

# H-Diplo | ISSF

Roundtable, Volume IX, No. 13 (2017)

[issforum.org](http://issforum.org)

H-Diplo/ISSF Editors: **Thomas Maddux and Diane Labrosse**

H-Diplo/ISSF Roundtable and Web/Production Editor: **George Fujii**

---

**Eric Grynaviski. *Constructive Illusions: Misperceiving the Origins of International Cooperation*.** Ithaca: Cornell University Press, 2014. ISBN: 978-0-8014-5206-2 (hardcover, \$39.95).

Published on **6 March 2017**

Shortlink: [tiny.cc/ISSF-Roundtable-9-13](http://tiny.cc/ISSF-Roundtable-9-13)

Permalink: <http://issforum.org/roundtables/9-13-constructive-illusions>

PDF URL: <http://issforum.org/ISSF/PDF/ISSF-Roundtable-9-13.pdf>

## Contents

---

Introduction by Deborah Larson, University of California, Los Angeles .....	2
Review by Arjun Chowdhury, University of British Columbia .....	4
Review by Steven Ward, Cornell University .....	8
Author's Response by Eric Grynaviski, George Washington University .....	15

## Introduction by Deborah Larson, University of California, Los Angeles

---

The study of perception and misperception in international relations is often driven by the implicit belief that understanding the source of biased judgments will lead to mutual understanding and better relations between states. But what if cooperation between two states is based on mutual misperception? Is collaboration dependent on positive illusions?

These questions have not received much attention in international relations theory.<sup>1</sup> Eric Grynawiski's *Constructive Illusions* argues that U.S.-Soviet détente in the 1970s would not have been established if not for incorrect beliefs about the other state's goals and intentions. Whereas constructivists emphasize intersubjective beliefs and understandings, Grynawiski proposes that False Intersubjective Beliefs (FIBs) may also be quite important. He defines a FIB as an erroneous belief that a belief is shared. (29) In the book, he applies theory to several cases—the 1972 Basic Principles Agreement, the 1973 Arab-Israeli ceasefire, the Anti-Ballistic Missile (ABM) Treaty, and the decline of détente.

While pointing out that the focus of the book on détente in the 1970s is relatively narrow, both reviewers regard the book as having made a theoretical and conceptual contribution in developing the concept of FIBs and showing how mutual misperception could contribute to collaboration.

Nevertheless, the reviewers question whether the book lives up to some of the claims made by the author. Steven Ward points out that constructivists do not argue that intersubjective beliefs lead to cooperation, but merely that they lead to regularized patterns and practices. Consequently, the existence of superpower cooperation despite the absence of shared beliefs does not undermine the constructivist position. Both reviewers are not persuaded that cooperation based on mutual misperception can be sustained, especially once the illusions have been revealed. Arjun Chowdhury suggests that mutual misperception is likely to promote cooperation under a fairly restrictive set of conditions: “when decision-makers prefer cooperation for different reasons, and when shared knowledge of these reasons would impede cooperation.” If this is the case, then false intersubjective beliefs would have no impact if two states opposed cooperation.

Despite these reservations, the reviewers judge that Grynawiski has written a challenging and sophisticated analysis of how misperception can be a necessary condition for policy cooperation and that his book has highlighted the relationship between misperception and cooperation as an area warranting further investigation.

### Participants:

**Eric Grynawiski** is an Associate Professor of Political Science and International Affairs at George Washington University. He is the author of *Constructive Illusions* (Cornell, 2014), winner of APSA's 2015 Jervis-Schroeder best book award in International History and Politics. His research on sociological approaches to international cooperation has appear in the *European Journal of International Relations*, *International Organization*, *International Theory*, and *Security Studies*.

---

<sup>1</sup> For an exception, see Arthur A. Stein, “When Misperception Matters,” *World Politics* 34:2 (1982): 505-526.

**Deborah Welch Larson** is professor of political science at the University of California, Los Angeles. She received her Ph.D. in political science from Stanford University. Her publications include *Origins of Containment: A Psychological Explanation* (Princeton: Princeton University Press, 1985); *Anatomy of Mistrust: US-Soviet Relations during the Cold War* (Ithaca: Cornell University Press, 1997); and “Status Seekers: Chinese and Russian Responses to U.S. Primacy,” *International Security* 34, no. 4 (Spring 2010): 63-95 (with Alexei Shevchenko). She is currently doing research with Alexei Shevchenko on the role of status-seeking in the foreign policy of China and Russia, using social identity theory.

**Arjun Chowdhury** (PhD. Minnesota) is an Assistant Professor of Political Science at the University of British Columbia. His book project explains why the modern nation-state, despite its manifest failure to monopolize violence and provide public goods in most of the world, persists as the central unit of world politics.

**Steven Ward** is an Assistant Professor of Government at Cornell University. He received his Ph.D. from Georgetown University in 2012. He is working on his first book manuscript, “Thwarted Ambitions: Status and the Origins of Radical Revisionism in Rising Powers,” and has previously published related research in *Security Studies*.

Review by Arjun Chowdhury, University of British Columbia

---

Barbara Tuchman's seminal book on the Great War, *The Guns of August*, identified the European powers' misperceptions of each other's intentions and misunderstanding of their own military strategy as the cause of a war that was much longer and deadlier than its protagonists desired. During the Cuban Missile Crisis, President John Kennedy drew on Tuchman's thesis, telling his advisors: "the great danger and risk in all of this is a miscalculation, a mistake in judgment."<sup>1</sup> Kennedy's fear epitomizes a seemingly obvious intuition underpinning explanations for why states cooperate, which is the target of Eric Grynawski's *Necessary Illusions*. The intuition: as conflict is often driven by misunderstanding and disagreement, cooperation must indicate the absence of misunderstanding and disagreement (12). In logical terms, this runs close to the fallacy of denying the antecedent. However, the concepts used to explain cooperation—intersubjectivity in the constructivist literature or common knowledge in rationalist analyses—implicitly evince this logic (17-28), and sometimes do so explicitly (12). The premise that cooperation must stem from shared understanding has practical implications in the design of international regimes, namely that these be set up to share information and reduce the possibility of misunderstanding (28). Indeed, the Hot Line between Washington and Moscow was set up precisely for this reason.

