

A Conceptualization of Trust in International Relations

AARON M. HOFFMAN
Purdue University, USA

This article provides a conceptualization of trust in interstate relations in terms amenable to research. In the field of International Relations trust is often equated with the willingness to take risks on the behavior of others. However, this approach is problematic because it does not provide a basis for distinguishing between trusting and non-trusting relationships. In contrast, I propose that trust implies a willingness to take risks on the behavior of others based on the belief that potential trustees will 'do what is right'. Recognizing that trust involves particular beliefs about the motivations of others distinguishes it from the broader category of risk and enables trusting relationships to be identified more precisely. After elaborating my definitional approach, I discuss indicators designed to gauge the presence of trusting interstate relations in empirical settings.

KEY WORDS ♦ cooperation ♦ institutional design ♦ measurement ♦ obligation ♦ trust

Academics and policy-makers have long recognized a link between trust and the peaceful resolution of interstate disputes. For example, over 40 years ago, Karl Deutsch identified trust as part of the cement upon which peaceful relations are built (Deutsch et al., 1957). More recently, United Nations Secretary General Kofi Annan characterized the December 2000 peace agreement between Ethiopia and Eritrea as an opportunity for both sides to 'build trust and work for reconciliation' (Associated Press, 12 December 2000). Similarly, the 1993 Joint Declaration on Peace between the British and Irish governments explicitly linked trust to the achievement of 'an agreed and peaceful time'.

However, it is clear from repeated failures to advance the cause of peace that our understanding of how trusting interstate relationships develop requires attention. But unless more attention is paid to developing valid measures of trusting relationships intellectual breakthroughs are unlikely to occur in this area.¹ The traditional approach to measuring trusting relationships assumes that trust is a necessary condition for cooperation. Yet, this supposition winds up overestimating the incidence of trusting relationships, implying, for instance, that trust operates even when cooperation is coerced.

The goal of this article is to develop a series of measures that are more sensitive to the presence of trusting relationships than existing indicators. At root, this is a theoretical project because measures reflect the ways abstract constructs are conceptualized (Adcock and Collier, 2000; Cook and Campbell, 1979). The first part of this article, then, is devoted to considering the definition of trust. By and large, scholars agree on the elements that a definition of trust must include, with one exception — some equate trust with the willingness to take risks, while others suggest that trust involves the willingness to take risks *and* the expectation that others will honor particular obligations. The distinction is critical because it directly affects the way in which we classify empirical cases. My analysis suggests that equating trust with the willingness to take risks disables our capacity to distinguish between cases of trust and cases of distrust. For this reason, I argue that studies of trust are better served by definitions that combine the concepts of risk and obligation.

This article concludes with a discussion of three measures designed to identify trust empirically. The first detects trusting relationships by connecting the decision-making record to the policy choices of leaders. The second approach examines the types of oversight actors employ to monitor each other's behavior and the third focuses on the restrictiveness of rules actors develop to constrain each other's activities. All three measures assume that trust manifests itself in the form cooperation takes and that identifying trusting forms of cooperation requires indicators that are sensitive to its particular features. Efforts to cooperate can be motivated by a variety of factors, including trust, but cooperation based on trust has distinctive features that can be exploited for the purposes of measurement.

Conceptualizing Trust

How should we define trust? Currently, no single definition serves as a focal point for research.² Nevertheless, substantial agreement exists about the elements any definition should either include or imply. First, scholars agree that *trust* refers to an attitude involving a willingness to place the fate of

one's interests under the control of others. This willingness is based on a belief, for which there is some uncertainty, that potential trustees will avoid using their discretion to harm the interests of the first (e.g. Baier, 1986; Brenkert, 1998; Coleman, 1990; Gambetta, 1988; Govier, 1997; Hardin, 1993; Larson, 1998; Luhmann, 1979).

Second, scholars agree that *trusting relationships* are behavioral manifestations of trust.³ Trusting relationships develop when actors grant others discretion over their interests based on the belief that those interests will not be harmed (Snijders, 1996). In interstate relations trusting relationships develop when leaders enact policies that delegate control over their states' interests based on the belief that their counterparts are trustworthy. The undefended border between Canada and the United States (US) and the 1976 agreement between Thailand and Malaysia giving the Malaysian military the right of 'hot pursuit' across the Thai border are examples of policies that create vulnerabilities between states. Both could be 'trusting policies' depending upon the reasons leaders had for establishing them. Note that trust, the attitude, is not a sufficient cause of trusting relationships. For example, even though on the eve of his first official visit to Washington, Mexican President Vicente Fox asserted that US–Mexican relations had been infused with trust (*New York Times*, 4 September 2001: A27), nevertheless, Mexico was not rewarded immediately with new bilateral agreements regulating immigration and the drug trade because the terrorist attacks of 11 September 2001 pushed these issues off the US foreign policy agenda.

Third, the intensity and scope of trust and trusting relationships are capable of variation. With respect to trust, intensity refers to the strength of an actor's perception of the trustworthiness of others. With respect to trusting relationships, intensity refers to the amount of discretion trustors grant trustees over their interests. The scope of trust and trusting relationships can vary as well, although the extent of this variation is disputed. Hardin, for example, argues that trust is never unconditional, that it always implies a three-part relation, as in 'A trusts B to do x (or matters x)' (Hardin, 1998: 12). In contrast, many suggest that trust between parties can be general, stretching over the totality of their interactions with one another. Studies that treat trustworthiness as a kind of expense account that actors can use under any circumstance to convince others of their reliability fall into this category (e.g. Coleman, 1990; Fukuyama, 1995; Putnam, 1988, 2000), as do psychological studies that attempt to discern trusting from distrusting personality types (e.g. Rotter, 1980).

Theoretically, Hardin has the stronger argument. It is difficult to imagine a relationship in which one actor trusts another unconditionally. For example, doctors are often entrusted with their patients' lives, but even in

this situation relationships between doctors and patients are limited to the areas in which doctors are competent. In contrast, Hardin's formulation makes it seem that the list of issues over which A trusts B is known in advance, but in practice this is rarely the case. Consider the issue of trust in government. Citizens are unlikely to trust a government willing to use lethal force arbitrarily against its own population, but citizens might trust their government to regulate health care even though it has not done so previously. In short, there are always limits on trust, but those limits are not necessarily obvious.

Fourth, trusting others involves making predictions about their future actions (Coleman, 1990; Gambetta, 1988; Giddens, 1988; Luhmann, 1979; Seligman, 1997; Sztompka, 1999). However, the discretion or 'leeway' (Bianco, 1994) trustees receive prevents trustors from being certain about the fate of their interests, which introduces calculations of risk into the decision-making process (Coleman, 1990; Majone, 1996). This is an important point. Many students of international politics presume that trusting interstate relationships emerge when potential trustee states 'commit' themselves to particular outcomes (e.g. Kydd, 1997, 2000; Maoz and Felsenthal, 1987; Wendt, 1999). Commitments create certainty about future behavior by making it 'impossible for one of the players to choose some of the strategies' (Rapoport, 1974: 228). Typically, this means that actors in Prisoner's Dilemma-like situations undertake actions that make it impossible for them to defect in the short-run.

However, mechanisms that create certainty about a potential trustee's future behavior replace the need for trust in relationships. Actors that entrust their interests to others always run the risk of betrayal (Baier, 1986; Coleman, 1990; Luhmann, 1979; Seligman, 1997) — if there were no threat of betrayal, the 'problem of trust' (Seligman, 1997) would be no problem at all. For this reason, binding commitments are incompatible with trusting relationships because these devices make betrayal impossible.

The same may be said for expectations based on an actor's 'dependable habits' (Baier, 1986; Hardin, 2001). My belief in your trustworthiness must be based upon my assessment that you will not betray my interests, despite your capacity to do so. If you operate by rote, then the possibility of betrayal is effectively negated because you never recognize the possibility for gain at my expense. Your habitual behavior occurs whether I rely on you or not; my interests never factor into your decisions. But my belief in your trustworthiness depends upon my expectation that you will take my interests into account prior to acting (Baier, 1986; Hardin, 1998; Lieberman, 1967). Thus, statements such as, 'trust Audrey to throw a party on Saturday' are not really about trust at all (Giddens, 1988; cf Larson, 1998).

