H-**Diplo** | **ISSF** Roundtable XI-4

Nicholas J. Wheeler. *Trusting Enemies.* New York: Oxford University Press, 2018. ISBN 9780199696475 (hardcover, $39.95).

Contents

Introduction by Robert Jervis, Columbia University

Although every negotiator I have talked to has stressed the importance of the personal relations with his or her opposite numbers, most academic theorizing ignores this dimension entirely. Nicolas Wheeler joins Marcus Holmes, whose *Face-to-face Diplomacy* will soon be reviewed on ISSF,[1] in arguing that academic research has paid a steep price for neglecting what practitioners understand. At a time of both heightened international tensions and an American president enamored of summit meetings, Wheeler's *Trusting Enemies* is especially welcome. In it, he brings together literatures that are often separate: interpersonal relations on the one hand and inter-state rapprochements on the other; how states signal on the one hand and how they perceive on the other. Having written on both of the latter topics, I can testify to the fact that they are not only usually treated separately, but that rational choice is the common approach for the former while social psychology and constructivism predominate in the latter endeavor.

Our two reviewers are well positioned to judge this ambitious project. Keren Yarhi-Milo has applied psychological theories to how states judge whether others are threats and has also examined the characteristics of leaders that makes them particularly sensitive to how others perceive them.[2] Janice Stein has co-authored a major book on the Cold War (the end of which is one of Wheeler's case studies) as well as having written several studies of decision-making and a key article on threat perception.[3]

Both reviewers find the book "important" to those concerned with "diplomacy and cooperation in world politics" (Yarhi-Milo) and, related, to those interested in "when and how leaders of enemy states accurately read signals of peaceful intentions" (Stein). Both reviewers also praise Wheeler's detailed and careful research into important cases, drawing on not only secondary sources but primary ones and interviews.

A great strength of Wheeler's book is its focus on face-to-face meetings between leaders and his argument that these can lead to what he calls "bonding trust," which he distinguishes from the more familiar form of calculating trust that is analyzed by rational choice theorists. This focus has drawbacks, however. As Yarhi-Milo notes, it cannot "present a theory or empirical evidence that solves the aggregation issue" because while national leaders may be the most important actors in their countries, they are rarely the only ones, and if rapprochements are to endure, they must take root in a wider community. Bureaucracies, both civilian and military, can play spoiler roles and, as Wheeler's own case studies show, powerful figures who are not included in the summit meetings and who therefore do not experience the personal relations that can produce trust may voice powerful dissents. Furthermore, as Stein argues, the "enabling conditions and dynamics that [Wheeler] identifies are extraordinarily demanding," starting with the requirement that the leaders be open to the possibility that they are caught in a security dilemma and willing to reframe their relations and even identities as the interaction proceeds. It is then discouraging but not surprising that only one of Wheeler's three cases results in bonded trust and agreement.

Despite these objections and limits, the reviewers find *Trusting Enemies* to be an ambitious and important project that deserves serious and sustained attention.

---

[1] Marcus Holmes, *Face-to-face Diplomacy: Social Neuroscience and International Relations* (New York: Cambridge University Press, 2018). See the H-Diplo/ISSF Roundtable Review at https://issforum.org/ISSF/PDF/ISSF-Roundtable-10-30.pdf.

[2] *Knowing the Adversary: Leaders, Intelligence, and Assessments of the Adversary's Intentions* (Princeton: Princeton University Press, 2014); *Who Fights for Reputation? The Psychology of Leaders in International Conflict* (Princeton: Princeton University Press, 2018).

[3] Richard Ned Lebow and Janice Gross Stein, *We All Lost the Cold War* (Princeton: Princeton University Press, 2001); Robert Jervis, Lebow, and Stein, *Psychology and Deterrence* (Baltimore: Johns Hopkins University Press, 1985); Stein, "Threat Perception in International Relations," in Leonie Huddy, David Sears, and Jervis, eds., *The Oxford Handbook of Political Psychology* (New York: Oxford University Press 2003), 364-394.

In his response Wheeler deeply engages with the reviewers, elaborating on some of his points, clarifying others, and illustrating the arguments and issues with a discussion of the face-to-face meetings of North Korean Supreme Leader Kim Jong-un and President Donald Trump.

Participants:

Nicholas J. Wheeler is Professor of International Relations and Director of the Institute for Conflict, Cooperation, and Security at the University of Birmingham. His publications include (with Ken Booth) *The Security Dilemma: Fear, Cooperation, and Trust in World Politics* (Palgrave Macmillan, 2008) and *Saving Strangers: Humanitarian Intervention in International Society* (Oxford University Press, 2000). He co-edits with Professors Christian Reus-Smit and Evelyn Goh the *Cambridge Series in International Relations*.

Robert Jervis is Adlai E. Stevenson Professor of International Politics at Columbia University and Founding Editor of ISSF. His most recent book is *How Statesmen Think* (Princeton University Press, 2017). He was President of the American Political Science Association in 2000-2001 and is the founding editor of the International Security Studies Forum. He has received career achievement awards from the International Society of Political Psychology and ISA's Security Studies Section, the Grawemeyer Award for the book with the Best Ideas for Improving World Order, and the National Academy of Science's tri-annual award for behavioral sciences contributions to avoiding nuclear war.

Janice Gross Stein is the Belzberg Professor of Conflict Management in the Department of Political Science and the Founding Director of the Munk School of Global Affairs and Public Policy at the University of Toronto. Her most recent publications are "The Micro-Foundations of International Relations: Psychology and Behavioral Economics," in *International Organization* 71 (2017) and "Loss Avoidance and Negotiation Outcomes: Understanding the End Game," forthcoming 2019. She is an Honorary Foreign Member of the American Academy of Arts and Science.

Keren Yarhi-Milo is a Professor of Political Science and International Affairs at Columbia University's Political Science Department and the School of International and Public Affairs. She is also the Associate Director of the Saltzman Institute of War and Peace Studies. Professor Yarhi-Milo's first book (Princeton University Press, 2014) titled *Knowing the Adversary: Leaders, Intelligence Organizations, and Assessments of Intentions in International Relations*, received the 2016 Furnnis Award for best book in the field of international security. Also, it is Co-Winner of the 2016 Best Book Prize from the Diplomatic Studies Section of the International Studies Association. Her new book, *Who Fights for Reputation? The Psychology of Leaders in International Conflict* (Princeton University Press, 2018) won the best book award on foreign policy from the American Political Science Association. Professor Yarhi-Milo's articles have been published or are forthcoming in *International Studies Quarterly*, *International Organization*, *International Security*, *Journal of Conflict Resolution*, *British Journal of Political Science*, and *Security Studies*. She holds a Ph.D. and a Master's degree from the University of Pennsylvania, and a B.A., summa cum laude, in Political Science from Columbia University.