To this commonsensical understanding, Grynawski provocatively counters: what if important instances of cooperation occurred not just despite mutual misunderstanding, but because of it? If so, shared understanding is not a necessary condition for cooperation. *Détente*, an intense period of diplomacy in the 1970's between the United States and the Soviet Union where the two signed important arms agreements, among other deals, is Grynawski's case material for suggesting that cooperation can derive from misunderstanding. *Détente* is an important case because it defies two alternative explanations—social-psychological and materialist—that either would not predict cooperation at all, or would not predict the end of cooperation when it actually occurred. Throughout the Cold War, the superpowers did not trust each other, leading to numerous missed opportunities where both might have preferred cooperation but could not trust the other to abide by the agreement.<sup>2</sup> Distrust thus could not explain cooperation in the 1970's, as top decision-makers on either side continued to not trust each other. The materialist explanation, that as the military balance became equal in 1970's both sides had incentives to cooperate, does predict the beginning of *détente*, but, as the military balance remained the same into the 1980's, cooperation should have continued rather than ended, which is what happened (142-147).<sup>3</sup> Neither distrust, which remained constant, nor a change in the military balance can explain why the superpowers cooperated in the 1970's, and not in the 1960's or the 1980's. This puzzle allows Grynawski to zoom in on the beliefs of decision-makers in the 1970's to see if cooperation was the

---

<sup>1</sup> Quoted in Ronald Carpenter, *Rhetoric in Martial Deliberations and Decision-Making: Cases and Consequences* (Columbia: University of South Carolina Press, 2004), 85.

<sup>2</sup> Deborah Welch Larson, *Anatomy of Mistrust: U.S.-Soviet Relations during the Cold War* (Ithaca: Cornell University Press, 1997).

<sup>3</sup> The alternative materialist expectation, that cooperation is more likely when there is an imbalance of power because the weaker side should compromise, is also not supported by *détente*. If this were true, the superpowers should have cooperated before 1970, and not cooperated in the 1970's and 1980's.

## H-Diplo/ISSF Roundtable 9-13

result of shared beliefs. The historical analysis supports his argument in the main, but qualifies it in important respects.

In three chapters, Grynaviski takes us from the beginning of détente to the Anti-Ballistic Missile (ABM) Treaty, to the decline of détente in the late 1970's. In these chapters, he identifies key moments where mutual misunderstanding facilitated cooperation, also, to test his claims, addressing the counterfactual question as to whether cooperation would have occurred if the decision-makers had correctly understood each other's positions to test. Rather than misperception and misunderstanding writ large, Grynaviski focuses on a particular instance of misunderstanding, a False Intersubjective Belief (FIB; pun, I assume, intended). A FIB is a situation where both sides believe an idea to be shared, even when it is not (29). Under this logic, if state A with belief X wishes to cooperate with state B with belief Y, cooperation will occur when A believes that B shares the belief X, and B believes that A shares the belief Y. This mutual misunderstanding is a FIB. If A and B correctly perceive that B believes Y and A believes X, respectively, cooperation will not occur. It is less clear what occurs if the misunderstanding is one-sided: if A correctly perceives that B does not share its belief X, but B (mis)perceives that A shares its belief Y. Here A may have incentives to misrepresent its own beliefs if it prefers cooperation, even without shared understanding, to a situation where B share its beliefs (36). In this sense, one-sided misperception as well as mutual misperception can facilitate peace or cooperation, as anticipated by, among others, Robert Jervis.<sup>4</sup>

At the beginning of détente, the U.S. and USSR recognized that military capabilities were roughly in balance, and thus cooperation was necessary, but each believed the other to share its own understanding of why cooperation was necessary (55). Soviet decision-makers believed the U.S. was cooperating because it finally accepted the USSR as a political equal. This (false) belief, which was crucial to their willingness to cooperate, led to two expectations of U.S. behavior: that the U.S. would not link issues in negotiation as it had in the past (for example, make arms limitations conditional on Soviet policy in Vietnam), and that the U.S. would not actively impede the USSR's effort to expand its influence in the Third World. But U.S. decision-makers, primarily President Richard Nixon and Secretary of State Henry Kissinger, had diametrically different beliefs on all three of these dimensions, and incorrectly assumed the Soviets shared their beliefs.

Nixon and Kissinger did not come to détente with the view that military parity made the USSR a political equal, as the Soviets saw it. Rather, military parity compelled the U.S. to pursue foreign policy through more creative, non-military means (60-62). For the U.S., cooperation was not the result of mutual recognition between equals, but an effort to maintain political superiority through cooperation. Thus, U.S. decision-makers did not expect to alter their prior policy of linking issues in negotiations, such as tying arms control, which the Soviets wanted but wished to negotiate in isolation, to trade deals. Indeed, they expected to continue these linkages to manage the rise of Soviet power. Nor did they expect to alter their policy of limiting Soviet influence and weakening Soviet allies in the Third World, and continued efforts to destabilize Left regimes in southern Africa and elsewhere.

These FIB's—for the Soviets, that military parity meant the U.S. understood it had to negotiate with the Soviets as equals; for the U.S., that military parity meant the Soviets understood that negotiations were necessary despite the U.S. not seeing them as equals—were the necessary conditions for détente (44, 50). If the U.S. and USSR had correctly perceived each other's quite different beliefs – that is, in the absence of FIB's—

---

<sup>4</sup> Robert Jervis, "War and Misperception," *Journal of Interdisciplinary History* 18:4 (1988): 675-700.

détente would not have occurred. At a more fine-grained level, several instances of cooperation, such as the Moscow Summit of 1972 where the ABM treaty was signed, would not have occurred if both sides shared the same understandings. Given that détente progressed through a fortuitous series of coinciding misunderstandings, it stands to reason that cooperation based on FIB's is fragile: at any moment, FIB's might be cleared and cooperation cease. Grynaviski does not pursue the question of how enduring cooperation based on mutual misunderstanding can be, relative to its alternative, cooperation based on mutual understanding. Logically, one might expect the latter to be more enduring, but, equally, beliefs can change. This might ultimately be an empirical question of which occurs more quickly: the clearing up of mutual misunderstandings, or changes to beliefs.

