardin 2002: 28). For instance, if player 1 thinks there is a t_2 chance that player 2 has Assurance Game preferences, we can think of t_2 as player 1's level of

[6]For similar analyses of trust in international relations, see Snyder 1971; Bennett and Dando 1982, 1983; Wagner 1983; Plous 1985, 1987, 1988, 1993; and Glaser 1997: 184.

trust for player 2. Similarly, player 2 will think there is a t_1 chance that
player 1 is trustworthy, and has Assurance Game payoffs, and a $1 - t_1$
chance that player 1 is untrustworthy and has Prisoner's Dilemma payoffs.
The greater t_1 and t_2, the more likely the other side is to be trustworthy,
and the higher the level of trust.

Finally, cooperation is possible when the level of trust for the other
exceeds a *minimum trust threshold* for each party (Luhmann 1979: 73).
The minimum trust threshold will depend on the party's own tolerance
for the risk of exploitation by the other side. In the example above
based on the Prisoner's Dilemma and the Assurance Game, if each side
anticipates that the other side will cooperate if they are an Assurance
Game type and defect if they are a Prisoner's Dilemma type, coop-
eration gives player 1 (if it is an Assurance Game type) a payoff of
$t_2 \times 4 + (1 - t_2) \times 1$, while defection yields $t_2 \times 3 + (1 - t_2) \times 2$; so
cooperation will make player 1 better off than defection if the level of
trust, t_2, exceeds 0.5 (the same calculation holds for player 2). Thus if the
other side is at least 50 percent likely to be trustworthy, it is worthwhile
cooperating with them, but if the level of trust falls below 50 percent,
trustworthy actors will defect because of mistrust. The minimum trust
threshold for the actors in this case is therefore equal to 50 percent, and
cooperation is possible if the level of trust exceeds the minimum trust
threshold.

To trust someone, then, as I will use the concept, is to believe it rela-
tively likely that they would prefer to reciprocate cooperation. To mistrust
someone is to think it is relatively likely that they prefer to defect even if
they think one will cooperate. This conception of trust is related to but
distinct from others in the literature. Two in particular are especially preva-
lent in the study of international relations: trust as belief that the other will
cooperate in a Prisoner's Dilemma and trust as a belief about anticipated
behavior rather than about preferences.[7]

The Prisoner's Dilemma and similar extensive form games are the most
common models used to analyze trust (Deutsch 1958; Dasgupta 1988;
Kreps 1990: 65–72; Gibbons 2001; Camerer 2003: 83–92). However,
these models fail to provide an adequate framework for understanding
trust. In the single shot Prisoner's Dilemma and related extensive form
games there is a dominant strategy to exploit the other side, and, hence,

---

[7]Three others are more tangentially related. Trust can be thought of as an equilibrium
selection device in games with multiple equilibria like the Assurance Game. Trustworthiness
can be thought of as a propensity to tell the truth, as in cheap talk models (Farrell and Rabin
1996; Sartori 2002). Finally, trust can be thought of as a form of social capital (Fukuyama
1995, 1999).

no reason to trust anyone.[8] This makes trust irrational by definitional fiat, and forces those using such models to conceive of trust as a form of naiveté and trustworthiness as a species of irrationality (Camerer 2003: 85). It presupposes that no rational self-interested actor could possibly prefer mutual cooperation to exploiting the other side's cooperation. Yet examples of such a preference ordering are easy to come by, for instance a state that just wishes to be secure might rationally prefer not to develop an expensive new weapons system if it were assured that its neighbors would show similar restraint. That is, rational self-interest can support cooperation even in single shot games.

The Prisoner's Dilemma framework also makes it difficult to investigate the uncertainty which is at the heart of trust problems. In these games, strictly speaking, there is no uncertainty about motivations or behavior since everyone has a dominant strategy to defect. As a result, to attain any degree of realism, analysts must smuggle uncertainty in through the back door. Coleman, in his influential discussion of this type of trust game essentially adds incomplete information without drawing the game tree (Coleman 1990: 91–116) and many other treatments add uncertainty about what the other side will do without treating it formally. Since trust is fundamentally concerned with this kind of uncertainty, uncertainty needs to be at the center of the model, not left as an informal addendum to a complete information game.

These problems also dog the repeated Prisoner's Dilemma. In the indefinitely repeated Prisoner's Dilemma, cooperation can be sustainable if the players care enough about future payoffs because they will fear that attempts to exploit the other side will be met with retaliation (Axelrod 1984). This framework has been used to analyze trust in experimental settings (e.g., Deutsch 1958; Swinth 1967; Kollock 1994; Parks, Henager, and Scamahorn 1996), as well as in more philosophical discussions. Russell Hardin argues that "I trust you because your interest encapsulates mine, which is to say that you have an interest in fulfilling my trust" where the source of this interest is the desire to keep mutually beneficial relationships going over time (Hardin 2002: 3). However, the repeated Prisoner's Dilemma model suffers from the same problem as the one shot game. There is no uncertainty in the game about whether the other side prefers to sustain the relationship. Either future payoffs are valued highly enough to make sustained cooperation worthwhile, or they are not and the parties will rationally defect. Trust is therefore perfect or nonexistent. To adequately model trust in the context of the repeated

---

[8]The unsuitability of the Prisoner's Dilemma as a model of trust was first pointed out, to my knowledge, by Gordon Tullock in a brief comment (Tullock 1967). See also Held 1968 and Birmingham 1969.