Review by Janice Gross Stein, University of Toronto

I n *Trusting Enemies* Nicholas Wheeler tackles a wicked problem in international relations theory and practice. He asks when and how a leader of a state is able to read accurately a signal by the leader of an enemy state that she is open to cooperation.

Why is this such a hard problem? Robert Jervis's pioneering work on the application of cognitive psychology to international relations provides a clear answer.[1] We are all biased cognitively to maintain our images. Once an image of an 'enemy' is established and embedded, we will discount information that is inconsistent with that information. We are overwhelmingly likely to explain away a signal of peaceful intent by a leader we have classified as an enemy. The bias is to conservatism and inertia. A leader of an 'enemy' state faces big obstacles when signalling a new interest in cooperation.

And yet, at times, these unexpected 'signals' do get through the screen of bias. Wheeler reviews a broad range of explanations of when and how these signals get through and finds them all incomplete. A large literature on "costly signalling," for example, argues that signals that impose clear costs on the sender, either through overt behavior or the imposition of domestic costs, are more likely to be believed because they are costly.[2] Wheeler is not convinced. He argues, as have others, that the costliness of a signal is in the eye of the beholder.[3] Whether or not a signal is perceived as costly and therefore credible is in the head of the receiver, not in the hands of the signaller.

Wheeler proposes to explain the accurate perception of a change of intent from hostile to peaceful by shifting the level of analysis. He finds both the state and individual level of analysis limiting, and shifts to the analysis of the interpersonal relationship between state leaders. He brings a theory of trust into conversation with signalling theory that, he claims, both explains and predicts when signals of peaceful intent will be accurately interpreted.

What is the theory of trust? Drawing from theories of trust in intimate relationships, he argues that trust in international relations is at the core of an interpersonal relationship that enables accurate interpretation of signals of peaceful intent. The most important 'mechanism' that explains the emergence of trust between leaders is a process of interpersonal bonding, made possible through face-to-face interaction. The bonding process changes their understanding of their own and the other's identity and interests through a process of positive and mutual identification. Face-to-face interaction, Wheeler argues, generates vivid non-verbal as well as verbal cues about someone's trustworthiness and intentions.

Because trust is about the future of a once-adversarial relationship, it is inherently uncertain. Betrayal is always possible. Leaders can develop an expectation that they will not be harmed based on calculation of the risks involved. Wheeler acknowledges what he calls "calculative trust" by rational egoists (4-5), but makes clear that his focus is on a much deeper transformation of identities and interests that is beyond calculation and emerges from interpersonal bonding. Here Wheeler joins his argument with some of the constructivist approaches to interpersonal politics. Two leaders who have bonded place a high value on the other's security and care about their adversary's security as well as their own. They display a sensitivity to

---

[1] Robert Jervis, *The Logic of Images in International Relations* (Princeton: Princeton University Press, 1970); Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976); and Jervis, "Signaling and Perception: Projecting Images and Drawing Inferences," in Jervis, *How Statesmen Think: The Psychology of International Politics* (Princeton: Princeton University Press, 2017), 107-125.

[2] James Fearon, "Domestic Political Audiences and the Escalation of International Disputes," *American Political Science Review* 88:3 (1994): 577-592; and Fearon, "Signaling Foreign Policy Interests: Tying hands versus Sinking Costs," *Journal of Conflict Resolution* 41:1 (1997): 68-90.

[3] Jervis, "Signaling and Perception," 115 and Keren Yarhi-Milo, "Tying Hands behind Closed Doors: The Logic and Practice of Secret Reassurance," *Security Studies* 22:3 (2013):405-435.

the dynamics of the security dilemma. When is this kind of interpersonal bonding likely to develop? Wheeler identifies two critical preconditions: a positive identification of interests, and humanization.

Wheeler illustrates his arguments in three richly detailed case studies of the United States and the Soviet Union in the days of Soviet leader Mikhail Gorbachev with presidents Ronald Reagan and then George H. W. Bush, India and Pakistan in the days of prime ministers Atal Bihari Vajpayee and Mian Muhammad Nawaz Sharif, and a shorter look at U.S.-Iranian relations from 2008-2009, in the early days of the first Obama administration. These cases are infused with archival material where it is available, discursive material, and interviews. This thick description is essential for Wheeler to persuade us that interpersonal bonding did occur and that leaders "inhabit a mental space of trust as suspension [of calculation]," a deeper level of trust than that experienced by rational egoists (8).

In the first case, interpersonal bonding and the trust that ensued permitted each leader to interpret the signals of the other accurately and to end the Cold War. In the second case Vajpayee and Sharif, as a result of their face-to-face meeting in Lahore, came to trust each other's intentions but their trust was sabotaged by Pakistan's military leaders who were spoilers. President Barack Obama reached out to Supreme Leader Ayatollah Ali Khamenei early in his presidency but there was no face-to-face interaction, and, Wheeler implies, therefore no interpersonal bonding and trust.

How persuasive are Wheeler's arguments and evidence? The originality of this book lies in the central question that it asks: what makes it possible for a leader to read the unexpected signals of peaceful intent by a leader of an enemy state accurately. The last several decades of research have focused on the dynamics of misperception. Wheeler turns the problem on its head.

To answer this question, Wheeler narrows the analytical focus and then narrows it again and again as he develops his argument. The set of enabling conditions and dynamics that he identifies are extraordinarily demanding – a sensitivity to the security dilemma, humanization of the other, face-to-face interaction, interpersonal bonding, and a reframing of identities and interests, which together create a state of trust devoid of calculation of risk. Although he does not specifically address this issue, the argument suggests that the collapse of even the weakest link in the logic chain would imperil the whole process. It is not surprising, then, that Wheeler finds that only in the Soviet-American case did this kind of trust develop and result in an agreement. In the other two, trust was created but sabotaged by actors outside Wheeler's analysis or it was not created at all. His own analysis, therefore, suggests a *de minimus* relevance to international politics. The argument may be elegant but largely irrelevant.

This progressive narrowing of the argument so that its relevance is tightly constricted comes at a considerable cost. Scholars of international politics are generally not interested in trust for its own sake, but as an enabler of conflict reduction that is stable over time. If what we wish to explain is conflict reduction, then some of the state-level variables that Wheeler excludes in a focus exclusively on interpersonal dynamics become material at the back end of the argument. As Keren Yarhi-Milo tells us, militaries and bureaucracies matter because they tend to read signals differently than do political leaders. [4] Leaders at the top generally do not manage conflict by themselves.

At the front end of the argument, Wheeler acknowledges but largely excludes trust that allows for calculation of risk by rational egoists. That too narrows the focus of his argument and, by extension, an analysis of that kind of trust on forging better relationships among adversaries.