Given that détente originated in incorrect beliefs that both sides agreed on why each cooperated, it should have ended when these incorrect beliefs were falsified by the interactions between decision-makers. This was not how détente ended, and analyzing the end of détente should, in my view, qualify Grynaviski's theoretical claims because it introduces a hitherto undiscussed variable: the preferences of decision-makers. Détente ended because after Nixon's impeachment, hawkish Republicans and Democrats critical of the USSR ascended to power. Unlike Nixon and Kissinger, these decision-makers preferred a more adversarial relationship with the USSR, whether because they suspected the USSR was cheating on arms agreements, like former Director of Policy Planning for the State Department (and principal author of NSC-68), Paul Nitze and the Team B that was tasked with evaluating CIA assessments of Soviet capability (133-135), or because they condemned the USSR's human rights record, like President Jimmy Carter (135-140). However, none of these policymakers had any more accurate assessments of Soviet intentions than Nixon and Kissinger did; they just interpreted Soviet behavior to be more aggressive than it actually was. Détente ended because a set of policymakers ascended to power who had no interest in cooperation, not because FIB's were falsified.

The historical analysis suggests that Grynaviski's theoretical claims are more conditional than presented early in the book, and consequently the book does not, despite its many insights, present a comprehensive theory of the effects of mutual misunderstanding on interstate relations. Early on, Grynaviski suggests that FIB's are necessary for cooperation, but he also acknowledges that in some cases, FIB's and other misperceptions can lead to conflict (50). This suggests that FIB's facilitate cooperation under a more restricted set of conditions, which the historical analysis clarifies. The conditional, and I would say more precise, argument is that when decision-makers prefer cooperation for different reasons, and when shared knowledge of these reasons would impede cooperation, FIB's are necessary for cooperation to occur. This qualification might suggest, *contra* Grynaviski, that FIB's are epiphenomenal, because if decision-makers prefer cooperation, it may not matter that they misperceive each other's beliefs. This would be incorrect: even if decision-makers prefer cooperation, they are doing so under the belief that the other side is cooperating for the same reasons they are. If they correctly perceive that the other side's motivations are different, they will cease to cooperate. FIB's thus have causal impact beyond the preferences of decision-makers, as Grynaviski persuasively argues.

These scope conditions mean that FIB's are unlikely to lead to cooperation in scenarios where decision-makers are opposed to cooperation. But might they amplify discord in these instances? This seems a reasonable expectation—states might mutually overestimate each other's aggressive intentions by projecting their own assessments on the other side—but the book does not venture into this territory. Consequently, *Constructive Illusions* is not a comprehensive theory of how mutual misunderstanding can facilitate cooperation or conflict. However, it should force scholars to approach instances of cooperation more critically, to ask two questions. First, was cooperation the product of mutual understanding or mutual

## H-Diplo/ISSF Roundtable 9-13

misunderstanding? Second, and consequently, how enduring (or not) is a particular instance of cooperation likely to be?

Review by Steven Ward, Cornell University

---

Eric Gyrnaviski's *Constructive Illusions* is an important, compelling effort to unsettle conventional wisdom about the role of misperception in international politics. Most approaches to misperception have treated the phenomenon as a cause of conflict. Security dilemmas operate because actors misperceive each other's true intentions (or, at least, cannot be sure that they have perceived them accurately); inaccurate beliefs about military technology and doctrine have been blamed for both of the twentieth century's world wars; according to the bargaining model of war, incorrect estimates of an adversary's valuation of what is at stake and the cost of fighting lead to armed conflict; and, according to neoliberal institutionalists, the inability of states in anarchy to credibly share information about their preferences and intentions is one reason that prisoners' dilemma-like situations often result in cooperation failure.<sup>1</sup> One of the core implications of several strands of International Relations (IR) scholarship, in short, is that if only we could find a way for states to understand each other correctly, we could eliminate many of the nasty outcomes that plague international politics.

Gyrnaviski offers an important corrective to this notion. His central argument is that under some conditions, cooperation *requires* misperception. Sometimes—if, for instance, two actors have fundamentally incompatible worldviews—holding accurate beliefs about each other's preferences or principles would obstruct efforts to coordinate policy. When this is the case, inaccurate beliefs about what the other side intends to do, or why it chooses to do so, can facilitate cooperation. Gyrnaviski substantiates this claim by showing that inaccurate beliefs about the other side's preferences and principles were a necessary condition for the emergence of détente between the United States and the Soviet Union during the early 1970s. Had President Richard Nixon and Soviet leader Leonid Brezhnev actually held accurate beliefs about each other's positions—had they, for instance, understood the nature of each other's conceptions of what was implied by the text of the 1972 Basic Principles Agreement, or held accurate beliefs about each other's true preferences with respect to anti-ballistic missile negotiations—cooperation would have died in the cradle.

This is a counterintuitive but persuasive argument, and Gyrnaviski has done a service to the field by developing it. The book's primary contribution is conceptual and theoretical. Gyrnaviski introduces a novel concept, the "false intersubjective belief," or FIB (5). An FIB is a false belief that two actors share another belief. If, for instance, I believe that God exists, you believe that God exists, and I believe that you believe that God exists, then, according to Gyrnaviski, the belief that God exists constitutes an intersubjective belief. But if, in reality, you do not believe that God exists, then I hold a "false intersubjective belief," since I believe that you believe something that you do not believe.

Gyrnaviski's focus in the rest of the book is on a particular type of FIB, which he calls "imagined intersubjectivity." This refers to a mistaken belief about what another actor is likely to do and about the principles or rationale that guide the other actor's behavior (2). The book's core argument is the novel theoretical claim that "imagined intersubjectivity" is sometimes necessary for cooperation, because in some

---

<sup>1</sup> For the classic statements of these arguments, see Robert Jervis, "Cooperation Under the Security Dilemma," *World Politics* 30:2 (1978), 167-214; Thomas Christensen and Jack Snyder, "Chain Gangs and Passed Bucks: Predicting Alliance Patterns in Multipolarity," *International Organization* 44:2 (1990): 137-168; James Fearon, "Rationalist Explanations for War," *International Organization* 49:3 (1995): 379-414; Robert Keohane, *After Hegemony: Cooperation and Discord in the International Political Economy* (Princeton: Princeton University Press, 1984).

cases having accurate information about the other actor's likely actions or principles would lead to conflict or at least obstruct policy coordination. Beyond making this important conceptual and theoretical contribution, Grynaviski also deploys a sophisticated qualitative research design, in which he carefully pairs process-tracing and counterfactual analysis to great effect.

Still, there are two elements of the book that are less than satisfying. First, Grynaviski pitches the argument as a broadside against constructivist accounts of international politics. This is unpersuasive, largely because Grynaviski's characterization of constructivism is incomplete. Second, Grynaviski does not provide a satisfying account of how misperception can lead to sustained cooperation—that is, cooperation that lasts beyond the revelation that the initial steps were based on a misreading of the other side's beliefs and intentions. The remainder of this essay elaborates these two problematic aspects of *Constructive Illusions*.