Prisoner's Dilemma one must introduce some uncertainty, either about preferences or about how much the parties value future interactions, so that the players can be rationally uncertain about whether the other side does prefer to reciprocate cooperation. The resulting game is essentially a repeated version of the mixed Assurance/Prisoner's Dilemma game just discussed.[9]

A second alternative conception of trust is to think of it as a belief about the probability that the other side will cooperate. That is, a belief about the likely behavior of the other side, not about their preferences. Diego Gambetta defines trust in this way, "When we say we trust someone or that someone is trustworthy, we implicitly mean that the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in cooperation with him" (Gambetta 1988: 217). Partha Dasgupta seconds this definition, "I am using the word 'trust' in the sense of correct expectations about the *actions* of other people" (Dasgupta 1988: 51)[10] as does Deborah Larson in her analysis of trust in the Cold War (Larson 1997: 12). Some scholars focus their inquiry on the *intentions* of other states, intentions to cooperate or defect (Edelstein 2000, 2002). One reason for defining trust as a expectations about behavior rather than beliefs about preferences is that behavior may seem more important than preferences. Behavior after all, is directly observable and clearly matters to others, while preferences are difficult to observe and do not directly affect others.

The problem with this conception of trust, however, is that trusted individuals sometimes fail to cooperate. In particular, if I am untrustworthy, I can anticipate that others who know this about me will fail to cooperate with me, even if they themselves are trustworthy. In international relations, a state like Hitler's Germany that has invaded its neighbors and committed atrocities in the past can expect that other states will not cooperate with it in the future. The reason is not that the other states are untrustworthy, the problem is that they think that Hitler is untrustworthy. Hitler might think Britain the most trustworthy state in the world, and yet realize that the likelihood that Churchill will cooperate with him is zero. As this example demonstrates, the motivations behind anticipated actions are crucial if one wishes to understand them. Therefore it is necessary to conceive of trust

---

[9]See Kydd 2000c: 412 for a version in which both players have identical Prisoner's Dilemma stage game preferences but there is uncertainty about how much they care about the future. Long-term thinkers are worried about being taken advantage of by fly-by-night operators. In terms of the definition discussed above, the long-haul types are trustworthy because they prefer to reciprocate cooperation while the fly-by-nighters are untrustworthy. See Hwang and Burgers 1999 for a similar but non-Bayesian analysis.

[10]Despite this Dasgupta actually focuses on motivations and presents one of the earliest incomplete-information games along the lines developed here (Dasgupta 1988: 52, 62).

in terms of underlying motivations, not just expected behavior.[11] With trust defined, we can turn to the place of trust in international relations theory.

TRUST AND INTERNATIONAL RELATIONS

The most obvious difference between international relations and domestic politics is that international relations take place in anarchy, whereas politics within states is conditioned by hierarchy. There has been considerable debate, however, about what, if anything, this distinction means in terms of the behavior that should be expected in each realm (Waltz 1979: 110; Milner 1991; Powell 1993). Much of this debate has its origins in conflicting ideas about the role of trust in a Hobbesian anarchy. I will first discuss trust under anarchy and the realist theories that are directly concerned with this problem, including the approach of this book. Then I will briefly discuss alternative liberal and constructivist perspectives on trust in international relations.

Anarchy and Trust

Thomas Hobbes was one of the first to theorize about the pernicious effects of anarchy. In the famous thirteenth chapter of *Leviathan*, he writes of the dreadful circumstances that prevail among men when, "there is no power able to over-awe them all."

> And from this diffidence of one another, there is no way for any man to secure himselfe, so reasonable, as Anticipation; that is, by force, or wiles, to master the persons of all men he can, so long, till he see no other power great enough to endanger him: And this is no more than his own conservation requireth, and is generally allowed. Also because there be some, that taking pleasure in contemplating their own power in the acts of conquest, which they pursue farther than their security requires; if others, that otherwise would be glad to be at ease within modest bounds, should not by invasion increase their power, they would not be able, long time, by standing on their defense, to subsist.
>
> (Hobbes 1968 [1651]: 184–45)

Here Hobbes argues that the best way to achieve security under anarchy is to destroy the power of others who might pose a threat. Preemptive or preventive attack is the best path to survival; one needs to attack

---

[11] The two conceptions could be reconciled if the behavioral definition was recast as an expectation that the other side will cooperate conditional on their having a high enough level of trust of one's self, or where they bear no risk from the transaction. But then we might as well talk about motivations rather than behavior.

others before they attack.[12] Hobbes understands that this might seem a bit extreme and that the reader may be wondering if it is really necessary to lash out at all potential threats. Why not simply maintain a strong defense, and only fight if attacked? Standing on the defense is inadvisable, Hobbes argues, because not everyone is motivated by security alone; some pursue conquest "farther than their security requires." The world is not composed solely of security seekers who would be happy to live and let live. In a world where other actors may be more aggressively motivated, the security seeker must attack preemptively and destroy the power of others because otherwise, eventually, he will be ground down by recurrent attacks (Kavka 1986: 97).