Two cases illustrate the costs of restricting the argument so tightly. Wheeler excluded from his analysis Egypt's President Anwar Sadat's visit to Israel to signal peaceful intent because it did not lead to a process of interpersonal bonding with Israel's Prime Minister Menachem Begin. He is correct that it did not lead to the bonded trust that overpowers calculus of risk. The two leaders disliked one another intensely, in large part because of differences in temperament, culture, and style.

---

[4] Keren Yarhi-Milo, *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations* (Princeton: Princeton University Press, 2014).

What Sadat's visit to Israel did do, however, was provoke an accurate reading of his intentions by most of Israel's senior leaders. This suggests that bonded trust is not the only route to the accurate reading of an adversary's signals.

That accurate reading enabled Israel's leaders to reframe their calculation of risk, so that a withdrawal from the Sinai in exchange for recognition and permanent agreement became an active option for the first time. Two years of hard bargaining followed, but that bargaining would have been inconceivable without the jarring impact on the mental frames deeply embedded by decades of reinforcement during the bitter conflict. The agreement that they reached, with extensive help from the United States, has endured over three decades despite multiple challenges. The solidity of the agreement over time has deepened to the point where the leaders of both countries now work actively together on shared security problems. Stable conflict reduction gradually grew out of an agreement produced by flinty-eyed calculus of risk, with very limited face-to-face interaction, but it took decades.

A second case is also instructive, although for very different reasons. The extraordinary face-to-face meetings between U.S. President Donald Trump and North Korea's Supreme Leader Kim Jong-un are almost unprecedented in the history of summitry. Generally summits are carefully prepared and only held when officials are reasonably confident that leaders can come to some kind of agreement and avoid the damaging optics of failure. Here, both leaders went around that process and trusted in face-to-face meetings to generate the kind of interpersonal dynamics that could break the logjam.

Almost every kind of process has been tried over the last twenty years to persuade leaders of the Democratic People's Republic of Korea to refrain from developing and then to give up their nuclear weapons. Before he became president, Trump was warned by his predecessor that North Korea was the most dangerous threat that he would face. After an initial experiment with coercive diplomacy, the president concluded after some back channel exchanges that, almost exactly as Wheeler argues, a face-to-face meeting could generate the trust between the two leaders that would lead to a reframing of their interests. Trump spoke repeatedly and publicly about his peaceful intentions and the enormous economic opportunities awaiting North Korea if it gave up its nuclear weapons. Kim Jong-un in turn spoke of the disarming of the entire Korean peninsula and signalled through the president of the Republic of Korea his deep interest in an accommodation.

This process can fairly be described as one where both presidents displayed sensitivity to the security dilemma and humanized one another. They met twice in face-to-face meetings, first in Singapore and then in Hanoi, and expressed their confidence in one another. This is at least an approximation of the early stages of the process that Wheeler outlines in *Trusting Enemies*.

It was not enough. Their second meeting ended in failure and both presidents left disappointed. A few months later, Kim Jong-un tested a short-range missile to signal to Trump that he would not wait indefinitely. The morning the short-range missile was fired, the president tweeted:

"Anything in this very interesting world is possible, but I believe that Kim Jong Un fully realizes the great economic potential of North Korea, & will do nothing to interfere or end it. He also knows that I am with him & does not want to break his promise to me. Deal will happen!"[5]

The words are remarkably personal and attest to the president's belief that a relationship of trust had developed. It is possible that the interpersonal dynamics built sufficient insulation into the relationship that makes it resilient enough to survive the failure to reach a substantive agreement. The final chapter has not yet been written. What the story tells us thus

---

[5] Donald J. Trump, Twitter post, 4 May 2019, 6:42 AM, https://twitter.com/realdonaldtrump/status/1124670603179565056.

far is that a trusting personal relationship is not yet enough to overcome the disagreements over substance and eliminate the calculation of risk.

These two cases are mirror images of one another. The first tells us that interpersonal dynamics were not enough to create trust but did tilt the framing enough to change calculations of risk. That change enabled the two leaders and their advisors to begin a difficult process of bargaining over substantive issues that produced a resilient agreement over time. The second suggests that interpersonal dynamics generated bonded trust that was, in and of itself, not sufficient to produce substantive agreement. To be fair to Wheeler, he is not trying to explain conflict reduction or stable agreement in this book, but I suspect that his interest in bonded trust is a step along the way to an explanation of enduring agreement.

Wheeler makes a major contribution to an important question: when and how leaders of enemy states accurately read signals of peaceful intentions. His answer to that question needs to be inserted into the larger literature on the determinants of enduring agreements among enemy states. There are multiple pathways to agreement, some independent of one another but some cumulative. It is no coincidence that scholars frequently find that an outcome is 'over-determined,' and that more than one pathway leads to a similar outcome. It is precisely because reaching agreement among enemy states is such a complex problem that multiple pathways converge and cumulate.

Wheeler's answer goes to the deepest interpersonal dynamics that occur between enemy leaders. But they are not the only dynamics. There is a tendency among scholars to locate explanations at only one level of analysis and to claim that explanations developed at other levels are flawed. A more productive approach might be to look for complementarities and cumulation across levels. Understanding how bureaucracies and militaries perceive enemy intentions matters when they have vetoes over agreements or large scope in implementation. Understanding domestic constituencies helps to delimit the bargaining space. And understanding the substantive issues at stake—their centrality, their vividness to leaders and publics—helps to understand the obstacles that even leaders who are sensitive to the security dilemma, open to interpersonal diplomacy, and accurately read their adversary's signals of peaceful intent still have to overcome.

Review by Keren Yarhi-Milo, Columbia University

*T*rusting Enemies is an important book for all students of diplomacy and cooperation in world politics. Wheeler's book offers a significant addition to the growing body of literature on the role of face-to-face diplomacy by scholars such as Marcus Holmes, Todd Hall, and me, among others.[1] *Trusting Enemies* is also in conversation with a broader literature on how enemies credibly communicate intentions, build trust, and enforce cooperation with each other. The book is impressive not only because of the novel theoretical claim it advances, but also because its empirical chapters are rich and carefully studied.

At the heart of Wheeler's book is a theory of bonding trust. Bonding trust is achieved when two state leaders who have peaceful intent accurately interpret each other's signals of peaceful intent as a result of the relationship that has formed between them. Put differently, the ability of leaders to accurately perceive signals of peaceful intent depends on a prior process of identity transformation between them. The mechanism by which leaders achieve this bonding trust, according to Wheeler, is interpersonal relationships. Social interactions between leaders in face-to-face meetings therefore have the potential of transforming the identities of leaders, leading them to a mental state of mind in which they place a high value on the other's security and care about that person's well-being as an end in itself. Importantly, in Wheeler's account, is that while the interpersonal component is necessary for the development of bonding trust, face-to-face interactions might prove insufficient to producing bonding trust among leaders. In such cases, Wheeler agues, leaders would not completely trust each other and their behavior will be more akin to calculated risk analysis.