Finally, actors assess the risks of entrusting their interests to others using

subjective estimates of the probability their trust will be honored. These estimates are subject to a host of calculative errors, misperceptions and manipulations that exacerbate the dangers associated with delegating control over one's interests to others. Actors that fail to accurately assess their counterparts' reliability are more likely to have their interests betrayed.

The danger of misplaced trust can be alleviated, but not eliminated, by improving the amount and quality of information actors have about one another (Giddens, 1988; Hardin, 1993; Luhmann, 1979). Given the right circumstance (e.g. a single-shot Prisoner's Dilemma), even normally reliable actors can be untrustworthy (Dasgupta, 1988; Hardin, 2001). 'Collaboration games' (Stein, 1984), such as the Prisoner's Dilemma and Stag Hunt, raise questions about the trustworthiness of actors because each of these games rewards opportunism in the short-run (Axelrod, 1984; Oye, 1986). In contrast, 'coordination games' (Snidal, 1985) do not raise questions about an actor's trustworthiness — once actors decide how to coordinate their behavior, cooperation is self-enforcing because there are no rewards for opportunism (Krasner, 1991; Snidal, 1985). For example, selecting which side of the road to drive on is a kind of coordination game. Once established, the principle 'always drive on the left side of the road' provides rational drivers no incentive to violate the rule.

Beneath the consensus, however, lurks disagreement between those that define trust as confidence in expectations (e.g. Coleman, 1990; Sztompka, 1999) and those that define trust as confidence in expectations that others will 'do what is right' (Hollis, 1998; Lieberman, 1967). Proponents of the first definition (the predictive approach) argue that A trusts B when A predicts that B will at least do no harm in a circumstance in which A's interests depend on B's behavior (e.g. Bianco, 1994; Coleman, 1990; Dasgupta, 1988; Gambetta, 1988; Larson, 1998). For example, predictive definitions suggest that the US government's decision to rely on Syrian forces during the Gulf War to guard the Second Marine Division's left flank during the latter's push into Kuwait (Lake, 1999) required trust. On predictive accounts, entrusting one's interests to another is equivalent to gambling (Coleman, 1990; Sztompka, 1999). As Coleman notes, 'the elements confronting the potential trustor are nothing more or less than the considerations a rational actor applies in deciding whether to place a bet. The actor knows how much money may be lost, how much may be gained, and the chances of winning' (1990: 99).

Proponents of the second definition of trust (the fiduciary approach) combine the concept of obligation with the idea that trust involves risk. Trust involves more than predicting the behavior of others. It includes trustors' perceptions that their trustees have a responsibility to fulfill the trust placed in them even if it means sacrificing some of their own benefits

(Barber, 1983; Hollis, 1998; Kegley and Raymond, 1990; Lieberman, 1967). For example, Article 73 of the United Nations (UN) Charter instructs member states that have ‘assumed responsibilities for the administration of territories whose peoples have not yet attained a full measure of self governance’ to ‘recognize the principle that the interests of the inhabitants of these territories are paramount’ (the text of the UN Charter can be found at <http://www.un.org/aboutun/charter>).

Both the predictive and fiduciary approaches define trust in ways that are consistent with the received wisdom discussed above. Both highlight the predictive element of trust, both implicitly recognize that actors use subjective estimates to gauge the reliability of their counterparts, and both understand the directional nature of the trust relationship. Adherents to both approaches also recognize that the problem of trust depends on the context in which interaction takes place.

Yet the two definitions are not equally capable of distinguishing situations involving trust from those that are better described as involving suspicion, insecurity or fear. Consider the Cold War relationship between the US and USSR under the ‘balance of terror’. As long as the superpowers maintained second-strike capabilities, both predicted (with some uncertainty) that the other would not launch their nuclear weapons. Based on this expectation, the US and Soviet Union exposed themselves to attack. Although threat-based relationships are not usually described as trusting, the US–Soviet nuclear relationship fits predictive definitions of trusting relationships — decision-making under conditions of risk, exposure of interests to the possibility of harm, and the calculation that the other side will refrain from purposely harming those interests in spite of their capacity to do so. Thus, the definition admits an example it should not.

Even Russell Hardin (1993, 1998, 2001), who provides the most sophisticated predictive definition of trust, fails to provide a basis upon which to reject the conclusion that the US–Soviet nuclear relationship involved trust.⁴ Hardin argues that actors trust one another when they suspect that the interests of potential trustees ‘encapsulate’ their own interests. As Hardin explains,

I trust you with respect to some action if your reason for doing it is to take me into account in some relevant way. Typically your reason will be that it is in your interest to maintain our relationship. Hence, my trust in you is typically encapsulated in your interest in fulfilling my trust. (Hardin, 1998: 12)

However, Hardin’s explanation does not fit the US–Soviet case. The threat of retaliation, not concern for maintaining relations, restrained the US and USSR from using their nuclear weapons against one another. Both superpowers pursued strategies (e.g. arms build-up) designed to favorably

transform their nuclear relationship without tripping an armed response, suggesting that each aspired to dominate the other rather than maintain the state of rough parity. In short, even the most developed predictive approaches to trust require additional data to distinguish trust from other factors that motivate actors to take risks.

Definitions of trust that focus solely on predictability are flawed because they cannot distinguish between trust and factors, such as coercion, that also encourage actors to take risks (Hardy et al., 1998). The underlying problem for predictive definitions is that trust and trusting relationships are subsets of risk and risk-taking behavior — trust implies risk, but risk-taking does not necessarily imply trust (see Williamson, 1993). Other factors produce risk-taking behavior, but predictive definitions cannot distinguish between risks taken based on trust and risks taken for alternative reasons. Something more than a willingness to gamble on the behavior of others must characterize the trust relationship, but most predictive definitions provide few clues about what this might be.

Fiduciary approaches suggest that the concept of obligation successfully distinguishes trust from the broader category of risk.⁵ When actors place their trust in others, they do so in the belief that their trustees are obliged to fulfill their trust (Messick and Kramer, 2001). When we grant trustworthy actors discretion over our interests, we do so because we view them in a normatively positive light as ‘upright’, ‘honorable’, ‘truthful’, ‘loyal’ and ‘scrupulous’ (Holzner, 1973; Landau and Bogus, 1987) rather than strictly as ‘a good bet’. In other words, trustworthy individuals are those deemed likely to uphold the obligations our trust bestows on them.

Of course, the expectations of trustors do not automatically bind their trustees to a particular course of action. Actors may not realize that they are being entrusted with something of value, may not care to take on the role of trustee, or may deliberately take advantage of the opportunity for gain at the trustor’s expense. Trustors create trusting relationships; trustees determine the success of these relationships.

Returning to the US–Soviet example, we can now see why the nuclear relationship between the superpowers did not involve trust. Neither the US nor the Soviets made themselves vulnerable to nuclear attack based on a belief that the other had an obligation to avoid taking advantage of them. The threat of mutual destruction restrained the US and USSR from doing their worst.

Further bolstering the distinction between risk and trust is some suggestive evidence that people experience situations of trust differently than they experience situations of risk. Tyler, for example, finds that ‘trust in the motives of group authorities . . . encourages both voluntary deference to

their decisions and feelings of obligation to obey social rules' (2001: 286). Furthermore, Tyler argues that:

Although there is widespread support for the importance of expectations about the behavior of others in shaping cooperative behavior, there are also signs that this explanation is not a complete model of the psychology of trust. For example, Peter Brann and Margaret Foddy (1988) examine the cooperative behavior of people within a [collaboration game]. They find that those high in trust, as measured by Julian Rotter's (1966) Interpersonal Trust Scale — which measures general trust in others with whom one has no personal experience — feel a moral obligation to cooperate and do so irrespective of what others in their group are doing or are expected to do. (2001: 288)

Gratton (1973) also found that people did not experience trusting relationships in terms usually associated with decision-making under conditions of risk. In interviews, Gratton's subjects reported 'feeling bodily relaxation and loss of tension' (Gratton, 1973: 282) in conjunction with their decisions to enter into trusting relationships.