For the most part, *Constructive Illusions* is a narrow, careful book: it shows no more than that misperception can be necessary for short-term cooperation, and Grynaviski usually takes great care not to overreach. This caution is in tension with the very broad way in which Grynaviski situates the argument and raises the study's stakes: he argues that the book undermines our entire understanding of the way that information and meaning work in world politics. This implicates two influential analytical traditions: rational choice and constructivism.

Grynaviski's argument about rational choice approaches is fair, if perhaps overstated. It is true, as Grynaviski notes, that rationalist accounts of international politics usually suggest that a lack of information is an obstacle to cooperative interaction. While one can easily imagine situations in which accurate information is not necessary to avoid conflict within the context of a rationalist model (for instance, while underestimating another actor's reservation value is likely to lead to war, overestimating it is not), rationalist accounts of cooperation do typically revolve around improving information. *Constructive Illusions* is thus a powerful critique of rationalist accounts of the role of information in international politics.

The critique of constructivism, though, is more dubious. Grynaviski argues that constructivist accounts of cooperation are essentially the same as rationalist accounts, except that in the former intersubjective beliefs play the role that common knowledge does in the latter. For Grynaviski, in other words, intersubjective beliefs are largely functionally equivalent to common knowledge. They refer to shared understandings of a particular situation which facilitate cooperation by allowing actors to predict each other's behavior. Conflict, on the other hand, "stems from an information or meaning deficit" (28).

The real crux of Grynaviski's critique of constructivism—what gives *Constructive Illusions* its bite—is that constructivists have never thoroughly investigated the extent to which intersubjective knowledge actually characterizes international politics. Instead, they typically infer the existence of intersubjectivity from the observation that states quite often cooperate. Since—according to Grynaviski's depiction of constructivism—cooperation requires intersubjectivity, when constructivists see states cooperating, this leads them to the conclusion that states must share beliefs about the world. Grynaviski's aim in *Constructive Illusions* is to disabuse constructivists of this notion. By showing that cooperation does not require shared knowledge and in fact sometimes (and, Grynaviski implies, perhaps normally) requires false intersubjective beliefs, he hopes to clear the theoretical ground and set the stage for an approach to studying information in world politics that assumes that relations between states are characterized more by misunderstanding than by shared knowledge (10-13, 158-160).

This is a provocative but ultimately unconvincing argument. The problems begin with Grynawski's characterization of constructivist accounts of international politics. Constructivist theoretical frameworks do not typically focus on "common knowledge," but rather on what Alexander Wendt called "collective knowledge." The difference is that while "common knowledge is *nothing but* beliefs in heads," collective knowledge refers to "knowledge structures held by groups which generate macro-level patterns in individual behavior over time." Common knowledge refers strictly to individual beliefs; collective knowledge, on the other hand, refers to things like "social facts," discourses, and practices. These phenomena are distinct because they are not "*reducible* to individuals' beliefs;" rather, they exist—and confront individuals with real consequences—regardless of whether any particular individual believes in or participates in them.<sup>2</sup> The U.S. dollar, for instance, would not lose its value simply because I did not believe that it had any; similarly, racial differences do not lose their social, economic, and political consequences—do not cease to exist—simply because an individual says that he does not recognize them. In short, constructivist accounts of international politics are not so much about common knowledge as they are about collective ideas, discourses, and practices that construct, enable, and constrain agents regardless of whether any two particular agents agree on the exact content of any particular element of these structures.

This problem has implications for the significance of Grynawski's empirical findings. In Chapters two and three, Grynawski shows quite convincingly that American and Soviet leaders had important misperceptions about each other's beliefs and intentions. But since intersubjective knowledge is not reducible to private beliefs, this is not evidence of the absence of intersubjective knowledge. And it does not demonstrate that intersubjective knowledge played no role in constituting Soviet-American interaction during this period—indeed, Grynawski concedes as much at the end of both chapters (85, 116). Arguably, intersubjective knowledge structures—in the form, for instance, of diplomatic practices—were quite significant, as they are whenever representatives of states interact with one another.<sup>3</sup> This does nothing to undermine the book's narrow claim about the role of misperception in fostering policy coordination; it does, though, mean that Grynawski's argument does not undermine constructivist accounts of world politics in the way he suggests it does.

Another theoretical mischaracterization has to do with the consequences of intersubjective knowledge. Grynawski depicts constructivism as a kind of ideational liberalism. The development of stocks of shared knowledge has the same effect as the development of international institutions: it allows states to anticipate each other's behavior and thereby fosters cooperation. Conflict, on the other hand, results from the absence of shared knowledge.

This has particularly important implications for Grynawski's research design. *Constructive Illusions* uses a logic of inference based on process tracing in a small number of non-randomly selected cases. This means that in order to argue that his findings imply that there may be more false intersubjective knowledge than intersubjective knowledge in international politics in general, Grynawski has to use a critical case design. In

---

<sup>2</sup> Alexander Wendt, *Social Theory of International Politics* (Cambridge: Cambridge University Press, 1999), 161-164.

<sup>3</sup> On diplomatic practices, see Vincent Pouliot, *International Security in Practice: The Politics of NATO-Russia Diplomacy* (Cambridge: Cambridge University Press, 2010), and *International Pecking Orders: The Politics and Practice of Multilateral Diplomacy* (Cambridge: Cambridge University Press, 2016).

particular, he argues that U.S.-Soviet détente constitutes a “likely” case for constructivism (53, 85). It is a “paradigmatic” case of cooperation, and since cooperation implies shared beliefs, if we do not observe shared beliefs in this case, we should be skeptical that we will observe shared beliefs very often in other cases either (42). This provides an empirical warrant to claim that we should fundamentally reconsider the way we think about knowledge and meaning in world politics.