Wheeler's account of the emergence of bonding trust remains in the dyadic level throughout the book, but he allows the unit of analysis, at least theoretically, to shift from one of interpersonal trust between leaders to trust between communities. This is discussed in chapter 5 of the book, titled "From Interpersonal Trust to Security Communities." This chapter could perhaps have been the more innovative part of the theory given that most existing accounts of face-to-face interactions put the individual leaders at the center of the analysis, leaving open the important question of aggregation. Wheeler's claims that his theory of the sources and dynamics of interpersonal trust is necessary in order to understand how trust can spread to form security communities not just between leaders, but also between elites and even between the people of those states. Nevertheless, despite the attempt in this chapter to think about this aggregation problem, the theory and empirical work remain firmly grounded at the dyadic-interaction level between leaders, and not beyond. Thus, in the end, Wheeler's book does not present a theory or empirical evidence that solve the aggregation issue. This is a common limitation of existing work on this topic, and one that offers an important and promising venue for future research.

Trusting Enemies offers an important contribution to the IR literature on diplomacy not because it argues that interpersonal interactions between leaders matter. Indeed, my own work in *Knowing the Adversary* as well as Marcus Holmes's recent book *Face-to-Face Diplomacy* both showed that these type of interactions are crucial to how leaders read each other's intentions. *Trusting Enemies* is innovative because it adds the concept of trust, specifically bonding trust, to our theories of

---

[1] Keren Yarhi-Milo, *Knowing the Adversary: Leaders, Intelligence, and Assessment of Intentions in International Relations* (Princeton: Princeton University Press, 2014); Todd H. Hall, *Emotional Diplomacy: Official Emotion on the International Stage* (Ithaca: Cornell University Press, 2015); Marcus Holmes, *Face-to-Face Diplomacy: Social Neuroscience and International Relations* (New York: Cambridge University Press, 2018); Todd Hall and Keren Yarhi-Milo, "The Personal Touch: Leaders' Impressions, Costly Signaling, and Assessments of Sincerity in International Affairs." *International Studies Quarterly* 56:3 (2012): 560-573; Marcus Holmes and Keren Yarhi-Milo, "The Psychological Logic of Peace Summits: How Empathy Shapes Outcomes of Diplomatic Negotiations," *International Studies Quarterly* 61:1 (2016): 107-122. See also Deepak Nair, "Saving Face in Diplomacy: A Political Sociology of Face-to-Face Interactions in the Association of Southeast Asian Nations," *European Journal of International Relations* (January 2019); Seanon S. Wong, "Emotions and the communication of intentions in face-to-face diplomacy," *European Journal of International Relations* 21:1 (2015): 144-167.

why and when leaders are able to read peaceful intentions during face-to-face interactions. Thus, Wheeler's works departs from my own in that his key mechanism is the process by which bonding trust leads to identity transformation.

This begs the important question of when, or under what conditions, face-to-face interaction between leaders leads to relationships of bonded trust. The answer offered by Wheeler is an important one. He argues that there are two preconditions for a process of bonding trust to emerge: First, leaders should possess security dilemma sensibilities, meaning that they allow for the possibility that the other side is acting out of fear and insecurity. Second, the two leaders acquire an index of each other's trustworthiness through face-to-face interaction. If both conditions are present, then the process of bonding trust can begin.

Reflecting on those conditions one is left with several significant questions. First, should security dilemma sensibilities be considered as qualities or traits of leaders/individuals, and if so, are those related to dispositions such as emphatic capacity? Or are security dilemma sensibilities something that leaders acquire with experience, and if so, should we overall observe the emergence of bonding trust to occur between dyads of experienced leaders, compared to inexperienced ones? *Trusting Enemies* does not theorize or test propositions concerning the micro-foundation of security dilemma sensibilities, but I believe that future research can use the insights offered here to explore this question further.

Reflecting on the second precondition for the development of bonding trust—that both leaders acquire an index of trustworthiness of each other—also raises several important questions. For example, I was unclear whether this is a precondition or a result of bonding trust. On page 272, Wheeler notes that "it is relationships of bonded trust and not calculative trust that are causally *prior* to the accurate interpretation of signals." If so, then acquiring an index of trustworthiness cannot possibly be a *pre*condition for bonding trust.

The theory put forward in this book is compelling. But perhaps the theory's greatest strength is also its greatest weakness. Indeed, an interactionist theory of the sort *Trusting Enemies* offers is powerful because of its ability to explain processes and mechanisms of trust building and cooperation. As such it offers strong and convincing accounts of past successful (and unsuccessful) interactions between leaders. But, at the same time, the theory is limited in its predictive power, and is not designed to generate systematic expectations of when or under what conditions face-to-face interactions between leaders will result in bonding trust. The theory also cannot explain how that bonding trust travels from leaders to their communities.

Given the nature of the theory, these limitations are understandable. Notwithstanding them, *Trusting Enemies* significantly advances our understanding of the *process* by which leaders come to trust each other, and the fundamental implications of bonding trust on the transformation of leaders' identities and interests. I also believe that there are many insights in this book—especially those pertaining to individual micro-foundational assumptions—on which future scholarship could build, both theoretically and empirically.

Response by Nicholas J. Wheeler, University of Birmingham

I'm very grateful to Janice Gross Stein and Keren Yarhi-Milo for their thought-provoking and highly engaging reviews of my book, *Trusting Enemies: Interpersonal Relationships in International Conflict.* I also want to thank Robert Jervis for organising the forum and writing such a positive introduction. Jervis's work has been a towering influence on me ever since I first read *Perception and Misperception in International Politics* as an undergraduate, and his ideas permeate the pages of *Trusting Enemies*.[10] Stein's and Yarhi-Milo's work has also been highly influential, and it is a privilege to have such eminent scholars engage so deeply with one's work. Each of them raises challenging questions, some of which, as I discuss below, I am taking up in current research. But as Jervis notes in his Introduction, both "reviewers find this an ambitious and important project that deserves serious and sustained attention."

I cannot hope to respond to the richness of all the points made in the reviews, but will restrict my comments to the following three key themes and issues: (i) the narrowness of the theory of "bonding trust"; (ii) the components of a process of interpersonal bonding; and (iii) the problem of aggregation.