Applied to international relations, the Hobbesian argument says that given anarchy and mistrust, security seeking states will pursue aggressive policies up to and including war. This conflict will sometimes be tragic, because, in some cases, both sides will be motivated by security, a defensive consideration, not aggression. Hence, there is said to be a "security dilemma," a term coined by John Herz at the dawn of the Cold War (Herz 1950; Jervis 1978). Insecure states will pursue power to make themselves more secure; this renders other states less secure, and their efforts to catch up in turn render the first state less secure in a vicious circle. International conflict is a tragic clash between states with fundamentally benign desires to survive. In the latter years of the Cold War, Kenneth Waltz founded his influential structural realist theory of international relations on the security dilemma, and used it to argue for the virtues of bipolar systems and the irrelevance of economic interdependence (Waltz 1979).

A key component of the Hobbesian explanation of conflict is mistrust, the belief that other states may be aggressively motivated. Scholars disagree, however, on the precise role that trust plays in the security dilemma. Two existing schools of thought, offensive and defensive realism, both descended from Waltz's structural realism, contribute important insights about trust and international relations, but have serious limitations. To address these problems, I develop a new approach that I call *Bayesian realism* because it starts from the core realist assumption of the state as a unitary rational actor and relies on a game theoretic analysis of beliefs and behavior based on the Bayesian theory of belief change.[13] The three schools of thought and their key assumptions are shown in Table 1.1.

[12] Preemptive attack is motivated by the expectation of an imminent attack by the other. Preventive war is motivated by a longer term fear that one's enemy is gaining ground which will leave one open to coercion or attack in the future. See Levy 1988.

[13] In terms of the intrarealist debate, the approach is essentially a marriage of the neoclassical realist focus on states with different motivations (Schweller 1994, 1996, 1998; Kydd 1997b; Rose 1998) and the defensive realist concern with signaling motivations.

TABLE 1.1
Views on Anarchy and Trust

	State Motivations	Level of Trust
Offensive Realism	Security	Low
Defensive Realism	Security	Variable
Bayesian Realism	Mixed	Variable

OFFENSIVE REALISM

The most loyal Hobbesians argue that there is an irreducible level of
mistrust between states that prevents cooperation. Adherents to this view
include John Mearsheimer and other offensive realists.[14] Offensive realists
assume that "survival is the primary goal of great powers" (Mearsheimer
2001: 31). Since motivations are hard to discern, however, there is always
uncertainty about the intentions of other states, a permanent state of dis-
trust (Waltz 1979: 91–92, 118, 1988: 40; Mearsheimer 2001: 31–32;
Copeland 2000: 15).[15] In effect, offensive realists treat uncertainty and
distrust as a permanent background feature of the international system—
an element of the structure. Mistrust is a constant, like anarchy, not a
variable, like relative power. As Mearsheimer puts it, "There is little room
for trust among states. Although the level of fear varies across time and
space, it can never be reduced to a trivial level" (Mearsheimer 1994/5: 11,
1990: 12, 2001: 32). In effect, states make worst case assumptions about
other states' motivations (Keohane and Martin 1995: 43–44; Brooks 1997:
447–50). Other states may be expansionist, and one cannot tell for sure,
so there is no point attempting to differentiate between states based on
their underlying motivations. From an offensive realist standpoint, there is
little use in studying international trust, because the level of trust is never
high enough to affect behavior.

[14]There are currently many strands of realist thought operating under a variety of labels.
Waltz's (1979) theory was called "neo-realism" or "structural" realism. Snyder (1991: 12)
introduced the term "aggressive" realism which has been transmuted to "offensive" realism,
which was embraced by Mearsheimer for his refinement of Waltz. In my view, offensive realism
is closest in spirit to Waltz's theory because Waltz argues that states of differing motivations
behave the same, a key offensive realist claim. Others consider Waltz closer to defensive
realism because he claims that states maximize security, not power. However, Waltz provides
no compelling argument as to why power maximization is not behaviorally equivalent to
security maximization under his theory.

[15]Waltz's famous dictum that states "at a minimum, seek their own preservation and, at a
maximum, seek universal domination" suggests that the diversity of state goals is important,
but his analysis focuses exclusively on security seeking states.

Offensive realists also argue that mistrust in the international system never gets cleared up because security seekers behave the same as more aggressive states. Offensive realists claim that the violent nature of the international system—the constant competition and war—is not a result of the individual natures of states. As Waltz puts it, "in an anarchic domain, a state of war exists if all parties lust for power. But so too will a state of war exist if all states seek only to ensure their own safety" (Waltz 1988: 44). Expansionist states will invade for gain; security seekers to protect themselves; both will extend their power as far as they can. According to Mearsheimer, "states seek to survive under anarchy by maximizing their power relative to other states, in order to maintain the means for self defense" (Mearsheimer 1990: 12, 2001: 30–40; Labs 1997). A key implication of this argument is that since security seekers and aggressive states behave the same, they cannot be distinguished by their behavior, and, hence, the mistrust that causes conflict cannot be overcome. Mistrust and conflict form a self-reinforcing cycle in which the mistrust causes conflict that reinforces the mistrust.