In contrast, the imagery policy-makers attach to situations of risk often emphasizes their sense of uncertainty. For example, Robert Schuman, the Foreign Minister of France, described the plan bearing his name that launched European integration as 'a leap in the dark' (Community, 1983: 11). More recently, after the Ulster Unionist Party voted to accept a power-sharing arrangement in Northern Ireland, David Trimble publicly challenged Sinn Fein leader Gerry Adams — 'Now Mr. Adams it's over to you. We've jumped. You follow' (*New York Times*, 28 November 1999).

An important objection to the fiduciary approach to trust is that the concept of obligation conflicts with the notion that actors are self-interested. In its most extreme rendering, this objection suggests that there is a basic incompatibility between self-interest and trusting relationships.⁶ This could mean either that the self-interested delegation of authority by trustors is incompatible with trusting relationships or that the self-interested reciprocation by trustees is incompatible with trusting relationships. Both possibilities are considered below.

In the case of trustors, why do they entrust to others that which they value if not for self-interested reasons? It is hard to conceive of an answer to this question. Trust could be an attitude borne of emotion rather than rational calculation, but there is no reason to suppose that emotional responses violate an actor's self-interest. In fact, most theories of emotion assume that they are functional responses to various stimuli (Crawford, 2000). Moreover, there is strong agreement that the decision to entrust one's interests to others is usually based on the belief that the fulfillment of that trust will make the trustor better off (e.g. Baier, 1986; Bianco, 1994; Coleman, 1990; Dasgupta, 1988; Deutsch, 1958; Dore, 1983; Fukuyama, 1995; Gulati,

1995; Hardin, 1993; Hardwig, 1991; Kydd, 1997; Larson, 1998; Lorenz, 1993; Lyons and Mehta, 1997; Sako, 1998; Swinth, 1967). This conclusion seems rather difficult to reject — actors with some foresight about the possible consequences of their actions often can predict how particular decisions will affect their well-being. Unless actors cease caring about themselves, they will inevitably prefer those strategies that satisfy, rather than undermine, their self-interest.

The case of baby-sitting is a good example for demonstrating the consistency of trust and self-interest. Parents, exhausted and at their wits-end from caring for their newborn for six months without a break, ask their neighbor's responsible teenage son to baby-sit for a couple of hours on a Saturday night. Their request is, of course, driven by their own self-interest — the parents fear losing their sanity if they do not get out of the house, an outcome they want to avoid. But are we to conclude that because these two parents are driven by a self-interested desire to see a movie they have not entrusted the care of their child to their neighbor? Clearly, this example involves trust and self-interest.

Self-interest seems like a compelling reason for actors to entrust their interests to others, but delegation and trust are not synonymous. Actors often delegate without considering the reliability of those to whom they transfer authority. For example, members of the US Congress often delegate responsibility to the President in order to avoid blame for policy failures that might hurt their re-election chances (Kiewiet and McCubbins, 1991). In IR, states often attempt to pass the buck of defense to their allies in an effort to avoid paying the costs of fighting (Christiansen and Snyder, 1990). Trust is not at issue in these cases because the reliability, i.e. trustworthiness, of the recipients is not at issue. Also, trust is not at issue when actors care little about the fate of that which they transfer to others. An example of this kind of delegation may be found in the practice of providing diplomatic posts in 'out of the way' places to political supporters. Because the costs of betrayal are nonexistent the 'problem of trust' never emerges.

Perhaps the problem is that self-interest and trustworthiness are mismatched ideas. Do we really want to call it trust when actors rely on others to do that which is in their interests? Once again, if we rule out self-interest, it is difficult to imagine another basis upon which to explain reciprocation by trustees. Most arguments that assume actors are self-interested recognize that obeying rules and norms satisfies their long-term interests (e.g. Abbott and Snidal, 2000; Axelrod, 1984; Bull, 1977; Krasner, 1983; Locke, 1980), even though rule and norm violations often bring short-run rewards. Also, we would not want to conclude that there is a basic conflict between the ideas of self-interest and obligation. Empirically, there is a good deal of

evidence that self-interested actors establish institutions that assume obligations will be upheld. For example, one of the keystones of international law, which was created by self-interested actors, is the principle of *pacta sunt servanda*, the idea that agreements create obligations for the signatories that must be fulfilled in good faith (Kegley and Raymond, 1990; Lukashak, 1989).

It appears difficult to sustain the claim that trusting relationships and the transfer of control over one's interests for self-interested reasons are incompatible. This is significant because it means that fiduciary definitions of trust, which are superior to predictive definitions, can be used without sacrificing prominent assumptions about what motivates state behavior. Although developing a definition of trust is significant, it remains to be seen if this conceptualization can be measured for the purposes of research. To this task, I now turn.

Operationalizing Trusting Relationships: Rules for Measurement

With an approach to defining trust in hand, the next step is to develop rules that permit these relationships to be identified empirically. Of all the problems facing trust scholarship, measurement issues present the most formidable barriers. We are faced with the difficulties of assessing an inherently relational phenomenon that lacks obvious material indicators signifying its presence. Given the complexity of the measurement problem, I offer only a first cut toward the goal of developing sensitive indicators of trusting interstate relationships. The operational procedures I suggest improve on traditional measurement strategies because, consistent with fiduciary approaches, they recognize that the mere presence of cooperation is an unreliable indicator of trust. Trusting relationships are special forms of cooperation involving discretion-granting policies and leaders that view one another as trustworthy.

In contrast, measurement strategies premised on predictive definitions of trust assume that trust is a precondition for cooperation (e.g. Coleman, 1990; Deutsch, 1958; Gambetta, 1988; Larson, 1998).⁷ Actors that delegate control over their interests to others run the risk that those interests will be betrayed and since predictive definitions do not distinguish between risk and trust, all efforts to cooperate must involve trust because they all involve risk.

However, as I argued above, equations of trust and risk rest on a faulty premise. Trust involves risk, but cannot be reduced to risk — trust and risk are separable constructs. The implication is that the necessary link between trust and cooperation cannot be sustained because some efforts to cooperate

emerge for no other reason than the willingness of actors to gamble on favorable or even poor odds.

The US–Soviet nuclear relationship is an example of cooperative risk taking without trust. Nevertheless, scholars using measures premised on predictive definitions argue that the US and USSR had to trust one another in order to conclude and maintain their arms control agreements (e.g. Krass, 1985; Larson, 1998). Yet, the superpowers' use of highly invasive and regular checks of each other's nuclear weapons stockpiles to monitor compliance hardly seems like the kind of behavior in which trusting partners engage. In contrast, neither the US nor Canada actively patrols their common border. If either country invaded the other, alarms would be sounded by citizens unlucky enough to be in the way, but only after the encroachment occurred. Active surveillance of the type employed by the US and USSR suggests that the superpowers were concerned cheating would occur absent this oversight.

Indicators that recognize that trust is manifest in the design of cooperation must be developed to isolate trusting relationships empirically. I suggest three measures consistent with fiduciary definitions of trust. The first involves connecting the perceptions of leaders to their policy choices. The second measure relies on information about the mechanisms states use to monitor cheating, and the third examines the restrictiveness of rules used to structure cooperation. I outline the three approaches below.

Measure One: Discretion-granting Policies and Decision-making Data

Trusting interstate relationships emerge when leaders believe their counterparts are trustworthy and, based on this perception, enact policies that make their states vulnerable to the actions of their counterparts.⁸ This definition suggests that measuring trusting relationships involves (1) identifying policies that grant other states discretion over outcomes previously controlled by the first; and (2) demonstrating that the leaders responsible for enacting such policies did so at least in part because they believed that their counterparts were trustworthy.