The Narrowness of the Theory of Bonding Trust

Stein considers that *Trusting Enemies* "makes a major contribution to an important question: when and how leaders of enemy states accurately read signals of peaceful intentions." However, she finds the theory too narrow and limiting in that the conditions I identify are "extraordinarily demanding." Given the demands of my theory, it is not surprising, as she and Jervis point out, that only one of the three cases in the book—the U.S.-Soviet case—results in a "relationship of bonded trust." This is the point at which risk is suspended in the interactions of two individual leaders, making accurate signal interpretation possible. Stein identifies two cases where, she argues, the theory of bonding trust falls short. The first is Egyptian President Anwar Sadat's visit to Israel in November 1977 which she views as an example where an accurate reading of an adversary's intentions was achieved in the absence of a relationship of bonded trust. The second is the summitry between President Donald Trump and his North Korean counterpart Kim Jong-un. Here, she argues that Trump and Kim have developed a relationship of bonded trust, but this has not been sufficient to achieve an agreement over North Korea's nuclear weapons. I argue below that the theory of bonding trust is less limited in explaining each case than Stein suggests.

Starting with the agreement that was achieved over the Sinai, Stein claims that Sadat's move was a costly signal that communicated his trustworthiness to Israeli decision-makers. As a result, they began to actively explore withdrawal from the Sinai as a way of securing a long-term peace with Egypt. If ever there was a dramatic "costly" signal of peaceful intent, it was Sadat's decision to travel to the capital of his country's sworn enemy and publicly recognize the state of Israel before the Knesset. However, the Israeli Prime Minister, Menachem Begin, did not interpret Sadat's visit as evidence of his peaceful intent. I discuss the case in Chapter 1 of my book to show the limits of costly signalling theory and to drive home my core thesis that a relationship of bonded trust is prior to accurate signal interpretation. As Stein appreciates, it took U.S. mediation in the form of President Jimmy Carter's Camp David peace process to achieve a peace treaty between Israel and Egypt. Sadat's historic and courageous decision to go to Jerusalem was a key enabling factor in Carter's intervention, but this did not grow out of an accurate reading of Sadat's intentions on the part of Begin.

Stein implies that it is an either-or choice between theories of "calculative trust" and "bonding trust." It is true that I counterpoise a theory of bonding trust to that of calculative trust in the book. But I also argue that trust which emerges out of a process of social bonding can co-exist alongside continuing risk calculations. It is only when actors enter a mental or psychological state of trust as suspension which becomes possible through a full-blown transformation of identity and

---

[10] Robert Jervis, *Perception and Misperception in International Politics* (Princeton: Princeton University Press, 1976).

interests that actors cease calculating risks in relation to the intent and integrity of the individual with whom they have bonded.

A good example of a mix of bonding and risk calculation at work in an interpersonal relationship is President George H.W. Bush and Soviet Premier Mikhail Gorbachev's personal interactions that I discuss in Chapter 6 of *Trusting Enemies*. The interpersonal dynamics between these two leaders did not reach the point of suspension; instead, the process of bonding between the two state leaders, and the trust that emerged out of it, was not incompatible with continuing risk calculation about the other's future trustworthiness (intent and integrity). Bush left the 1989 Malta summit with a positive view of his relationship with Gorbachev, but he continued, in the months following the summit, to build probabilistic risk calculations into his decisions regarding the future intent and integrity of the Soviet leader. One important illustration of this was Bush's growing concerns about whether he could trust Gorbachev when he said that there would be no violence to end the movements for political change in the Baltic republics in the early months of 1990. Consequently, the theory I develop in the book can accommodate interpersonal interactions between decision-makers that vary from purely egoistic calculations at one end to the suspension of risk calculations at the other.

The second case that Stein uses to show the lack of generalizability of my theory is the interpersonal dynamics between Trump and Kim. Stein argues that the case shows that a relationship of bonded trust between the two leaders is not enough to reach an agreement, but the reply here is that it is too soon to make such a definitive assessment of (i) the nature of the bond and (ii) the effect of any interpersonal bond between Trump and Kim on the conflict. I argue in *Trusting Enemies* that there are two observable implications of a theory of bonding trust. First, the presence of discursive materials that evidence a process of bonding and second, the absence of risk calculation between the two leaders who have developed a bond. Stein focuses on the discursive materials, and as she points out, there is evidence of a bonding process. Trump claimed at his June 2018 Singapore summit with Kim that they had developed a "very special bond"[11] and later in an interview with George Stephanopoulos of ABC news he said, "I do trust him."[12] Kim has also, according to South Korean officials, expressed his "unwavering trust for President Trump."[13] These claims deserve to be taken seriously in the absence of countervailing evidence, but the difficulty is that there is no other data available (e.g. memcoms, other classified materials, memoirs, personal letters, interview material etc.) to triangulate against the verbal utterances of Trump and Kim.

Turning to the second observable implication of the theory, namely, the absence of risk calculation between Trump and Kim, it is even harder to know what is really going on. The evidence does not exist to say authoritatively whether Kim and Trump have suspended risk calculations in relation to each other. What can be said is that the U.S. and North Korean governments are continuing to build probalistic risk calculations into their military contingency planning. But state behaviour of this kind is not incompatible with a theory of bonding trust. Instead, what would show the limits of the theory in this case is evidence that Trump or Kim had not suspended risk calculations in relation to the intent and integrity of the other.

---

[11] Kevin Liptak, "Trump's North Korean gamble ends with 'special bond' with Kim," *CNNpolitics*, https://edition.cnn.com/2018/06/11/politics/trump-kim-summit-singapore/index.html, 12 June 2018, last accessed 24 July 2019.

[12] Meghan Keneally, Bill Hutchinson and Karma Allen, "'I do trust him.' Trump opens up about Kim after historic summit," *abcnews*, https://abcnews.go.com/US/exclusive-trust-trump-opens-kim-historic-summit/story?id=55815265, 12 June 2018, last accessed 24 July 2019.

[13] Yoonjung Seo & James Griffiths, J. North Korea's Kim has 'unwavering trust' in Trump, South Korea says. *CNN*. Retrieved from: https://edition.cnn.com/2018/09/06/asia/north-south-korea-trump-kim-moon-intl/index.html, 6 September 2018, last accessed 22 July 2019.

One intriguing possibility raised by the case, and potentially an important extension of the theory, is that this is an example of what Marcus Holmes and I call "fake bonding."[14] We explain this as a situation where one or both leaders deliberately fake a bonding process. A number of possibilities present themselves in relation to this case. First, Trump or Kim, or both of them, maybe faking a process of bonding in the hope of extracting a better deal from the other. If both leaders believe that this is what the other is doing, and continue to participate in the sham, then neither will gain any advantage from this attempted orchestration of the bond. More worryingly is the possibility that one or both leaders are faking the bond, but the other is not aware of this, with the aim of lulling the other into a false sense of security that can then be ruthlessly exploited. How far, for example, is Kim faking his admiration of Trump, and the sense of personal chemistry between them, in order to lull the U.S. president into believing that Kim is sincerely committed to ending North Korea's nuclear programme, despite the absence of concrete and verifiable assurances, in the belief this will lead Trump to make concessions that Kim can then exploit? A third possibility is that both leaders are faking the bond, with each believing that the other knows this, and over a period of time, the fake bond will turn into a genuine bond.[15]

If both leaders have, or do in the future, enter a mental state of trust as suspension, then the theory claims that this makes possible the de-escalation and ending of conflict at the interstate level. Stein is right that there is little sign of this so far, but as she recognizes, the ending of the story has not yet been written. Nevertheless, as I discuss below in relation to the concept of security dilemma sensibility (SDS), there is some evidence to support the claim that the interpersonal dynamics between Trump and Kim has led to a limited process of de-escalation.