A final implication of offensive realism is that to explain international events, attention should be focused not on the elements of the structure that are constant, anarchy and mistrust, but on those that vary, most importantly, relative power and the number of great powers (Mearsheimer 2001: 43). Wars are caused by shifts in relative power, miscalculations about power due to multipolarity, unbalanced power, the ease of conquest, etc. Long rivalries such as the Cold War are a simple result of the fact that the United States and the Soviet Union were the two strongest powers—the only ones that could hurt each other. While the Cold War was a product of mistrust, in the same way that it was a product of anarchy, it was not because the United States and Soviet Union mistrusted each other for some special reason, or more than they distrusted anyone else. Pick any two states, make them the most powerful states in the world, and they will mistrust each other enough to fight a cold war, if not a hot one.

The offensive realist approach is consistent with a static theory of international relations given pessimistic assumptions about initial levels of trust. As I show in chapter 2 in a static model of trust and cooperation, if the level of trust is too low, states will not cooperate regardless of their own motivations, and conflict will result between security seekers. However, once the game is made dynamic, so that states can make reassuring gestures to build trust over time, offensive realism is undermined. In a dynamic version of the security dilemma, modeled in chapter 7, rational security seeking states can reassure each other and cooperate regardless of how low the initial level of trust is. As a result, the level of trust can be raised sufficiently to support cooperation. This poses a serious problem for offensive realism, suggesting that powerful states need not end up in conflict if they are rational security seekers.

DEFENSIVE REALISM

In contrast to offensive realism, other scholars argue that mistrust does not always prevent cooperation, although it does occur frequently and is responsible for much international conflict. This perspective is often identified with Robert Jervis who coined the term "spiral model" (Jervis 1976: 62) and it is also associated with defensive realism, particularly the work of Charles Glaser (1992, 1994/95, 2002).[16]

Defensive realism shares the basic Hobbesian framework focusing on unitary actors operating in an anarchic environment. Like the offensive realists, defensive realists assume that states are security seekers. Glaser asserts, "states are motivated only by the desire for security" (Glaser 2002: 4). Adherents of the spiral model believed that the Soviets were security seekers like the Americans (Jervis 1976: 64,102). Charles Osgood, an early spiral modeler, called those who believe the Soviets were aggressive "neanderthals," while making it plain that, in his opinion, both sides in the Cold War sought security (Osgood 1962: 29). Stephen Van Evera argues that his version of realism assumes, "that states seek security as a prime goal, for reasons rooted in the anarchic nature of the international system" (Van Evera 1999: 11). Conflict between states is therefore genuinely tragic rather than a clash between good and evil (Spirtas 1996).

Defensive realism diverges from offensive realism in its analysis of how pervasive the mistrust is, and whether it is possible to do anything about it. Where offensive realists see mistrust as pervasive and constant, defensive realists see it as variable and amenable to change. Some states trust each other enough to cooperate. These states can have normal relations, enjoying mutual security. Other states, unfortunately, develop deep levels of mutual distrust for each other. To these states, the security dilemma logic applies, and conflict results.

Defensive realists and spiral modelers differ in their analysis of what produces the breakdown in trust and what can be done about it. Defensive realists adhere to the rational actor assumption and focus on signaling (Glaser 1994/95: 67). States sometimes engage in competitive arms racing behavior which can lower their mutual level of trust. To address the problem, they can signal their true motivations to each other by engaging in cooperative gestures that reassure. Factors affecting their ability to signal include the nature of military technology, or the "offense-defense balance" (Jervis 1978; Van Evera 1999: 117). When offensive weapons predominate, security seekers will have to develop offensive capabilities like more aggressive states; when defensive technologies are strong,

---

[16]For an overview of defensive realism, also derived from Waltz and coined by Snyder, see Taliaferro (2000/01).

security seekers can invest in defense and signal their nonaggressive goals.

Spiral modelers abandon the rational actor assumption and focus their analysis on psychological biases (Jervis 1976: 67; Larson 1997: 19). One such bias is the tendency of actors with benign self-images to believe, without justification, that others share this benign image, so that if others engage in hostile behavior it must be a result of malevolence on their part. Spiral modelers also have a psychological theory of reassurance, often known as Graduated Reciprocation in Tension-reduction (GRIT) for the version proposed by Osgood (Osgood 1962). GRIT argues that unilateral cooperative gestures can build trust and establishes a set of conditions that will maximize their effectiveness. While both of these theories are presented as nonrational or psychological, we will see in chapters 3 and 7 that they contain rational cores that can easily be modeled game theoretically.