States enact discretion-granting policies when they transfer the capacity to determine political outcomes to others. Examples of discretion-granting policies include the 'blank check' Germany granted Austria-Hungary on the eve of World War I, and the US food aid program that, prior to the mid-1970s, permitted recipients to sell US agricultural products for local currency and use the revenues for a variety of domestic purposes. The government of South Vietnam, for instance, used the revenues it generated through the food aid program to cover some of its military expenses (Martin, 2000). The 1986 decision by European Community member states

to relinquish their vetoes in favor of a system of qualified majority voting also qualifies as a discretion-granting policy. As a consequence of this decision, the states least interested in pursuing further integration lost their capacity to define the pace and shape of future European collaboration.

Note that discretion-granting policies do not include those that simply recognize the authority of states to make decisions. Discretion-granting policies are those that grant actors control over outcomes previously controlled by others, where control is defined as the ability to make and implement decisions; arrangements that do not allow actors to both make and implement decisions on behalf of others do not count as discretion-granting. For this reason the 1949 North Atlantic Treaty would be excluded from further consideration by this initial test. Although Article V of the treaty leaves decisions about how to defend against armed attacks to individual NATO members, signatories to the agreement have long possessed this right. And whereas the NATO treaty involved no delegation of decision-making authority from one state to another, the 1950 Treaty of Paris that launched the European Coal and Steel Community (ECSC) did not provide the coal-steel pool implementing powers. Even though the High Authority appeared to have wide latitude over the governance of the ECSC (see Community, 1957), opposition to its policies by member states usually sufficed to stymie the Community's progress (Gillingham, 1991).

Once discretion-granting policies are identified, the next step is to uncover the motives that lead to those policy decisions. Consistent with the argument of this article, trusting relationships emerge when policy-makers grant others discretion based on the belief that their counterparts are trustworthy. The most direct method of gauging the subjective considerations underlying policy choices is to survey the attitudes of the decision-makers responsible for enacting them. For example, in 1787 the willingness of the American states to begin the process of amalgamation was influenced by their belief that George Washington would be a trustworthy leader (Lipset, 1963; Wood, 1998): when Washington arrived in Philadelphia for the constitutional convention he was met by 'the acclimation of the people, as well as more sober marks of the affection and veneration which continues to be felt for his character' (see Madison's letter of 15 May 1787 in Farrand, 1966: v. 3 p. 21).

The claims leaders make about their motives are an important and valuable source of evidence regarding trusting relationships, so long as care is taken when relying on political speech. For instance, public statements about the trustworthiness of others should be treated with greater skepticism than private statements because there is a greater incentive for actors to lie in public than in private. Thus, less significance should be placed on US President George W. Bush's claim during his joint news conference with

Russian President Vladimir Putin that he ‘looked (Putin) in the eye [and] found him to be very straightforward and trustworthy’ than should be placed on James Madison’s private correspondence cited above (see Khong, 1992, and Welch, 1995, for good discussions of the precautions one should take when relying on verbal and written statements as sources of evidence). Still, there are limits to what we should expect to find in the historical record. For example, leaders are unlikely to be motivated to create trusting relationships solely by their perception that others are trustworthy. Concerns about traditional political issues, such as security and prosperity, are also likely to factor into the development of trusting relationships.

Also, although some evidentiary sources may be devoid of statements by leaders attesting to their beliefs about their counterparts’ trustworthiness, the absence of such statements is not enough to indicate that policy-makers do not trust one another. Once again, the records of the US constitutional convention are instructive. During the convention, delegates made few overt references to their trust (or distrust) in one another (see Farrand, 1966). Yet, the bonds of trust between many of the convention participants are a palpable part of the historical record. As Farrand notes, ‘there is ample evidence to show that there was not a little social intercourse among the delegates’ (Farrand, 1913: 62). George Washington’s diaries, for example, reveal that he often dined with other delegates and spent his leisure time in their company (see the documents in Hutson, 1987).

Leaders may not find appropriate moments to reveal their trust in others in private or public statements. Alternatively, trust may operate as an unacknowledged background condition in the sense that leaders see no need to record their obvious faith in one another. Under these conditions, behavioral indicators of trust may be more useful than looking for direct statements. Friendships intuitively involve trust (Wieselquist et al., 1999) and evidence of these relationships can provide useful clues about underlying attitudes.⁹ And when we are unable to locate documentary evidence in which leaders proclaim their trust in one another, the evidence of friendships offers a useful substitute.

In addition to being the most direct method of identifying trusting relationships, variations on this strategy may be used to gauge the intensity of trusting relationships. For example, one might capture the intensity of trust by examining the frequency with which decision-makers refer to trust as a motivation for their policy choices. Alternatively, the number of decision-makers that express their faith in another state also may provide clues about the intensity of trusting interstate relationships. These are admittedly crude strategies, nonetheless we need some method of tapping the intensity of trusting relationships. If trust can vary from very high to

very low levels, i.e. trust is a continuous variable, then the measure of trust should be able to vary as well.

Measure Two: Oversight Indicators

The second measurement strategy, which begins with the identification of discretion-granting policies, concentrates on the mechanisms states use to oversee the exercise of that discretion. All things being equal, the oversight mechanisms actors employ can be used to assess if relationships are trusting. Methods of oversight that permit actors more leeway to make decisions are consistent with trusting relationships. I develop this argument below.¹⁰

In general, states structure the oversight of agreements in one of two ways — either oversight occurs before decisions are implemented (‘before-the-fact’ oversight) or oversight takes place after decisions are implemented and their consequences known (‘after-the-fact’ oversight).¹¹ Before-the-fact oversight tends to be more ‘centralized, active, and direct’ (McCubbins and Schwartz, 1984: 166) than after-the-fact oversight, which requires comparatively less centralization and involves less active and direct intervention by monitors than the former (McCubbins and Schwartz, 1984). The most widely discussed forms of before-the-fact and after-the-fact methods of oversight are, respectively, ‘police patrols’ and ‘fire alarms’ (McCubbins and Schwartz, 1984) — police patrol oversight involves the examination of a sample of an actor’s activities by monitors with the aim of preventing certain outcomes from obtaining. Fire alarm oversight requires affected third parties to report the consequences of decisions to monitors, who then find ways to sanction those that engage in undesirable behavior.

Examples of police patrol and fire alarm oversight are not difficult to find in world politics. The International Atomic Energy Association (IAEA) uses police patrols to sniff out efforts by non-nuclear states to develop nuclear weapons. In contrast, victims of human rights abuses sound alarms to alert the international community about their suffering. However, police patrols and fire alarms are not the only forms of before- and after-the fact oversight available to states. For example, actors sometimes employ what we might call ‘privy council’ oversight. This approach involves a group of monitors charged with the duty of evaluating and approving executive decisions before they are made and implemented. In contrast to police patrols, privy counselors do not search for possible violations. Instead, privy councilors wait for those they monitor (who are rule-bound) to seek their permission before undertaking particular actions. The requirement in the US Constitution that the Commander-in-Chief receive permission from Congress, through a declaration of war, before committing US troops to battle is an example of privy council oversight.

There are also other forms of after-the-fact oversight. Fire alarm oversight assumes the presence of monitors possessed of the capacity to punish 'bad' behavior. Yet, some after-the-fact oversight does not involve special monitors at all. Protests, for example, can be a form of oversight in which affected parties work to bring decision-making in line with popular will by threatening to remove those responsible for particular decisions from office.

Both before-the-fact and after-the-fact methods of oversight are capable of restraining action, but theorists generally assume that, given the choice, actors prefer after-the-fact oversight. Rational actors are concerned with the efficient use of their resources and after-the-fact oversight is less costly for them to implement and more effective at uncovering abuses than before-the-fact oversight. The reasons are clear — before-the-fact oversight, which must be created and paid for by the principals to an agreement, faces practical restrictions on the number of decisions it can monitor. By comparison, the costs of after-the-fact oversight are borne by affected third parties and, as long as third parties are willing to report, fire alarm oversight can catch every instance in which actors misuse their discretion. Politicians prefer this system because it enables them to take credit for preventing mischief without having to expend resources to find instances of cheating (McCubbins and Schwartz, 1984).