The Components of a Process of Interpersonal Bonding

Ken Booth and I define SDS as "an actor's intention and capacity to perceive the motives behind, and to show responsiveness towards, the potential complexity of the military intentions of others. In particular, it refers to the ability to understand the role that fear might play in their attitudes and behaviour, including, crucially, the role that one's actions may play in provoking that fear."[16] I argue that SDS is a key precondition for a process of interpersonal bonding and that face-to-face diplomacy is a critical modality of interpersonal interaction for testing out the intuition that another state's behaviour is driven by fear and insecurity, and crucially, that one's own state's actions have played a key role in creating this. In a conference paper with Marcus Holmes we advance previous discussions of SDS by conceiving it as a variable that changes depending on an actor's "intention and capacity" to exercise SDS (previous studies have posited a binary where actors either possessed or did not possess SDS) and the extent to which the other party possesses, or is open through new forms of interpersonal interaction, to developing and exercising SDS.[17] As Holmes explains it in his response to a recent ISSF Forum on his book, *Face-to-Face Diplomacy: Social Neuroscience and International Relations,* "SDS is a variable and leaders will vary in the extent to which they possess it (what we call "inward SDS") and believe that the other side possesses it ("outward SDS")."[18]

---

[14] Nicholas J. Wheeler and Marcus Holmes, "Fake bonding? Will Trump and Putin's Helsinki meeting replay Singapore?" https://blog.bham.ac.uk/socialsciencesbirmingham/2018/07/16/helsinki-summit/, 16 July 2018, last accessed 22 July 2019.

[15] I owe this insight to Marcus Holmes.

[16] Ken Booth & Nicholas J. Wheeler, *The Security Dilemma: Fear, Cooperation and Trust in World Politics* (Palgrave Macmillan, 2008), 7.

[17] Nicholas J. Wheeler and Marcus Holmes, "Overcoming the Four Horsemen of Diplomacy: Explaining Variation in Face-to-Face Engagement," paper presented to a workshop on Trust Research, Namur, Belgium, 17 January 2019.

[18] Author's response by Marcus Holmes, H-Diplo/ISSF Roundtable 10-30 on *Face-to-Face Diplomacy: Social Neuroscience and International Relations,"* https://issforum.org/roundtables/10-30-neuroscience, 1 July 2019, last accessed 22 July 2019.

Stein argues that both Trump and Kim have "displayed sensitivity to the security dilemma." Again, the evidence is limited here, but a plausible example of Trump exercising SDS is his decision after the Singapore summit to cancel the annual U.S.-South Korean military exercises that have been viewed by successive North Korean leaders as evidence of the United States' malign intent. In justifying this move which the U.S. president took without consulting South Korea, Japan, or even the Pentagon, Trump declared that he was suspending the exercises because they are "very provocative," echoing terms used by North Korea itself, adding that his decision would save the United States "a tremendous amount of money."[19] If this is evidence that Trump possesses a capacity for inward SDS, then Kim's decision to refrain from further nuclear and long-range ballistic missile testing could be cited as evidence of Kim's capacity for inward SDS. Exercising inward and outward SDS can be critical to de-escalating conflicts when both sides have what Andrew Kydd calls "Assurance Game" preferences (that is, they will reciprocate and not exploit cooperative moves).[20] But it becomes a seriously misguided, and potentially dangerous strategy to pursue, in cases where another state leader has malign intent. Face-to-face interaction is a key modality of interaction by which decision-makers can test out an intuition of SDS, but there are no guarantees that such intuitions are correct ones.

Trump and Kim met for a second summit in Hanoi in February 2019, but this ended in failure. Patience is running out in Pyongyang at the lack of progress and Trump's apparent exercise of inward SDS in cancelling the exercises in 2018 has not particularly reassured Kim about U.S. intentions.[21] One worrying indicator here from a North Korean point of view and a clear retreat from Trump's earlier commitment is that the United States and South Korea have continued to operate joint exercises, albeit not on the scale of previous exercises. The North Koreans have perceived even these drills as evidence of U.S. bad faith. On May 9, North Korea broke its missile testing moratorium and launched two short-range missiles. Asked whether this constituted a "breach of trust," Trump said, "I don't consider that a breach of trust at all. And, you know, at some point I may. But at this point no."[22] Kim's decision to test short-range missiles provides valuable data to North Korean scientists and technicians in developing a nuclear capability that can hit the United States. Thus, it is evident that despite the warm personal interactions between Trump and Kim, U.S. anxieties about the nuclear threat from the Democratic People's Republic of Korea (DPRK) have not been significantly abated.

---

[19] David S. Cloud, "Trump's decision to halt military exercises with South Korea leaves Pentagon and allies nervous," *Los Angeles Times*, 12 June 2018.

[20] Andrew H. Kydd, *Trust and Mistrust in International Relations* (Princeton: Princeton University Press, 2005).

[21] At his summit meeting on 26 April 2019 with Russian leader, Vladimir Putin, Kim Jong-un was reported by the Korean Central News Agency (KCNA) as saying in relation to the breakdown of talks, the situation "may return to its original state as the U.S. took an unilateral attitude in bad faith," "North Korea summit: Kim accuses US of 'bad faith,'" *BBC News*, https://www.bbc.co.uk/news/world-asia-48062071, 26 April 2019, last accessed 26 July 2019. Writing in 1967, Ole Holsti developed the concept of an "'inherent bad faith' model" in relation to adversarial relationships. Actors who operate with such a mindset explain negative behaviour on the part of an adversary in terms of inherent traits whereas any positive actions are discounted as explainable in terms of situational pressures. The result is to create a negative psychological bias towards an adversary (D. Finlay, O. Holsti, and R. Fagen, *Enemies in Politics* (Chicago: Rand McNally, 1967), 26. It is a symptom of a "bad faith model" that decision-makers will attempt to maintain cognitive consistency and avoid what psychologists call 'cognitive dissonance' (L. Festinger, A Theory of Cognitive Dissonance (Stanford: Stanford University, 1957). See also Jervis, *Perception and Misperception in International Politics*, 382-406. As a result, they are likely to reject conciliatory signals as a trick or a sign of weakness (see Janice Gross Stein, "Image, Identity, and Conflict Resolution," in Chester A. Crocker, Fen Osler Hampson, and Pamela Aall, eds., *Managing Global Chaos: Sources of and Responses to International Conflict* (Washington, D.C.: United States Institute of Peace Press, 1996), 93, 98; Deborah Welch. Larson, *Anatomy of Mistrust, US–Soviet Relations during the Cold War* (Ithaca: Cornell University Press), 22.