Defensive realism, like offensive realism, makes some important contributions but has certain shortcomings. The main improvement over offensive realism is the openness to the possibility that the level of trust varies in important ways; it may sometimes be high enough to sustain cooperation and may sometimes be low enough to prevent it. However, the focus on security seeking states is problematic for three reasons. First, it prevents defensive realism from developing a complete strategic theory of international relations. For instance, Stephen Walt's theory of alliances argues that states balance against the threat presented by "other" states that may be aggressive in order to improve their security (Walt 1987: 18). But if the other states may be aggressive, it is not clear why the state under advisement is not, and it would seem that a complete theory of international relations must give full consideration to the incentives and constraints facing both security seeking and aggressive states (Schweller 1994). Otherwise it is impossible to capture the strategic dynamics facing states which are uncertain about each other's motivations.[17]

Second, the assumption of security seeking states biases the analysis in favor of explanations involving psychological bias. States are assumed to be uncertain about each other's motivations. For the theorist to also assume that states are security seekers is to imply that a state that thinks some other state is expansionist is making a mistake. The subject of inquiry then becomes how to explain such mistakes, and this makes psychological theories associated with the spiral model quite attractive. If instead we assume that states may be expansionist as well as security seeking, then mistrust is not necessarily a mistake and the bias towards psychological theories is eliminated.

---

[17] For this reason, defensive realism is perhaps best thought of as a theory of the foreign policy of security seeking states, rather than a theory of international relations.

Third, to build into the theory the assumption that states are security seekers is to treat as an assumption something that should be the subject of empirical inquiry. States are uncertain about each other's motivations. As analysts, we wish to investigate these beliefs and understand their origins. We therefore need to make our own judgments about state motivations, in order to determine whether a state's beliefs are correct or not. We will doubtless find cases where states are mistaken and others in which they are substantially correct. To arrive at these judgments requires empirical inquiry; an examination of the record of events and any documents that are available. We cannot simply assume the answer to such a question before we begin.

BAYESIAN REALISM

The approach I develop here shares certain assumptions with offensive and defensive realism, but departs from them in important ways. I start from the same Hobbesian framework and assume that states in anarchy may face threats to their survival so that security becomes an important goal. However, I also assume that states have many other goals and some may be willing to pursue conquest "farther than their security requires" as Hobbes puts it. Thus, some states may be interested primarily in security, others may be more aggressively motivated. I further assume that because there has been enough variation in state motivations historically, and because motivations are difficult enough to discern, states may be rationally uncertain about the motivations of other states. That is, despite whatever shared culture they may have, states can reasonably wonder what the motivations of other states are.

With this foundation, I then apply Bayesian theory to the question of how states form beliefs about each other's motivations, and how they behave in response to these beliefs. Instead of assuming certain motivations and then characterizing beliefs that fail to reflect them as mistakes, I simply ask how states of various motivations behave in situations with varied beliefs about each other. Bayesian theory provides a framework for answering such questions assuming the decisionmakers behave rationally given their beliefs and change their beliefs rationally in response to new information.

The Bayesian framework supports the defensive realist claim that states with benign motivations that believe each other to be benign can get along. Well justified trust can sustain cooperation. Conversely, states that have malevolent motivations and know this about each other will not get along. Well justified mistrust will lead to conflict and possibly war.

Contrary to defensive realism, however, Bayesian analysis does not reveal an inherent tendency towards unjustified mistrust in international relations. Rather, it indicates that *convergence on correct beliefs is more likely than*

convergence on incorrect beliefs. That is, although the learning process is noisy and prone to errors of all kinds, beliefs over time and on average are more likely to converge towards reality than to diverge from it. This implies that if a state is a security seeker, other states are more likely to eventually discover this rather than to remain convinced that it is aggressive. If a state is interested in power or expansion for its own sake, other states are more likely to come to believe this than to think that it is a defensively motivated security seeker. Mistaken beliefs may arise, whether unjustified trust or unjustified mistrust, but over time they are more likely to be corrected than to remain or be further exaggerated.

This claim about beliefs, combined with the fact that state motivations are assumed to vary freely, implies that of the conflicts we observe, a relatively small percentage will be driven by mistaken mistrust. If we see a conflict, it is more likely to have arisen because one or more of the parties has genuine non-security-related motivations for expansion, or is untrustworthy. Therefore we should become more confident that one or both sides was aggressive, and less confident that both sides were motivated primarily by security.

However, Bayesian realism does not assume that states know each other's motivations perfectly, or find them out easily. Nor, as will be discussed in chapter 7, can one simply ask if a state is aggressive or not. A Hitler has every bit as much incentive to pretend to be modest in his ambitions as someone who is genuinely uninterested in world conquest. There are many obstacles in the way of rational learning about the motivations of other states. However, there are also tremendous incentives to get it right. Misplaced trust can lead to exploitation; misplaced distrust can lead to needless and costly conflict.

The mechanism that enables states to learn about each other's motivations is cooperation. Because of the importance of avoiding unnecessary conflict, states are often willing to take a chance by cooperating with another country, in hopes of establishing mutual trust. These gestures provide the occasion for rational learning because they help to distinguish trustworthy states from non-trustworthy ones—security seekers from the more aggressive. In circumstances where security seekers and more aggressive states behave differently, trust can be built by observing what states do.

Preferences, Identity, and Trust

Bayesian realism posits that state preferences vary, some states are basically security seekers and others are more expansionist. In the next chapter, in the context of a model of the security dilemma, I will show how a set of underlying structural variables including relative power and the costs

of conflict help account for this variation. However, other, less "realist," factors have an important influence on state preferences as well, and are therefore important in understanding trust in international relations. This underlines the necessity of integrating realist theories with other theories of preference and identity formation (Legro 1996; Moravcsik 1997). While that task will not be central to this book, I will briefly discuss how two prominent schools of thought, liberalism and constructivism, contribute to such an analysis in ways that are particularly important in the case of the Cold War.