However, there are reasons to believe that the choice among methods of oversight is driven by more than an actor's concern about the relative costs of implementation. For example, scientists forced to help governments develop weapons, citizens that organize around the principle of nonproliferation, dissatisfied members of the military and traitors of all stripes are available to pull fire alarms to alert monitors about violations of the Nuclear Nonproliferation Treaty. Nonetheless, the signatories to this agreement prefer to use the International Atomic Energy Association to oversee compliance. We see a similar pattern with respect to domestic crime prevention. Although there are affected third parties (e.g. victims and their friends and relatives) that will sound the alarm when a crime takes place, states continue to pay for comparatively costly police patrols.

These examples suggest that actors concerned about how their counterparts will exercise their discretion prefer before-the-fact methods of oversight to after-the-fact methods. When it works, before-the-fact oversight limits discretion more effectively than after-the-fact oversight because the former denies actors opportunities to exercise their freedom of action any way they see fit. In contrast, when actors are less concerned that their counterparts will harm their interests, they opt for after-the-fact methods of oversight. This is true even when the 'costs of opportunism' (Lake, 1999) are high. For example, the costs to states that fail to defend their borders

against possible incursions from their neighbors are usually high. At a minimum, states risk losing control over portions of their territory. At worst, states risk being subjugated or, as the Melians discovered, destroyed (Thucydides, 1972). Yet, despite the apparently high costs to be suffered at the hands of opportunists, some contiguous states (e.g. the US and Canada and Denmark and Sweden) make no effort to defend their common borders. Other states conclude agreements that permit their neighbors' military to cross their shared borders freely. The US and Mexico concluded such an agreement in 1882, which permitted troops of both countries to cross their joint border in 'hot pursuit' of bandits, cattle thieves and Indian raiders (Gonzalez and Haggard, 1998).

Some will no doubt protest using oversight mechanisms to indicate trusting relationships. If actors trusted one another, they would not need to monitor their partners' behavior. The impulse to use oversight mechanisms suggests that actors are suspicious of one another's motives and sensitive to the costs of opportunism. This objection rests on the assumption that some agreements cannot be overseen. If actors are unable to structure agreements that preclude the possibility of oversight, then we cannot conclude that the presence of these mechanisms indicates suspicion because oversight would be an element of every agreement, trusting and distrusting alike.

Yet, *all* agreements contain at least the potential for overseeing and sanctioning actor performance. As long as oversight includes both systematic and unsystematic methods of uncovering violations, the behavior of actors in every agreement can be monitored. The term 'whistle blowing' refers to the most disorganized forms of oversight. Similarly, as long as systems contain more than two actors, cheaters always face potential reputation costs for violating agreements because victims can tell all who will listen about the harm they suffered. Some agreements do contain explicit clauses detailing the methods of oversight and punishment used to induce compliance and others lack these clauses. However, the distinction between explicit and implicit oversight and punishment is an unreliable guide to trusting relationships. Often, the terms of agreements are written down to enable decision-makers that did not take part in negotiations to evaluate and approve agreements. On the flip side, trusting partners are not the only actors willing or able to conclude informal arrangements. For example,

The US decision to pursue containment rather than 'rollback', even at the height of Cold War tensions, was a tacit acknowledgement of the Soviet sphere of influence in Eastern Europe . . . In a related area of unspoken agreement, each side has been careful to avoid any direct engagement of military forces, despite frequent involvement of Soviets and Americans in limited wars around the world. (Lipson, 1991: 496)

The issue is not *whether* states monitor one another's behavior, but *what kind* of oversight do states employ? The proactive nature of before-the-fact oversight carries with it greater potential to limit the discretion of actors than after-the-fact oversight. Using the principle that where there is more discretion there is more trust, after-the-fact oversight is more consistent with trusting relationships than before-the-fact oversight, *ceteris paribus*. In particular, scholars should be on the lookout for evidence of changes in the type of oversight actors employ to enforce agreements. All things being equal, shifts from before-the-fact oversight to after-the-fact oversight should indicate changing levels of trust within a relationship.

Measure Three: Rule Indicators

The third and final method of detecting trusting relationships discussed here focuses on the types of rules actors employ in written agreements (i.e. treaties). Similar to the argument made in the previous section, the intuition here is that different kinds of rules provide actors with varying degrees of decision-making freedom. All things being equal, rules that provide actors the most leeway indicate trusting relationships.

Broadly speaking, actors can opt for two kinds of agreements — framework-oriented agreements or statute-oriented agreements. *Framework-oriented* agreements are dominated by constitutive rules that specify basic structure, institutional forms, procedures and rights. These 'constitutive rules' (Dessler, 1989) create and define legitimate behavior between parties. In contrast, *statute-oriented* agreements are dominated by specific codes that regulate the behavior of actors under specific circumstances.

Articles drawn from the Rome Statute of the International Criminal Court provide useful illustrations of the distinction between constitutive and regulative rules. Article VII of the treaty, '[t]he Court shall have jurisdiction in respect of war crimes in particular when committed as part of a plan or policy or as part of a large-scale commission of such crimes' is a constitutive rule. It defines part of the court's legal jurisdiction. In contrast, Article XVI specifically regulates the conditions under which the court may exercise its jurisdiction — 'No investigation or prosecution may be commenced or proceeded with under this Statute for a period of 12 months after the Security Council, in a resolution adopted under Chapter VII of the Charter of the United Nations, has requested the Court to that effect; that request may be renewed by the Council under the same conditions' (for text of the agreement, see <http://www.un.org/law/icc/statute/rome fra.htm>).

With respect to discretion, framework-oriented agreements provide actors more latitude than statute-oriented agreements because the rules they depend on define modes of interaction without specifying when these modes

must be employed or to what end.¹² Regulative rules create specific injunctions on behavior, governing not only how actors should behave, but also when particular behavior should be undertaken (or avoided). Again, on the assumption that where there is more discretion there is more trust, framework-oriented agreements are more consistent with trusting relationships than statute-oriented agreements.

However, relying on data from treaties is controversial in trust scholarship because written agreements are thought to signal that partners are suspicious of one another (e.g. Baier, 1986; Zucker, 1986). Kegley and Raymond, for example, argue that 'alliance agreements would not need to be committed to writing if there were not a perceived need to reduce apprehensions about the reliability of allies in times of crisis' (1990: 47). If accurate, this claim would invalidate the study of treaty design to determine whether relations are trusting or not.

In IR, the claim that treaties are inconsistent with trusting relationships stems from the view that states face credibility problems in anarchic systems. Lack of third-party enforcement creates incentives for parties to renege on their agreements. These incentives make it difficult for states to convince others they are reliable (Fearon, 1995). The credibility problem is not insurmountable, however. States can signal that they will keep their promises by making commitments that are so costly to violate only those intent on honoring their pledges would make them (Fearon, 1994; Kydd, 2000; Schelling, 1966).

Written treaties count as costly commitments for two reasons. First, by undertaking public agreements, states place their reputations at risk. Parties that abandon their public agreements will be branded unreliable, making others hesitant to cooperate with them in the future. Second, by hard-wiring provisions, such as monitoring devices and penalties, into written agreements, states communicate their willingness to faithfully implement their treaty obligations by making cheating too costly to undertake (Keohane, 1984). In short, costly commitments communicate reliability through the manipulation of devices designed to deter cheating by raising its costs, and since deterrence is inconsistent with trusting relationships, treaties are inconsistent with trusting relationships (Deakin and Michie, 1997; Deakin and Wilkinson, 1998).

However, it is unclear that states create written agreements for the purpose of deterring the behavior of potentially unreliable partners. There are many ways to publicize agreements without writing them down, which suggests that writing is unnecessary in creating costly commitments. A more compelling argument for explaining the use of written agreements to structure interstate relations rests on the state's political and organizational need for records of its activities. The political process in many, if not most,

states requires agreements to be ratified by decision-makers that did not take part in the negotiations (Putnam, 1988). Writing facilitates this process by providing explicit and detailed records of the terms of an agreement that can be evaluated by those responsible for ratification.