[22] Jerome Cartillier, "North Koreas's missile tests not 'breach of trust': Trump," Agence-France-Presse, https://news.abs-cbn.com/overseas/05/12/19/north-koreas-missile-tests-not-breach-of-trust-trump, last accessed 20 July 2019.

The question of how far Trump and Kim are exercising inward and outward SDS in the U.S.-North Korean conflict, and accurately interpreting the other's exercise of SDS, raises wider issues that Yarhi-Milo draws attention to regarding the "micro-foundation of security dilemma sensibilities." She invites future research around the question as to whether the possession of SDS on the part of leaders depends on "dispositions such as empathic capacity" or whether SDS is something that leaders develop with experience. Existing research has identified factors that inhibit the exercise of SDS, crucially Ole Holsti's "bad faith model" of the adversary and what has been called "peaceful/defensive self-images."[23] What explains why some leaders are able to escape these barriers to possessing and exercising SDS when others are imprisoned by them, however, remains undertheorized.

Holmes and I are developing a theory that seeks to explain why some individuals are predisposed towards the exercise of inward/outward SDS. We focus on three individual-level traits: (i) generalized vs. particularized trust; (ii) high vs. low self-monitoring; and (iii) high vs. low reliance on intuition.[24] The difference between generalized trustors and particularized trustors is that the latter confine their trust to narrowly defined moral communities whereas generalized trustors are open to the possibility that those beyond their own particular moral community might be capable of trusting behaviour.[25] Self-monitoring is a personality trait that refers to the ability to monitor expressive and affective behaviours in social situations. High self-monitors care about what others think of them and look for ways to understand and empathize with how others perceive them. In terms of SDS, high self-monitors will be more attuned to how an adversary perceives them (outward SDS) and crucially, how their specific actions are perceived by the other (inward SDS). By contrast, low self-monitors are less worried how they are perceived by the other and are less likely to develop inward SDS, for their concern is not focused on how their actions will be perceived. The third trait explaining variation in inward and outward SDS is a leader's reliance on intuition. Those who rely heavily on intuition are more likely to act, despite the lack of cognitive beliefs and evidence, whereas low-reliance intuitional thinkers are more likely to eschew their instincts in favor of conscious, reasoned decision-making. This variation is of critical importance for taking action based on inward and outward SDS because sometimes leaders have little more than intuition to go on. A good example of the importance of a leader's intuition in making possible a diplomatic breakthrough is President Ronald Reagan's reflection in his memoir that he had an intuition that if he could only get "in a room alone and try to convince him [a Soviet leader] we had no designs on the Soviet Union and Russians had nothing to fear from us"[26] then it might be possible "to break down the barriers of mistrust that divided our countries."[27]

In our paper, Holmes and I test these propositions in relation to individual decisions to pursue face-to-face summitry, arguing that different levels of inward and outward SDS explains the variation in why some leaders pursue this diplomatic option when others eschew it. Our illustrative cases are President John F. Kennedy and Soviet Premier Nikita Khrushchev's decisions to meet in 1961 at the Vienna summit, Reagan and Gorbachev's decisions to engage in 1985 in Geneva, and to provide variation on our dependent variable, the lack of interpersonal face-to-face diplomacy between President Barack

---

[23] Finlay, Holsti, and Fagen, *Enemies in Politics*, 26. Booth and Wheeler, *The Security Dilemma*; see also Jervis, *Perception and Misperception*, Robert Jervis, "Cooperation Under the Security Dilemma', *World Politics* 30: 2 (1978): 167-214.

[24] Wheeler and Holmes, "Overcoming the Four Horsemen of Diplomacy."

[25] The concepts of generalized and particularized trust were originally developed by Eric Uslaner in *The Moral Foundations of Trust* (New York: Cambridge University Press, 2002). Brian Rathbun has applied the concepts to International Relations (IR). See *Trust in International Cooperation: International Security Institutions, Domestic Politics and American Multilateralism* (Cambridge: Cambridge University Press, 2012).

[26] Ronald Reagan, *An American Life: The Autobiography* (New York: Simon and Schuster, 1990), 589; see also 11, 588-589, 603, and 636.

[27] Reagan, *An American Life*, 12.

Obama and Kim Jong-il in Obama's first term.[28] Future research, as Yarhi-Milo suggests, could test for differing levels of diplomatic experience in predisposing leaders to inward and outward SDS. Hitherto, gender has been excluded from discussions of SDS, but this is an omission, along with the question of racialization, that also needs to be included in theorizing the micro-foundations of SDS.

The second precondition for a process of social bonding is what I call an "index of trustworthiness." Jervis's developed the concept of the index in *The Logic of Images in International Relations*[29] which, in turn, draws on earlier work by the micro-sociologist Erving Goffman.[30] Jervis explains the index as statements or actions that carry some inherent evidence that the image projected is correct because they are believed to be inextricably linked to the actor's capabilities or intentions."[31] I build on Jervis's work, fusing the concept of the index with that of trustworthiness. I use the index of trustworthiness to denote signals of inherent credibility that are received in face-to-face interaction as to the trustworthiness of others. Yarhi-Milo pushes me on the question as to whether the index is "a precondition or a result of bonding trust." The answer in the book is that the index is a precondition of a process of interpersonal bonding, but the ambiguity in the text that she rightly draws attention to revolves around my claim that "it is relationships of bonded trust and not calculative trust that are causally *prior* to the accurate interpretation of signals." By signals here I mean a receiver's interpretation of another leader's action at the state level (such as Sadat's visit to Jerusalem) that is intended by the sender to convey peaceful intent. I do not mean the "vivid indicators"[32] that are exchanged in face-to-face interaction and which give actors an index of the other's trustworthiness. Bonding trust, and the suspension of risk this makes possible, requires the two preconditions of SDS and the index.

Trusting Enemies identifies the components—SDS—index of trustworthiness—positive identification of interests—and humanization—that are necessary for trust as suspension to emerge in face-to-face interaction. What it does not do, as Yarhi-Milo points out, is generate a predictive theory of "when or under what conditions face-to-face interactions between leaders will result in bonding trust." Put differently, I explain the impact of social bond formation in terms of the emergence of trust, but not how it is that these conditions become possible in some face-to-face interactions, but not others. Sometimes this variation in social bonding extends to the same dyadic pairing as in my case study of the interpersonal dynamics between the Indian leader, Prime Minister Atal Bihari Vajpayee and his Pakistani counterpart Prime Minister Nawaz Sharif. In this case, social bonding did not take place at their first two formal meetings, but it did happen at their third encounter, a private luncheon held during the UN General Assembly in September 1998.