The problem is that constructivist accounts of world politics do not imply that short-term episodes of cooperation—like détente—are the most likely contexts in which to look for intersubjective beliefs, discourses, or practices. Indeed, there is nothing about constructivist accounts of international relations that implies that there is any relationship at all between intersubjective knowledge and cooperation. Wendt was explicit about this point in *Social Theory of International Politics*, and others have shown that shared ideas, discourses, and practices can constitute conflict just as easily as they constitute cooperation.<sup>4</sup> The claim at the core of the constructivist project is rather that “stable patterns of behavior” (which are not necessarily cooperative) are made possible by stocks of shared knowledge and meaning.<sup>5</sup>

The question of critical case selection looks quite different from this perspective. It does not matter that détente was an episode of cooperative interaction, since constructivist accounts of world politics do not imply that intersubjective knowledge should make cooperation any more likely than conflict. What matters, instead, is whether détente constituted an episode of stable, patterned interaction. Did it? Perhaps, but it could also be interpreted as a transitional and fleeting episode that led from one period of enmity to another, which is not the kind of case in which one would expect to find deeply shared stocks of knowledge. And in any event, it is certainly not the most likely context in which to observe the kinds of shared knowledge structures that are at the heart of *Social Theory* and other constructivist work: if one is looking for extreme cases of “stable patterns of behavior” in order to show that they are based on misunderstandings and thereby undermine constructivist accounts of international relations, then one should analyze cases of deep, enduring enmity, rivalry, or friendship, not periods of transition between them.

Once again, this does nothing to threaten the important argument at the heart of *Constructive Illusions*: that misperception can be a necessary cause of policy coordination. But, in this reader’s opinion, the book does not have the sweeping implications for the study of meaning in world politics that the author claims.

The second problematic area involves the book’s claim about the significance of FIBs for long-term, stable patterns of cooperative behavior. *Constructive Illusions* makes a very persuasive argument that misperception can be a necessary condition for short-term policy coordination. But this argument raises important questions about how sustainable—how robust—cooperation built on misunderstandings is. After all, if misperception is a necessary condition for cooperation in some cases, then in those same cases the absence of misperception is logically a sufficient condition for cooperation failure. In short, cooperation based on misperception may last only until actors realize that they have been operating according to false assumptions about each other. This

---

<sup>4</sup> Wendt, *Social Theory*, 251, 254; Jennifer Mitzen, “Ontological Security in World Politics: State Identity and the Security Dilemma,” *European Journal of International Relations* 12:3 (2006), 341-370.

<sup>5</sup> Wendt, *Social Theory*, 251.

## H-Diplo/ISSF Roundtable 9-13

sort of story is still interesting, but it does threaten Grynawski's ability to claim that misperception plays a critical role in the construction of stable cooperative patterns of interaction.

This tension is present from the very beginning of the book: Grynawski opens with the story of Captain Cook's arrival in Hawaii in 1778. Initially, the Hawaiians mistook Cook for the god Lono, and treated him accordingly. After leaving the island, though, Cook's mast broke which forced him to return. This disabused Hawaiians of the notion that Cook was Lono, and they wound up killing him (1-2, 160-161). Other examples of cooperation based on misperception that Grynawski offers to illustrate his argument similarly last only as long as the misperception does. Two people meet up for dinner, one of them thinking that the get together is a date, the other not. Things go smoothly until the end of the evening, when the hopeful romantic leans in for the kiss, creating an awkward moment that is likely to prevent future encounters. Two children are engaged in playing a game, though they disagree about the rules. If, by some chance, they happen to realize that they disagree about the rules, they are likely to begin arguing and stop playing (5).

In all three of these examples, cooperation persists only so long as misperception does. The problem is that as actors continue to interact, the likelihood that they will realize that they were wrong about fundamental aspects of a situation increases. This means that cooperation based on misperception may be inherently unstable and transitory.

Grynawski is aware of this problem and proposes, in the book's introduction, four pathways through which misperception might produce sustained, robust cooperation (8-10). Unfortunately, he never fully develops these arguments, and offers little if any compelling evidence in their support.

One of these pathways—when an FIB is never revealed—does not explain how cooperation can survive the revelation of misperception. A second—when an FIB contributes to cooperation that over time changes actors' identities and interests—is more plausible and worth exploring. This pathway does not, however, make an appearance in the book after page 10.

The two pathways that Grynawski claims are illustrated by the case of U.S.-Soviet détente relate to the development of trust and the influence of public commitments on domestic political incentives. First, Grynawski suggests that misperception can lead to first moves that are misinterpreted as concessions, which can lead to reciprocation. This can build trust, which may allow cooperation to persist “even if the initial concession is later revealed to have been a misperception” (8).

This pathway, Grynawski says, is illustrated by his account of anti-ballistic missile negotiations—and in particular by the record of the May 1971 agreement that set the stage for the signing of the final version of the Strategic Arms Limitation Treaty in Moscow a year later. The story that Grynawski tells about this episode revolves around a misunderstanding related to the sequence of negotiations: National Security Advisor Henry Kissinger wanted negotiations about an ABM treaty to occur simultaneously with negotiations about a freeze in Soviet construction of offensive weapons; the Soviets, on the other hand, wanted negotiations on offensive missiles to occur only after an agreement had been reached on an ABM treaty. The text of the May 1971 agreement was ambiguous enough that both sides thought the other had agreed to its position, which allowed negotiations to move forward. This is a compelling argument, but there is nothing in the discussion about the development of a cycle of cooperation that produced trust between Soviet and American leaders—indeed (unless I have missed it) the word “trust” does not even appear in that part of the

text (97-102). Grynaviski never suggests or offers evidence that American leaders updated their beliefs about the trustworthiness of their Soviet counterparts—or vice versa—during or following this episode.

Second, Grynaviski argues that misperceptions can prompt leaders to make public commitments to cooperation that then become difficult to renege on due to the fear of paying domestic ‘audience costs.’ Thus, even if the misperception is revealed, it will have produced a change in the domestic political calculus of one or both actors in a way that pushes toward continued cooperation. One problem with this argument is that the explanatory status of ‘audience costs’ has come under serious fire in recent years. It is far from clear that public commitments tie leaders’ hands in the way that Grynaviski suggests they do.<sup>6</sup>

Another problem is that the evidence Grynaviski offers in support of this pathway is fairly weak. The public commitment pathway, he says, is also adequately illustrated by the negotiations over the ABM treaty. The most problematic claim has to do with the 1972 Moscow Summit. Grynaviski argues that Nixon had no intention of attending the Summit because it would be awkward for him to have appeared in Moscow during the North Vietnamese ‘Easter Offensive.’ However, Nixon also thought that backing away from the Summit would be domestically costly, since he had already publicly committed to attending and had learned that cancelling might be unpopular. Apparently, though, he thought the cost of attendance would be higher than the cost of cancellation. Thus, the order of his preferences, according to the game tree that Grynaviski provides on page 114, was (from highest to lowest utility): have Brezhnev cancel the Summit; cancel the Summit himself; attend the Summit.