LIBERALISM

Liberal theories of international politics focus on how the domestic political system aggregates social preferences to generate national policy (Moravcsik 1997: 518). Such theories take as inputs the preferences and political power of important social actors, and the political institutions through which they interact. Variations in the interests and power of the different groups, or in the institutional environment, produce changes in policy that affect international behavior.

The most prominent liberal theory relating to trust is the well-known idea of the democratic peace. Pairs of states that are both democratic hardly if ever fight, while the same obviously cannot be said for pairs in which only one state is a democracy or in which neither is democratic (Rummel 1983; Doyle 1983a, b; Russett 1993; Chan 1997; Ray et al. 1998). Several explanations have been advanced to account for this finding, some of which relate to international trust. Kant made the argument that ordinary citizens are averse to war because they suffer the costs especially acutely (Kant 1991: 100). Democracy empowers the average citizen, at least in comparison to more restrictive regimes, so democracies may have a higher cost of fighting (Bueno de Mesquita and Lalman 1992: 153). Jack Snyder argues that cartelized political systems favor expansionist interests while democracy dilutes their influence and weakens their ability to promulgate expansionist myths (Snyder 1991: 39). Bruce Bueno de Mesquita and his colleagues argue that narrowly based regimes are more aggressive because they focus on doling out private goods to their retainers, while more broadly based regimes must focus on providing public goods to mass audiences (Bueno de Mesquita et al. 1999). Nondemocracies, therefore, may have a higher evaluation of the gains from conquest.

Putting these ideas together, liberal theory can be said to support two basic points related to international trust. First, because democracies find war costly and of little intrinsic benefit, democracies are more likely to be security seeking states. Conversely, nondemocratic states are sometimes more aggressive, less constrained by their citizens, more volatile. They may be security seekers, or they may not be; there is more variation in

the preferences of nondemocratic states. Second, given that democracy is a readily visible characteristic of a regime, other states will have relatively high confidence that a democracy is a security seeker (Bueno de Mesquita and Lalman 1992: 155). A pair of democratic states should have well-founded trust for each other because they are security seekers themselves and think the other side likely to be one too. Hence, there can be said to be a democratic security community, in which states correctly trust that disputes will be resolved short of force (Deutsch et al. 1957; Adler and Barnett 1998). Nondemocracies, having closed regimes, will not automatically be perceived as security seekers. Hence if they are to build trust, they must rely on other mechanisms, such as the type of signaling described in chapter 7.

Thus, liberal theory suggests that democracies are more likely to be security seekers than nondemocracies, and will have a higher degree of trust for each other than they do for nondemocracies. Both preferences and prior beliefs will be influenced by regime type. In the context of the Cold War, this suggests that U.S. trust for the Soviets should have been low, given their regime type, and that in the late 1980s the United States should have been sensitive to signs of democratization, such as the elections to the Congress of People's Deputies in 1988.

CONSTRUCTIVISM

The key concept in constructivist approaches is state identity. Identity has many definitions in the political science literature.[18] Two related conceptualizations focus on a set of norms that identify appropriate behavior and a type of state one aspires to be or group of states one aspires to belong to (Ruggie 1998; Wendt 1999; English 2000). Some constructivists focus on the domestic sources of state identity, looking at the culture and identities of groups within society (Hopf 2002). Others argue that state identities are affected by international factors. Aggressive behavior on the part of another causes fear and leads one to view it as an enemy, causing one to adopt the enemy role and in turn wish to do the other harm. Self-restraint and cooperation fosters trust and feelings of solidarity and, in turn, generates moderate preferences (Wendt 1999: 357). No pattern of identity is given in the nature of anarchy; anything can happen.[19]

Many constructivists would stress the incompatibility of their project with realism. However, there is little doubt that state identity affects state preferences, and, hence, the extent to which a state is trustworthy or not. A particularly important source of identity and prior beliefs in the context of the Cold War is ideology. Both sides had well articulated ideologies

[18] For an overview of definitions and conceptualizations of identity, see Abdelal et al. 2004.
[19] For a constructivist critique that reaffirms the primacy of Hobbesian enemy roles, see Mercer 1995.

that provided answers to questions about their own state's identity and the trustworthiness of other states on the international scene. As I will discuss in chapter 4, communist ideology, founded on class conflict, posited a relationship of general enmity between the socialist states and the capitalist world. Capitalist states were believed to be hostile both to the Soviet Union and to each other. Intracapitalist conflict could provide opportunities for tactical cooperation with capitalists, but the overall relationship would be hostile. Ideology, therefore, gave the Soviets a high level of gains from conflict, and a low level of trust. American anticommunism ultimately provided a similar set of ideological lenses, but, as we will see, this view was not dominant at the outset of the Cold War.