Answering the elusive question of why personal chemistry exists between some leaders and not others is the subject of my next book with Holmes. Our point of departure in developing a predictive theory of social bond formation is micro-

---

[28] Wheeler and Holmes, "Overcoming the Four Horsemen of Diplomacy."

[29] Robert Jervis, *The Logic of Images in International Relations* (Princeton: Princeton University Press, 1989). First published in 1970.

[30] Erving Goffman, *The Presentation of Self in Everyday Life* (London: Penguin Books, 1959), Erving Goffman, Strategic Interaction (University Park: University of Pennsylvania Press, 1967).

[31] Jervis, *The Logic of Images in International Relations*, 18. See also Todd Hall and Keren Yarhi-Milo, 'The Personal Touch: Leaders' Impressions, Costly Signaling, and Assessments of Sincerity in International Affairs," *International Studies Quarterly* 56:3 (2012): 560-573; Keren Yarhi-Milo, *Knowing The Adversary*: *Leaders, Intelligence, And Assessment of Intentions in International Relations* (Princeton: Princeton University Press, 2014); Holmes, *Face-to-Face Diplomacy*.

[32] Yarhi-Milo, *Knowing The Adversary*, 101.

sociology, and in particular the work of Randall Collins on interaction ritual chains.[33] In our forthcoming article in *International Theory*, we develop a model of social bond formation in interpersonal diplomatic interaction that develops from Collins's four ingredients of social bond formation—"bodily co-presence," "barriers to outsiders," "mutual focus of attention," and "shared mood." We highlight the importance of the interactional as against the dispositional traits of the actors and the material context in which the interaction takes place.[34] Collins argues that in the absence of these conditions, a positive social bond will not form, and that this may lead to individuals in dyadic interaction experiencing negative emotional energy. By contrast, when the conditions are present, individuals experience positive emotional energy and that shared feeling of personal chemistry. Collins argues that one of the effects of the bond that forms when individuals experience positive emotional connection is trust, which he explains in terms of "moral solidarity."[35] This is an exciting research project and next steps are to operationalize the model of diplomatic social bonding in specific cases and identify the observable implications of the causal mechanism of "collective effervescence." This idea refers to the simultaneity of thought and action in social interaction and was originally

theorised by Émile Durkheim.[36] Collins develops Durkheim's idea and explains collective effervescence as the mechanism by which the four ingredients result in a positive social bond.

Identifying observable and testable measures for the presence and absence of collective effervescence is critical to identifying when positive social bond formation takes place in face-to-face interaction.

The Problem of Aggregation

Both reviewers highlight the problem of aggregation as a limitation of the book, but each identifies different dimensions of the problem. For Stein, the problem is that I focus too narrowly on the interpersonal dimension of state behaviour, which leads me to exclude other actors in the decision-making process. My justification for developing the interpersonal as a discrete unit of analysis, distinct from IR's long-standing cognitivist focus on individual values and beliefs, is that interpersonal trust has been missing, or at best marginalized, in existing accounts of how state leaders interpret signals that are sincerely aimed at communicating peaceful intent. *Trusting Enemies* develops a new social interactionist account of trust emergence which has critical policy implications for how we think about ending interstate and intrastate conflicts.

Nevertheless, I share Stein's commitment to developing research that operates at multiple levels of analysis. I tried in the theoretical and empirical parts of the book to recognize how different actors within the domestic decision-making process constrained and enabled the possibilities for interpersonal trusting relationships to develop. For example, I show in the case study of prime ministers Vajpayee and Sharif how the Pakistani army chief, General Musharraf acted as a veto-player in relation to the personal trusting relationship that developed between the two leaders, culminating in Musharraf's decision to destroy the fledgling Lahore peace process between the two enemies by attacking India at Kargil.

Trusting Enemies introduces the concept of "trust as capacity vulnerability." I argue that "capacity"—the ability to deliver on promises and commitments—is a key dimension of an actor's trustworthiness. Capacity vulnerability has two key dimensions. The first refers to state leaders who trust in another leader's intent and integrity (the other two dimensions of

---

[33] Randall Collins, *Interaction Ritual Chains* (Princeton: Princeton University Press, 2004).

[34] Marcus Holmes and Nicholas J. Wheeler, "Social Bonding in Diplomacy," forthcoming *International Theory*.

[35] Collins, *Interaction Ritual Chains*, xv; see also David Boyns and Sarah Luery, "Negative Emotional Energy: A Theory of the 'Dark-Side' of Interaction Ritual Chains," *Social Sciences* 4:1 (2015): 148-170.

[36] Émile Durkheim, *The Elementary Forms of Religious Life* (Oxford: Oxford University Press, 2001, translated by Carol Cosman with an Introduction by Mark S. Cladis). First published 1912.

trustworthiness) but who do not trust in that leader's capacity to deliver on their promises. The flip side of capacity vulnerability is what occurred between Vajpayee and Sharif. Two leaders form a relationship of bonded trust, but one of the leaders, in this case the Pakistani Prime Minister Sharif, is unable to convince the Pakistani military—the veto-player - that they should trust the other leader (Vajpayee).

Yarhi-Milo's criticism is not that I ignore the aggregation problem, but that my attempt to scale up from the interpersonal to "trust between communities" fails because it does not go beyond the "dyadic-interaction level between leaders." My purpose in Chapter 5, which she argues had the potential to be the most innovative contribution of the theory, is to show the deficiency of existing causal accounts of the role of trust in security community formation, and the importance of supplementing these with an understanding of the dynamics of interpersonal trust in order to better explain how one-time enemies can become trusting partners. The fundamental problem that both Stein and Yarhi-Milo highlight is that if leader-to-leader bonded trust cannot be enlarged to encompass other actors in the bureaucracy, and the wider community, then how far can the theory travel? *Trusting Enemies* might, in Stein's words, go "to the deepest interpersonal dynamics," but if the theory is restricted to studying particular dyadic relationships, the aggregation problem will remain unsolved.

Solving the aggregation problem is the holy grail in IR theorizing, and social science theorizing more generally. It is certainly a challenge that Holmes and I are taking up from a micro-sociological perspective in our new book on social bonding. A theory that can explain and predict social bond formation from the dyadic to the inter-societal levels, and provide the empirics to support its claims, would be a major advance on the narrower, albeit important, trust-based theory of accurate signal interpretation developed in *Trusting Enemies*.