States also need written agreements for the purpose of compliance. In large organizations, such as states, those responsible for implementing agreements often are not the people that negotiated the agreements. Written records communicate rights and responsibilities to those that are required to implement and/or to comply with the agreement. Changes in the identity of office holders, especially when parties to an agreement expect an enduring relationship, provide another rationale for recording agreements. Organizational memory, in order to be resistant to personnel change, requires records for future office holders to refer to after the principal negotiators leave their positions.

The widespread use of treaties in interstate relations does not bespeak of a world in which suspicion is rampant. Some written agreements are created with the distrust of the parties in mind, but the mere existence of treaties signals neither trust nor distrust a priori. A focus on the kinds of rules states use to structure their agreements may help to distinguish treaties that codify trusting relations from those that establish distrustful relations. Agreements that rely primarily on constitutive rules are most consistent with trusting relationships because constitutive rules give states the greatest leeway to make decisions. Of course, many agreements contain both constitutive and regulative rules and, when these are encountered, there may be a temptation to classify the agreement based upon some sort of majority-rule principle. For many questions, this strategy may be appropriate. For others, an analysis of key provisions may be the best way to judge whether agreements are framework- or statute-oriented (see Hammons, 1999, for a formal strategy designed to distinguish between framework- and statute-oriented agreements).

Perhaps the best way to use the three measures of trusting relationships suggested above is in conjunction with one another as part of a multi-measure strategy. Although each measure relies on different indicators, the three should converge in the presence of trusting relationships. In the event that the measures diverge, greater emphasis should be placed on the readings provided by the first measure (combining discretion-granting policies and decision-making data) because it is more sensitive than the other two. The downside of Measure One is that it requires data that are often unavailable for scholarly examination. When this occurs, Measures Two and Three may be used on their own, but they are more error-prone than the first and should be used with care.

Conclusion

Although we frequently refer to the problem of trust, with respect to interstate relations it would be more precise to focus on trusting relationships. Trust refers to an actor's willingness to place its interests under the control of others based on the belief that those actors will honor their obligation to avoid using their discretion in a harmful manner. Trusting relationships are a behavioral manifestation of trust. They emerge when actors leave the fate of their interests to the discretion of others with the expectation that those actors will honor their obligation to avoid using their discretion in a manner harmful to the first.

In contrast to definitions that imply trust is necessary for cooperation, the conceptualization of trust developed above suggests that trust is manifest in the form cooperation takes. Empirically, the best way to pick out instances of trusting relationships is to connect the creation of discretion-granting policies with the attitudes of the leaders responsible for passing them. When decision-making data are unavailable or incomplete, second-best strategies that focus on the manner in which states design their cooperative agreements — the kinds of oversight mechanisms states utilize and the types of rules they use to structure their agreements — are useful substitutes. Ideally, all three indicators can be used together as part of a multi-measure strategy designed to enhance the accuracy of empirical inferences.

Of course, some might suggest that this article asks the wrong questions because the problem of trusting interstate relationships is not principally about whether scholars can identify them empirically. The important question is whether trusting relationships are capable of forming in the unregulated (anarchic) environment in which states find themselves. Constant struggle for power, caused by the anarchic nature of the international system, creates incentives for states to exploit opportunities to weaken potential adversaries and improve their relative power positions. These pressures create 'little room for trust among states because a state may be unable to recover if its trust is betrayed' (Mearsheimer, 1991: 148).

Arguments that focus on the 'security dilemma' are often difficult to reconcile with the idea of trusting relationships because they emphasize the willingness of states to take advantage of one another. Very simply, opportunism destroys the expectation of trustworthiness, i.e. the belief that trustees will advance and/or protect the interests of trustors in spite of their capacity to do harm. Absent basic confidence that others will do what is right, trusting relationships cannot develop.

Yet, a good deal of work points out that theoretical conclusions about the competitiveness of international life depend upon the accuracy of the opportunism assumption. Both Glaser (1997) and Kydd (1997, 2000) find

that assuming states are ‘pure security seekers’ reduces the competitive dynamic at the heart of security dilemma politics. The empirical record also suggests that the assumption of opportunism does not always describe the behavior of states. For example, Wolfers (1962: 95) points out that, prior to 1932, Great Britain pursued the self-abnegating strategy of ‘[d]isarmament by example’ in an effort to promote world peace. Similarly, recent research suggests that, in the aftermath of major wars, victorious powers, such as Great Britain in 1815 and the US in 1919 and 1945, search for ways to restrain their own power instead of augmenting it at the expense of lesser states (Ikenberry, 1998, 2001). Rather than being ‘nasty, brutish, and short’ (Hobbes, 1988: 186) international life is ‘what states make of it’ (Wendt, 1992).

As prominent efforts to establish peaceful relations break down, we are forced to examine the idea of trust and its relationship to and role in interstate relations more closely. However, we should remain mindful of the fact that trust is not a panacea for the world’s ills. The attack of 11 September 2001 on the World Trade Center and the Pentagon in the US could not have occurred were it not for the terrorists’ ability to trust in one another’s ability to remain secretive about their plans. But we should not allow the fact that ‘[e]xploitation and conspiracy, as much as justice and fellowship, thrive better in an atmosphere of trust’ (Baier, 1986: 231) to deter us — the goal of peace is an important one and trust a central part of the processes that enable actors to end their hostilities.

Notes

Earlier versions of this article were presented at the annual meeting of the International Studies Association, Los Angeles, CA, March 2000 and the annual meeting of the Midwest Political Science Association, Chicago IL, April 2000. Special thanks to Davis Bobrow, Charles S. Gochman, Andrew Kydd, Deborah Welch Larson, Bert Rockman, Alberta Sbragia, Alex Wendt and S. Laurel Weldon for their insightful comments on this article and to Friedrich Kratochwil, Heike Brabandt and the anonymous reviewers for their thoughtful evaluations of my work. Thanks as well to Jeffrey Checkel, Joseph Haberer, Iris Young, Michael Aleprete, David Hauser and Gary Schaub Jr for their help with my larger project on trust in International Relations.

1. An analysis of the factors that cause trusting relationships to emerge is beyond the scope of this article. For a recent effort to understand this question, see Hoffman, 2001.
2. Coleman’s (1990) definition comes close in the game-theoretic (e.g. Kydd, 1997) and social capital literatures (e.g. Putnam 1988).
3. Hardin (2001) uses the word ‘entrust’ to refer to any situation in which an actor delegates control over its interests to others. I prefer to use ‘entrust’ more narrowly, to refer to those instances of delegation that are based on an actor’s belief in the trustworthiness of its counterparts. As I will argue, actors delegate

control over their interests for reasons that are unrelated to trust. Given this orientation, using the term entrust more broadly would produce conceptual confusion.

4. Hardin (2001) objected to a similar example on the grounds that coercion creates a fully deterministic setting. This claim is surely incorrect. If the use of coercive techniques always creates a deterministic outcome, the Melians would have allied with the Athenians, the Warsaw Ghetto uprising would not have occurred, and Daw Aung San Suu Kyi would not have challenged the armed forces of Myanmar. The reason the system of Mutually Assured Destruction does not fit our intuitions about trust is that trust involves more than simply taking a chance on the future behavior of others. Trusting relationships imply particular attitudes about others that distinguish it from other risky choices.
5. Coleman (1990), a proponent of the predictive approach, also suggests that trusting relationships involve expectations of obligation, but his argument fails to establish a necessary connection between these two concepts.
6. It does not make sense to argue that self-interest and trust, the attitude, are incompatible since my belief in your trustworthiness to do x may have little to do with my interest in that outcome obtaining. For example, I know that my father can be relied on to buy milk if I tell him I need it, but since I am at the grocery store I will do it myself. The question is what makes me act upon my trust and what is the relationship between those motivating factors and self-interest?
7. Necessary relationships imply that the dependent variable, Y, is always preceded by the independent variable, X, but the presence of X does not always produce Y (Most and Starr, 1989).
8. In other words, trusting interstate relationships emerge when potential trustor states entrust their interests to potential trustee states.
9. As a reviewer of this article pointed out, friendships can be strategic and therefore they do not provide definitive proof of the presence of trust between individuals.
10. The *ceteris paribus* clause is important. Actors are not always able to choose among oversight mechanisms. Some forms of oversight, for example, may be infeasible with respect to particular issues.
11. I owe the phrases 'before-the fact' and 'after-the fact' monitoring to a presentation by Margaret Levi at the 2000 American Political Science Association Convention, Washington, DC.
12. Consider the rules of chess, which are largely constitutive. Players are instructed about the object of the game (i.e. checkmate) and about how the various pieces move, but they are not told when or why to use particular pieces.