Constructivists also argue that identity change was important at the end of the Cold War (Koslowski and Kratochwil 1994). Much of the debate about Soviet motivations in the 1980s was over whether they were being conciliatory because they were simply recognizing a temporary weakness, or because they had experienced a genuine transformation of identity into a state that no longer sought to expand its influence and subvert others. For instance, Robert English argues that the Soviets experienced a change in identity and became convinced that the Soviet Union should be a part of the community of Western democratic states (English 2000). Constructivist analyses have therefore been prominent in debates over the end of the Cold War and I will return to them in chapter 8.

Liberalism and constructivism offer insights into the preferences and prior beliefs that underly the strategic models of trust explored in subsequent chapters. A complete understanding of trust in international relations must integrate realist analysis of international strategic interaction with other theories of preference and identity formation offered by nonrealist theories.

METHODOLOGY

This book makes extensive use of game theoretic models of international relations. These are abstract, stylized representations of certain aspects of the world, expressed in the language of mathematics.[20] There are at least two important benefits of formal theorizing. First, it forces one to specify all the assumptions that one is making and verify that the logical connections

[20] See Powell 1990, 1999; Downs and Rocke 1990; Niou, Ordeshook, and Rose 1989; and Bueno de Mesquita and Lalman 1992. For methodological reflections on rational choice in international relations, see Lake and Powell 1999 and Sprinz and Wolinsky 2004. For a critique of the approach in IR, see Walt 1999 and the responses, subsequently collected in Brown et al. 2000. For discussion of rational choice in the study of American politics, see Green and Shapiro 1994 and Friedman 1996.

between these assumptions and the subsequent claims is ironclad. Second, it provides a common and rigorous theoretical language that enables others to check the soundness of one's results, making the theoretical enterprise more cumulative. The chief criticisms that have been leveled against it are that it tells us nothing that we did not already know and that its practitioners have failed to do adequate empirical work to validate it (Walt 1999).

I develop a set of closely related game theoretic models about trust, mistrust and reassurance. These models constitute the core of a rational choice theory of trust and conflict. In each chapter I derive general implications from the models about how the preferences, beliefs, and other attributes of the players influence their ability to cooperate under various conditions. In the historical chapters I describe the important actors, their preferences and beliefs, to the extent that they can be reconstructed, and how their beliefs and behavior correspond to the implications of the theory. For my data, I rely on published documents and secondary historical sources. The goal is to improve our understanding of the fundamental role that trust plays in international relations theory and of the role it played in the Cold War by making use of game theory's ability to clarify arguments and rigorously derive implications from assumptions.

How can the models be used to elucidate historical questions? This question is part of the broader issue of how political science, international relations, and the study of history intersect and interact. The connection between nonformal political science and history has been extensively debated (Elman and Elman 2001) and both sides have been enriched. The intersection of formal theory and statistical empirical work is also a focus of methodological research (Signorino 1999) with many applications (e.g., Bueno de Mesquita et al. 1999). However, formal theory and history have remained unfortunately estranged (Braumoeller 2003: 388; for exceptions see Bates et al. 1998 and Schultz 2001).

There are two ways models can contribute to historical analysis. First, models may make predictions about relations between observable indicators which can be verified by examining the historical data in a particular case, or set of cases, just as nonformal theories do (Levy 2001:48). This is the standard process of scientific inference where observable data are used to make inferences about unobservable causal relations posited by the theory (Hempel 1966: 6; King, Keohane, and Verba 1994: 75). For instance, in chapter 6, I compare the implications of three different theories of hegemony in the context of the post-World War II period. The theory I advance predicts active willing cooperation on the part of the European states, while the other two predict passivity or cooperation in response to U.S. coercion. The historical record provides many instances in which the Europeans willingly initiated cooperation, supporting the first theory.

A second way in which models can shed light on history is by helping us make inferences about other sometimes difficult to observe phenomena, namely state preferences and beliefs.[21] Some states are relatively transparent and data on their preferences can be gathered from foreign office records, interviews, and other public records. Others are more closed and give few direct visible clues about their motivations. In such cases, the state's behavior and the structure of the situation it finds itself in can be used to make inferences about its beliefs and motivations.

The effort to find such "revealed preferences" is widely argued to be inherently circular (e.g., observed cooperation suggests a state prefers to cooperate which explains why it cooperated) or bound to be frustrated by strategic considerations. Jeffry Frieden argues that, "where actors are strategic, we cannot infer the cause of their behavior directly from their behavior" (Frieden 1999: 48; see also Snidal 1986: 40–41). He gives an example in which a firm demands a tariff instead of the quota it really wants because it knows the government, which most prefers free trade, will reluctantly grant a tariff, but would reject a quota. In this case, the firm is not asking for what it most wants and the government is not implementing its favorite policy either. The inference is that the strategic setting prevents actors from revealing their true preferences through their behavior.

While it is true that the strategic setting certainly influences how we should interpret behavior, it is not the case that strategic considerations always prevent inferences about preferences. Models provide a set of variables, including the unobserved beliefs and preferences and other observable variables, and a theory of how the variables relate to each other. This structure may enable one to reason backwards from observed events to unobserved beliefs and motivations (Lewis and Schultz 2003; Signorino 2003).[22]

When can this process succeed in shedding light on state preferences and beliefs? In the kind of incomplete-information game theoretic models that are appropriate for such situations, there are broadly speaking two kinds of equilibria, *pooling* and *separating*. In pooling equilibria, actors with

---

[21]At a high enough level of generality, these tasks are the same. The general problem of inference involves creating a model based on certain assumptions and seeing which values of certain variable parameters maximize the relative likelihood of the realized data (King 1989: 22). Whether the parameters represent causal coefficients or preferences affects the structure of the model but not the nature of inference.