Until 11 or 12 May, Nixon and Kissinger both believed that Brezhnev was likely to cancel—thus they thought that Nixon could achieve his most preferred option: avoiding the Summit without paying the costs of canceling. On one of those dates, though, Kissinger became convinced that Brezhnev was not going to cancel. By this point, according to Grynaviski, it was too late: Nixon was forced to attend even though he preferred not to. This, importantly, was the result of a misperception that had been revealed before cooperation had been consummated. It thus seems to be evidence that misperception can tie leaders’ hands and promote cooperation even after it has been revealed for what it is.

The problem here is that Grynaviski never makes clear why Nixon—whose preferences, presumably, had not been affected by the realization that Brezhnev was not going to cancel—did not himself cancel the meeting sometime after May 12. In Grynaviski’s counterfactual analysis, he sets the situation up as a game of chicken. Nixon had to move first. Because he thought that Brezhnev would cancel, he choose the option not to cancel, which left him in position to achieve his highest payoff by not attending while avoiding the blame for cancelling. Once it turned out that he was wrong about Brezhnev’s preferences, he had already chosen his strategy: he was forced to follow through and accept his lowest payoff by attending the Summit. If, on the other hand, Nixon had understood Brezhnev’s preferences from the get go, he would have cancelled the Summit himself sometime during the first week of May (thereby avoiding ending up with his lowest payoff) and the SALT treaty would never have happened.

---

<sup>6</sup> See Jack Snyder and Erica Borghard, “The Cost of Empty Threats: A Penny, Not a Pound,” *American Political Science Review* 105:3 (2011), 437-456; Marc Trachtenberg, “Audience Costs: An Historical Analysis,” *Security Studies* 21:1 (2012), 3-42; and the symposium on audience costs in *Security Studies* 21:3 (2012).

But reality is not a game of chicken. In reality, actors are allowed to change their strategies after updating their beliefs about the preferences of other players. Nixon, in other words, could have cancelled the Moscow Summit after Kissinger alerted him to the fact that Brezhnev was planning on going through with it. This would have been entirely consistent with Grynawski's depiction of Nixon's preferences. Nixon thought that he would pay a heavy domestic price for cancelling, but—based on Grynawski's own analysis—it is clear that he had concluded that this price would have been lower than the cost of attending the Summit during a particularly intense period of the Vietnam War. In effect, Grynawski's analytical framework seems to have imposed an artificial constraint on Nixon's decision-making in a way that makes it look like misperception had an effect on cooperation that persisted even once it had been revealed as a mistaken judgment. Yet there is no discussion in the text that explains why exactly Nixon's hands were tied after May 12—why he was prevented from canceling a meeting he did not want to attend sometime after he realized that Brezhnev was not going to do it for him.

To his credit, Grynawski does attempt to directly take on the argument that cooperation based on misperception is fragile. In Chapter 4, he makes a strong case that while détente fell apart soon after it started, this had very little to do with any revelation that it had been founded on inaccurate beliefs. Instead, the end of détente was the result of domestic political changes in the United States that brought to power actors who preferred a harder line against Moscow. But showing that détente ended for reasons other than revealed misperceptions does not establish that cooperation can survive the revelation of misperception. For this, Grynawski would have to have identified an instance of robust, sustained cooperation, established that it was in part based on misperception, shown that it persisted after those specific misperceptions were revealed, and then explored the mechanisms that account for its persistence.

Like the unconvincing effort to undermine constructivist accounts of meaning in world politics, the failure to establish that misperception can contribute to sustained, robust cooperation does nothing to weaken the core of the argument at the heart of *Constructive Illusions*: that misperception can be a necessary condition for policy coordination. This is an important and novel argument, and Grynawski has made it in an accessible, sophisticated way. The book deserves a place on graduate syllabi, and demands attention from others who study information and cooperation in world politics.

## Author's Response by Eric Gynaviski, George Washington University

---

International Relations (IR) scholars often suggest that common knowledge or intersubjectivity is the bedrock for international cooperation. Rationalists point to common knowledge as the linchpin for cooperation; it makes states predictable in ways that makes cooperation possible. Constructivists prefer the term intersubjectivity. Common cultures, identities, or norms undergird cooperation. For both, transnational shared ideas are important for moving from cooperation to conflict.

The central aim of *Constructive Illusions* is to show that shared ideas are not important for international cooperation. States do not need to share ideas so long as they think they share ideas. This concept—where we incorrectly infer that elements of a relationship are common knowledge or intersubjective—I refer to as a false intersubjective belief, or FIB.

The basic logic is exhibited in kids' games. When kids from different schools meet and play freeze tag on the playground, they are often unfamiliar with the nuances in the rules played at other schools. Does a hug, tap, or crawl under the legs unfreeze someone? Does everyone need to be tagged once for the game to end, or does everyone have to be frozen? How many kids can be 'it,' especially if there are really small children playing?

When children meet, there are two possible ways to resolve rule disagreements. Under the first option, they are aware of their differences. In these situations, they meet in congress to determine the rules of the game. Often, disputes over what the rules should be, or the length of the meeting itself, undo their ability to play before the game even begins. Alternatively, they can just start playing, unaware that these nuances will matter later on. In the latter case, the children believe there is a shared set of rules; they can begin to cooperate because they misperceive others' beliefs. Once the kids are locked into the game and having fun, they iron out rule disagreements on the fly, perhaps never coming to a consensus on what all the rules mean.

This basic logic, *Constructive Illusions* shows, was at the heart of the détente between the superpowers, especially during the Nixon administration. The superpowers met and cooperated, signing about 150 agreements during the period. From arms control to trade, the superpowers engaged in a remarkable process of cooperation. The conventional wisdom holds that this was due to the development of common knowledge or intersubjectivity. *Constructive Illusions* shows that the inverse was the case. The superpowers only cooperated because mutual misunderstandings led them to overlook interest-based and ideational conflicts.

I want to thank Arjun Chowdhury and Steven Ward for their thoughtful reviews. They explore largely underdeveloped elements of the book at the theoretical and empirical level. I appreciate the opportunity to respond.