References

- Abbott, Kenneth W. and Duncan Snidal (2000) 'Hard and Soft Law in International Governance', *International Organization* 54(3): 421–56.
- Adcock, Robert, and David Collier (2000) 'Connecting Ideas with Facts: The Validity of Measurement', paper presented to the Annual Meeting of the American Political Science Association, Washington, DC.

- Axelrod, Robert (1984) *The Evolution of Cooperation*. New York: Basic Books.
- Baier, Annette (1986) 'Trust and Antitrust', *Ethics* 96(2): 231–60.
- Barber, Bernard (1983) *The Logic and Limits of Trust*. New Brunswick: Rutgers University Press.
- Bianco, William T. (1994) *Trust: Representatives and Constituents*. Ann Arbor: University of Michigan Press.
- Brenkert, George G. (1998) 'Trust, Morality, and International Business', in Christel Lane and Reinhard Bachmann (eds) *Trust Within and Between Organizations: Conceptual Issues and Empirical Applications*, pp. 273–97. New York: Oxford University Press.
- Bull, Hedley (1977) *The Anarchical Society: A Study of Order in World Politics*. London: Macmillan Education.
- Christiansen, Thomas J. and Jack Snyder (1990) 'Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity', *International Organization* 44(2): 137–68.
- Coleman, James S. (1990) *Foundations of Social Theory*. Cambridge: Cambridge University Press.
- Community, European (1957) 'Treaty Between the Federal Republic of Germany, the Kingdom of Belgium, the French Republic, the Italian Republic, the Grand Duchy of Luxembourg, and the Kingdom of the Netherlands Establishing the European Coal and Steel Community'. Belgium: United Nations Publications.
- Community, European (1983) *Opening of the Historical Archives of the European Communities to the Public*. Luxembourg: Office for Official Publications of the European Communities.
- Cook, Thomas D. and Donald T. Campbell (1979) *Quasi-Experimentation: Design and Analysis Issues for Field Settings*. Chicago: Rand McNally College Publishing Company.
- Crawford, Neta C. (2000) 'The Passion of World Politics: Propositions on Emotion and Emotional Relationships', *International Security* 24(4): 116–56.
- Dasgupta, Partha (1988) 'Trust as a Commodity', in Diego Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations*, pp. 49–76. New York: Basil Blackwell.
- Deakin, Simon and Jonathan Michie (1997) 'Contracts and Competition: An Introduction', *Cambridge Journal of Economics* 21(2): 121–5.
- Deakin, Simon and Frank Wilkinson (1998) 'Contract Law and the Economics of Interorganizational Trust', in Christel Lane and Reinhard Bachmann (eds) *Trust Within and Between Organizations: Conceptual Issues and Empirical Applications*, pp. 146–72. New York: Oxford University Press.
- Dessler, David (1989) 'What's At Stake in the Agent–Structure Debate', *International Organization* 43(3): 441–73.
- Deutsch, Karl W., Sidney A. Burrell, Robert A. Kann, Maurice Lee Jr, Martin Lichterman, Raymond E. Lindgren, Francis L. Loewenheim and Richard W. Van Wagenen (1957) *Political Community and the North Atlantic Area*. Princeton: Princeton University Press.

- Deutsch, Morton (1958) 'Trust and Suspicion', *Journal of Conflict Resolution* 2(3): 265–79.
- Dore, Ronald (1983) 'Goodwill and the Spirit of Market Capitalism', *British Journal of Sociology* 34(4): 458–82.
- Farrand, Max (1913) *The Framing of the Constitution of the United States*. New Haven: Yale University Press.
- Farrand, Max (1966) *The Records of the Federal Convention of 1787*. Vols 1–4, New Haven: Yale University Press.
- Fearon, James D. (1994) 'Domestic Political Audiences and the Escalation of International Disputes', *American Political Science Review* 90: 715–35.
- Fearon, James D. (1995) 'Rationalist Explanations for War', *International Organization* 49(3): 379–414.
- Fukuyama, Francis (1995) *Trust: Social Virtues and the Creation of Prosperity*. New York: Free Press.
- Gambetta, Diego (1988) 'Can We Trust Trust?', in Diego Gambetta (ed.) *Trust: Making and Breaking Cooperative Relations*, pp. 213–37. New York: Basil Blackwell.
- Giddens, Anthony (1988) *The Consequences of Modernity*. Stanford: Stanford University Press.
- Gillingham, John (1991) *Coal, Steel, and the Rebirth of Europe, 1945–1955: The Germans and French from Ruhr Conflict to Economic Community*. New York: Cambridge University Press.
- Glaser, Charles S. (1997) 'The Security Dilemma Revisited', *World Politics* 50: 171–201.
- Gonzalez, Guadalupe and Stephen Haggard (1998) 'The United States and Mexico: A Pluralistic Security Community?', in Emanuel Adler and Michael Barnett (eds) *Security Communities*, pp. 295–332. Cambridge: Cambridge University Press.
- Govier, Trudy (1997) *Social Trust and Human Communities*. Montreal: McGill-Queen's University Press.
- Gratton, Carolyn (1973) 'Some Aspects of the Lived Experience of Interpersonal Trust', *Humanitas* 9(3): 273–296.
- Gulati, Ranjay (1995) 'Does Familiarity Breed Trust? The Implications of Repeated Ties for Contractual Choice in Alliances', *Academy of Management Journal* 38(1): 85–112.
- Hammons, Christopher W. (1999) 'Was James Madison Wrong? Rethinking the American Preference for Short, Framework-Oriented Constitutions', *American Political Science Review* 93(4): 837–49.
- Hardin, Russell (1993) 'The Street-Level Epistemology of Trust', *Politics and Society* 21(4): 505–29.
- Hardin, Russell (1998) 'Trust in Government', in Valerie Braithwaite and Margaret Levi (eds) *Trust and Governance*, pp. 9–27. New York: Russell Sage Foundation.
- Hardin, Russell (2001) 'Conceptions and Explanations of Trust', in Karen S. Cook (ed.) *Trust in Society*, pp. 3–39. New York: Russell Sage Foundation.