[22]For a pioneering example in the context of the Cold War, see Gamson and Modigliani 1971. Unfortunately Gamson and Modigliani's decision model does not rigorously deduce connections between behavior and underlying preferences and beliefs, which in turn undermines confidence in the resulting inferences from observed behavior. For a related literature on inferring preferences from voting behavior, see Poole and Rosenthal 1997; Voeten 2000; and Martin and Quinn 2001.

different motivations behave similarly, so no light is shed on their underlying preferences. This corresponds to the offensive realist world in which security seeking and expansionist states behave the same, so no inferences can be made about state motivations. In separating equilibria, in contrast, actors with different preferences and beliefs behave differently. In this case, one can reason backwards from actions to preferences. The information provided by the actor's behavior may not be conclusive, and there will often still be room for doubt. However, in separating equilibria one ends up with beliefs that are, on average, more likely to be correct than the prior beliefs one had before observing the behavior. Consider a historical question such as what we can learn about Stalin's motivations from his behavior in the Berlin crisis. To claim that we can learn anything is to claim that Stalin faced a separating equilibrium in a game of incomplete-information in which, if he had more moderate motivations, say, he would have not initiated the crisis or would have ended it sooner. Should we infer from Gorbachev's 1985 test moratorium that Soviet preferences with respect to the arms race had shifted? If we believe that he was playing his part in a separating equilibrium in which a hard line Gorbachev would have rejected the moratorium idea, such an inference is supported. By better understanding how separating equilibria work in incomplete-information games we may be able to better ground our answers to these questions of historical interpretation. In this way, game theory may directly enhance our historical understanding.

Note that in making inferences about preferences and beliefs, we as scholars are doing exactly what the subjects we study were doing. The participants in any historical episode have some advantages over us and some disadvantages. Their advantages are their better understanding of their own motivations and perceptions, things which we must make inferences about from the writings they leave behind and from their actions. Their disadvantages may include a lesser degree of knowledge of the other side's motivations and beliefs, if documents have become available from the other side. But even if some documentary evidence is available, we are still in a position of making inferences about the beliefs of each side, just as the participants were.

A Road Map

This chapter and the next constitute the introduction to the book. The next chapter presents a model of trust in the context of the security dilemma, one of the foundational structures of international relations.

The main body of the book consists of three pairs of chapters. Chapter 3 focuses on the spiral model, a key component of defensive realism, and

analyzes how trust can be eroded by competitive behavior. A common explanation of conflict is that interstate competitions, such as arms races, worsen mutual mistrust by convincing states that their rivals are out to get them. The chapter addresses how this worsening of mistrust can happen and how likely it is that it will occur between states that are actually trustworthy. Chapter 4 looks at the beginning of the Cold War and the increasing distrust between the United States and the Soviet Union as an instance of this phenomenon. As mentioned earlier, the central argument is that this growing mistrust was probably justified and not a misperception. The origins of the Cold War are probably not a tragedy of misperceptions or uncertainty induced failures to cooperate. Rather, a more likely hypothesis is that each side, but particularly the Soviet Union, harbored non-security-related motivations for expansion that would have led them to exploit, not reciprocate the cooperation of their rival.

The next pair of chapters examine trust and cooperation under hegemony. Chapter 5 presents a trust game involving multiple actors, which provides the tools to analyze how trust impacts cooperation when there are many actors with different interests, geographical situations, and relative capabilities. Chapter 6 applies the model to post-World War II European cooperation between the United States, Germany, and the rest of Europe. While many have argued that hegemony facilitated cooperation in the postwar era, the role of trust in hegemonic theory is understudied. I show that for hegemony to promote cooperation, the hegemon must be relatively trustworthy, in comparison to other states.

The final pair of chapters focuses on reassurance, looking at how states starting from a position of mistrust can reassure each other about their motivations and promote cooperation. Chapter 7 presents a model of reassurance via costly signals that shows how states can reassure each other by running risks of exploitation by the other side. Chapter 8 analyzes the end of the Cold War in these terms, focusing on the process of reassurance that took place between East and West. Here I argue that the Soviet Union changed from an expansionist state to a security seeker, but that this change was not transparent. Therefore, Gorbachev implemented a policy of costly signaling to reassure the West.

Conclusion

Trust is a central issue in international relations, and that centrality is exemplified in the most important struggle of the second half of the twentieth century, the Cold War. When states can trust each other, they can live at peace, provided that they are security seekers, uninterested in expansion for its own sake. States that are security seekers therefore pay close attention

to the motivations of others, attempting to determine who is a fellow security seeker and who is more inherently aggressive. In the Cold War, the United States attempted to determine if the Soviets were security seekers who could be reconciled with the status quo if sufficiently reassured about U.S. intentions, or were more aggressive. The end of the Cold War was marked by reassuring gestures on the part of the Soviets designed to alter Western perceptions of their motivations. How these processes work is the subject of this book.