*What is Constructivism?*

Ward's primary concern is that I do not properly engage with constructivist theories of intersubjectivity. In particular, he claims that I mischaracterize constructivist scholarship because I mischaracterize Alexander Wendt's views.<sup>1</sup> This matters. One piece of this book is a criticism of the view that intersubjectivity is necessary for deep international cooperation. If this book gets the logic behind intersubjectivity wrong, then

---

<sup>1</sup> Alexander Wendt, *Social Theory of International Politics* (Cambridge: Cambridge University Press, 1999).

the book misses the target. In making his argument, however, Ward gives a quite narrow view about what counts as constructivist scholarship. Since I am not specifically targeting this narrow view throughout the book, Ward concludes that *Constructive Illusions* does not implicate constructivist scholarship. This section thus briefly explains the wider view that Ward is missing.

### *What is Intersubjectivity?*

IR scholars often argue that intersubjective agreements are common in international politics. The international system appears orderly. How could there be order without common knowledge that makes the system predictable for the states that inhabit it? In a response to Dale Copeland, who argues that shared ideas play little role in world politics, Wendt writes that “far from facing profound uncertainty, states are confident about each other’s intentions almost all the time.” Why? Because without “a deep reservoir of common knowledge . . . the international system today would be far more chaotic and conflictual than it is—indeed there would not be an ‘international system’ at all.”<sup>2</sup>

This is one primary argument target for the book. Showing that states do not need common knowledge to generate orderly patterns of behavior means that there is no necessary reason to posit “a deep reservoir of common knowledge.” In doing so, the book clears a path for thinking about how diversity in ideas and beliefs is consistent with, and perhaps even constitutive of, international cooperation (12-13).

Ward believes that targeting common knowledge means that *Constructive Illusions* misses the mark. His view is that most constructivists are not interested in common knowledge—common beliefs that individuals knowingly share—but rather in collective knowledge. He points to Wendt’s seminal *Social Theory of International Politics*, in which Wendt explicitly points to an account of intersubjectivity that goes beyond beliefs to other forms of knowledge that are not reducible to intentional states.

I disagree. One of the aims of the first chapter of *Constructive Illusions* is to show the complicated ways that common knowledge—an idea familiar to rationalists—is also important for constructivist theories of cooperation. For example, most of the empirical work on socialization and compliance—a central debate—treats the diffusion of beliefs and preferences (the development of common knowledge) as an important question. The diffusion of beliefs, of course, refers to changes in individuals’ beliefs and preferences, belying the claim that agents’ beliefs do not matter. Work on international norms often posits that norms are standards of appropriate behavior, and work on the origins of norms points to them being ‘taught,’ which implies belief transmission.<sup>3</sup> More specifically, one target of the book is constructivist work on arms control.

---

<sup>2</sup> Alexander Wendt, “Social Theory as Cartesian Science,” in *Constructivism and International Relations: Alexander Wendt and His Critics*, ed. Stefano Guzzini and Anna Leander (New York: Routledge, 2006), 211; Dale Copeland, “The Constructivist Challenge to Structural Realism: A Review Essay,” in *Constructivism and International Relations: Alexander Wendt and His Critics* (New York: Routledge, 2006), 1-20.

<sup>3</sup> Jeffrey Checkel, “Why Comply? Social Learning and European Identity Change,” *International Organization* 55:3 (2001): 553-588; Martha Finnemore, *National Interests in International Society* (Ithaca: Cornell University Press, 1996); Martha Finnemore, “Norms, Culture, and World Politics: Insights from Sociology’s Institutionalism,” *International Organization* 50:2 (1996): 325-347; Alexandra Gheciu, “Security Institutions as Agents of Socialization? NATO and the ‘New Europe,’” *International Organization* 59:4 (2005): 973-1012; Audie Klotz, *Norms in International Relations: The Struggle Against Apartheid* (Ithaca: Cornell University Press, 1995).

## H-Diplo/ISSF Roundtable 9-13

Matthew Evangelista, for example, describes how the Cold-War rivals learned to share a set of strategic concepts through the work of epistemic communities, enabling cooperation.<sup>4</sup>

Wendt, whom Ward cites as the authority on constructivist scholarship, disagrees strongly with Ward's claim that common knowledge is not important for constructivist scholarship. As I note in *Constructive Illusions* (23-24), Wendt claims that most constructivists treat common knowledge as the dominant mode of knowledge in world politics. He writes explicitly that "the concept of common knowledge is equivalent to that of 'intersubjective understandings' favored by constructivists."<sup>5</sup> He goes on to outline how constructivists rely on the rationalist definition of the concept. He contrasts what he considers the standard constructivist view—intersubjectivity as common knowledge—with his own view premised on collective knowledge. In short, whereas Ward presents Wendt's theory of knowledge as the only theory of knowledge, Wendt describes it as a sophisticated alternative to the much larger body of empirical and theoretical work that defines intersubjectivity exactly as *Constructive Illusions* does.

*Why focus on the transition to cooperation?*

Ward incorrectly identifies a second mischaracterization as my turning constructivism into a kind of liberalism, where improvements in mutual understanding foster cooperation. As Ward points out, this would be wrong, I agree. I do not claim that enmity is the result of an inability to understand others. Rather, I simply make the claim that a transition from enmity to rivalry is not necessarily the product of better mutual understanding. There are few instances, I would claim, where cooperation or conflict is the result of deep intersubjectivity (more on what this means later).

Why focus then on cooperation? There are two reasons. First, it would not be novel to show that conflict is caused by misperception. Second, there is a cooperation bias in the constructivist literature. Martha Finnemore and Kathryn Sikkink, as early as 2001, wrote, "One consistent complaint about constructivist research has been its research focus on norms most of us would consider 'good,' such as human rights, protecting the environment, and promoting democracy."<sup>6</sup> On balance, there remains a cooperation-bias in the literature, which is sometimes explicit.<sup>7</sup> In general, the conventional wisdom, as I read the literature, is that increasing mutual understanding leads to cooperation. This strikes me as a reasonable target when engaging with constructivists.

In making this argument, Ward also posits that constructivists are not interested in political change, but focus instead on stable patterns of cooperation or conflict. In contrast, I believe that constructivists are *very*

---

<sup>4</sup> Matthew Evangelista, *Unarmed Forces: The Transnational Movement to End the Cold War* (Ithaca: Cornell University Press, 2002).

<sup>5</sup> Wendt, *Social Theory of International Politics*, 159–161.