- Hardwig, John (1991) 'The Role of Trust in Knowledge', *The Journal of Philosophy* 88(12): 693–708.
- Hardy, Cynthia, Nelson Phillips and Tom Lawrence (1998) 'Distinguishing Trust and Power in Interorganizational Relations: Forms and Façades of Trust', in Christel Lane and Reinhard Bachmann (eds) *Trust Within and Between Organizations: Conceptual Issues and Empirical Applications*, pp. 64–87. New York: Oxford University Press.
- Hobbes, Thomas (1988) *Leviathan*. C.B. MacPherson, trans. New York: Penguin Books.
- Hoffman, Aaron M. (2001) 'In States We Trust: A Study of the Evolution and Decline of Trust During the Formation of the United States, 1776–1860, and European Community, 1950–1986', University of Pittsburgh, Pittsburgh.
- Hollis, Martin (1998) *Trust Within Reason*. Cambridge: Cambridge University Press.
- Holzner, Burkart (1973) 'Sociological Reflections on Trust', *Humanitas* 9(3): 333–45.
- Hutson, James H. (1987) *Supplement to Max Farrand's The Records of the Federal Convention of 1787*. New Haven: Yale University Press.
- Ikenberry, G. John (1998) 'Constitutional Politics in International Relations', *European Journal of International Relations* 4(2): 147–177.
- Ikenberry, G. John (2001) *After Victory: Institutions, Strategic Restraint, and the Rebuilding of Order after Major Wars*. Princeton: Princeton University Press.
- Kegley, Charles W. Jr and Gregory A. Raymond (1990) *When Trust Breaks Down*. Columbia: University of South Carolina Press.
- Keohane, Robert O. (1984) *After Hegemony: Cooperation and Discord in the World Political Economy*. Princeton: Princeton University Press.
- Khong, Yuen Foong (1992) *Analogies at War: Korea, Munich, Dien Bien Phu, and the Vietnam Decision of 1965*. Princeton: Princeton University Press.
- Kiewiet, Roderick D. and Matthew D. McCubbins (1991) *The Logic of Delegation: Congressional Parties and the Appropriations Process*. Chicago: The University of Chicago Press.
- Krasner, Stephen D. (1983) 'Structural Causes and Regime Consequences: Regimes as Intervening Variables', in Stephen D. Krasner (ed.) *International Regimes*, pp. 1–23. Ithaca: Cornell University Press.
- Krasner, Stephen D. (1991) 'Global Communications and National Power: Life on the Pareto Frontier', *World Politics* 43(3): 336–66.
- Krass, Alan S. (1985) 'Focus On: Verification and Trust in Arms Control', *Journal of Peace Research* 22(4): 285–8.
- Kydd, Andrew (1997) 'Game Theory and the Spiral Model', *World Politics* 49(3): 371–400.
- Kydd, Andrew (2000) 'Trust, Reassurance and Cooperation', *International Organization* 54(2): 325–58.
- Lake, David A. (1999) *Entangling Relations: American Foreign Policy in its Century*. Princeton: Princeton University Press.

- Landau, Sidney I. and Ronald J. Bogus (1987) *The Doubleday Roget's Thesaurus in Dictionary Form*. New York: Doubleday.
- Larson, Deborah Welch (1998). *Anatomy of Mistrust*. Ithaca: Cornell University Press.
- Lieberman, Bernhardt (1967) 'i-Trust: A Notion of Trust in Three-Person Games and International Affairs', *Journal of Conflict Resolution* 8: 271–80.
- Lipset, Seymour Martin (1963) *The First New Nation: The United States in Historical and Comparative Perspective*. New York: Basic Books.
- Lipson, Charles (1991) 'Why are International Agreements Informal?', *International Organization* 45(4): 495–538.
- Locke, John (1980) *Second Treatise of Government*. Indianapolis: Hackett.
- Lorenz, Edward H. (1993) 'Flexible Production Systems and the Social Construction of Trust', *Politics and Society* 21(3): 307–24.
- Luhmann, Niklas (1979) *Trust and Power*. New York: John Wiley.
- Lukashak, I.I. (1989) 'The Principle Pacta Sunt Servanda and the Nature of Obligation Under International Law', *American Journal of International Law* 83(3): 513–18.
- Lyons, Bruce and Judith Mehta (1997) 'Contracts, Opportunism, and Trust: Self-Interest and Social Orientation', *Cambridge Journal of Economics* 21(2): 239–57.
- McCubbins, Mathew D. and Thomas Schwartz (1984) 'Congressional Oversight Overlooked: Police Patrols Versus Fire Alarms', *American Journal of Political Science* 28(1): 165–79.
- Majone, Giandomenico (1996) *Regulating Europe*. New York: Routledge.
- Maoz, Zeev, and Dan S. Felsenthal (1987) 'Self-Binding Commitments, the Inducement of Trust, Social Choice, and the Theory of International Cooperation', *International Studies Quarterly* 31: 177–200.
- Martin, Lisa L. (2000) *Democratic Commitments: Legislatures and International Cooperation*. Princeton: Princeton University Press.
- Mearsheimer, John J. (1991) 'Back to the Future: Instability in Europe After the Cold War', in Sean M. Lynn-Jones (ed.) *The Cold War and After: Prospects for Peace*, pp. 141–92. Cambridge, MA: MIT Press.
- Messick, David M. and Roderick M. Kramer (2001) 'Trust as a Form of Shallow Morality', in Karen S. Cook (ed.) *Trust in Society*, pp. 89–118. New York: Russell Sage Foundation.
- Most, Benjamin A. and Harvey Starr (1989) *Inquiry, Logic, and International Politics*. Columbia: University of South Carolina Press.
- Oye, Kenneth A. (1986) 'Explaining Cooperation Under Anarchy: Hypotheses and Strategies', in Kenneth A. Oye (ed.) *Cooperation Under Anarchy*, pp. 1–24. Princeton: Princeton University Press.
- Putnam, Robert D. (1988) 'Diplomacy and Domestic Politics: The Logic of Two-Level Games', *International Organization* 42: 427–60.
- Putnam, Robert D. (2000) *Bowling Alone: The Collapse and Revival of American Community*. New York: Simon and Schuster.

- Rapoport, Anatol (1974) *Fights, Games, and Debates*. Ann Arbor: University of Michigan Press.
- Rotter, Julian B. (1980) 'Interpersonal Trust, Trustworthiness, and Gullibility', *American Psychologist* 35: 1–7.
- Sako, Mari (1998) 'Does Trust Improve Business Performance?', in Christel Lane and Reinhard Bachmann (eds) *Trust Within and Between Organizations: Conceptual Issues and Empirical Applications*, pp. 88–117. New York: Oxford University Press.
- Schelling, Thomas C. (1966) *Arms and Influence*. New Haven: Yale University Press.
- Seligman, Adam B. (1997) *The Problem of Trust*. Princeton: Princeton University Press.
- Snidal, Duncan (1985) 'Coordination Versus Prisoners' Dilemma: Implications for International Cooperation and Regimes', *American Political Science Review* 79(4): 923–942.
- Snijders, Chris (1996) *Trust and Commitments*. Amsterdam: Thesis Publishers.
- Stein, Arthur A. (1984) 'Coordination and Collaboration: Regimes in an Anarchic World', in Stephen D. Krasner (ed.) *International Regimes*, pp. 115–40. Ithaca: Cornell University Press.
- Swinth, Robert L. (1967) 'The Establishment of the Trust Relationship', *Journal of Conflict Resolution*, 11(3): 335–44.
- Sztompka, Piotr (1999) *Trust: A Sociological Theory*. Cambridge: Cambridge University Press.
- Thucydides (1972) *History of the Peloponnesian War*. Rex Warner, transl. New York: Penguin Books.
- Tyler, Tom R. (2001) 'Why Do People Rely on Others? Social Identity and the Social Aspects of Trust', in Karen S. Cook (ed.) *Trust in Society*, pp. 285–306. New York: Russell Sage Foundation.
- Welch, David A. (1995) *Justice and the Genesis of War*. Cambridge: Cambridge University Press.
- Wendt, Alexander (1992) 'Anarchy is What States Make of it: The Social Construction of Power Politics', *International Organization* 46(2): 391–425.
- Wendt, Alexander (1999) *Social Theory of International Politics*. Cambridge: Cambridge University Press.
- Wieselquist, Jennifer, Caryl E. Rusbult, Craig A. Foster, and Christopher R. Agnew (1999) 'Commitment, Pro-Relationship Behavior, and Trust in Close Relationships', *Journal of Personality and Social Psychology* 77(5): 942–66.
- Williamson, Oliver E. (1993) 'Calculativeness, Trust, and Economic Organization', *Journal of Law and Economics* 36: 453–486.
- Wolfers, Arnold (1962) 'The Pole of Power and the Pole of Indifference', in *Discord and Collaboration*. Baltimore: The Johns Hopkins University Press.
- Wood, Gordon S. (1998) *The Creation of the American Republic 1776–1787*. Chapel Hill: The University of North Carolina Press.
- Zucker, Lynne G. (1986) 'Production of Trust: Institutional Sources of Economic Structure, 1840–1920', *Research in Organization Behavior* 8: 53–111.