<sup>6</sup> Martha Finnemore and Kathryn Sikkink, "Taking Stock: The Constructivist Research Program in International Relations and Comparative Politics," *Annual Review of Political Science* 4:1 (2001): 404.

<sup>7</sup> Alexander Wendt, "Why a World State Is Inevitable," *European Journal of International Relations* 9:4 (2003): 491.

interested in ideational change, and that the most cited and debated works in the field, such as those on the norm life-cycle, focus on change.<sup>8</sup> Constructivists are even interested in ideational change that is short-lived. Constructivists can be interested in Soviet New Thinking, even if it did not last in the face of Russian President Vladimir Putin's rise. They would also be interested in the growth of anti-torture norms, even if those norms were rolled back after 2001. And, they will likely be interested in whatever U.S. President Donald Trump does, even if he stays in office for only for four years.

Moreover, Wendt emphasizes short-term changes during periods of transition. Why might these short-term transitions matter? During periods of transition, the ideas in agents' heads are of vital importance. To return to Ward's example of money as a social fact to illustrate, imagine a world in which there are only two individuals who treat money as a social fact: only these two parties can and do exchange the good. If one were to stop counting green pieces of paper as dollars, then the social fact would cease to exist, because one party would refuse to accept green pieces of paper as a currency. Only after a transition occurs, and the knowledge of the dollar as a social fact diffuses, does it begin to exert the kind of force, irrespective of beliefs, that Ward claims.<sup>9</sup>

I selected détente as a case in part for these reasons. Beginning in 1969, the superpowers agreed to try to settle the deep disagreements that had divided them in international politics, and they developed rules to limit superpower competition. They talked about what they valued (e.g., equality) and how that should translate into broad principles (the Basic Principles Agreement), negotiating strategies (debates about linkage), and competition across the world. This strikes me as a good, but perhaps not perfect, case for evaluating whether shared ideas, which presumably should be produced by discussion, debate, and dialogue, contributed to the cooperation that later unfolded (53-54).

*Why should scholars who are interested in macro-historical work care?*

I want to step back from the specific focus on Wendt invoked by Ward, and ask what scholars interested in longer-term historical patterns would take away from this book. I believe there are three conclusions that emerge from my work. First, there are methodological implications for the way we study shared ideas. How do we know if intersubjectivity is present in the international system? One worry that scholars who are interested in longer-term historical patterns should have concerns the level of generality in their theories and whether it matches the empirical evidence. Consider the example of the Westphalian state system. Macro-historical constructivist scholars often cite the emergence of the modern state as an example of a long-term pattern that conditions international life. With the emergence of the modern state came important international norms, such as non-interference in others' affairs. Is the logic of sovereignty intersubjective?

The choice of method, I would argue, answers this question. Sovereignty, like détente, appears to be obviously intersubjective, because states, for hundreds of years, have shared ideas about states' rights and prerogatives. A

---

<sup>8</sup> Martha Finnemore and Kathryn Sikkink, "International Norm Dynamics and Political Change," *International Organization* 52:4 (Autumn 1998): 887-917.

<sup>9</sup> This plays with ideas in John Searle, *The Construction of Social Reality* (New York: Free Press, 1995). However, this argument was likely passed over too quickly in the book (54), and on rereading it I discover that it was jargon-laden.

macro-historical account that focuses on change over time in general rules would paint deep agreement. A finer-grained look at any specific period, however, shows massive disagreements between communities about the meaning of sovereignty. We see this from Westphalia, when different communities understood the requirements of Westphalia differently, into the post-Cold War era, when, for example, sovereignty meant very different things to actors in China versus the United States.<sup>10</sup> *Constructive Illusions* shows the dangers of using too macroscopic of a lens through which to study shared ideas.

A second lesson from *Constructive Illusions* is more theoretical. IR scholars, including Wendt, often talk about the international system using the same theoretical toolkit—shared identities, social facts, constitutive rules, socialization, social position, and so on—that anthropologists and sociologists use to describe the densest social systems on the planet. Using this toolkit explains why they search for shared ideas; the presence of shared ideas is an assumption within much of the literature on these kinds of societies. The punchline of the book for constructivists is that this is likely not the most productive approach to thinking about the social architecture of the international system. Instead, we should begin with the assumption that the international system is the most diverse social system on the planet. This diversity marks it as different in kind from systems studied in other disciplines. Therefore, when thinking about the relationship of work in our field to works in other fields, our starting point should be theories that highlight diversity, rather than sameness (158-160). *Constructive Illusions* is only a first step toward this view. It shows that cooperation, under certain conditions, is possible only in a diverse social world. In doing so, it undermines the argument that the presence of norms or cooperation is evidence of a shared international culture.

### *Can cooperation be sustained?*

A second question both reviewers asked about is the sustainability of cooperation. When misperceptions are revealed, they should undermine cooperation. When kids discover that their rules for tag differ, they often stop playing and return to fighting about the rules. In the same way, when the superpowers' actions revealed that they never understood one another, they should have returned to fighting. One would thus expect chapter five, which analyzes the decline of détente, to show that when superpowers learned more about each other, increasing their intersubjectivity, cooperation declined.

The historical record, however, undermines this view. The decline of cooperation instead occurred because hardliners came to power in the Ford and Carter administrations. These actors believed that the Soviet Union was an aggressive power that used détente as a fig leaf while expanding in the developing world and building more weapons. After Watergate, their pressure on the Ford and Carter administrations created institutional positions through which they could advocate their views and write legislation that undermined détente. These hawks did not understand the Soviet Union any better than the Nixon administration did (see chapter four).

Chowdhury believes this points to important scope conditions to the argument. He writes that FIBs promote cooperation only “when decision-makers prefer cooperation for different reasons, and when shared knowledge

---

<sup>10</sup> Andreas Osiander, “Sovereignty, International Relations, and the Westphalian Myth,” *International Organization* 55:2 (2001): 251-287; Allen Carlson, “Helping to Keep the Peace (Albeit Reluctantly): China’s Recent Stance on Sovereignty and Multilateral Intervention,” *Pacific Affairs* 77:1 (2004): 9-27.

of these reasons would impede cooperation.” Chowdhury’s scope condition is exactly right about détente in general (chapters 2 and 4). The growth of anti-cooperation sentiment scrapped cooperation.

There are theoretical instances, however, when cooperation founded on misperception may be durable in the face of initial anti-cooperation preferences. First, after cooperation begins, key political agents’ interests may change. But having cooperated, they get used to working together in ways that mute interest-based and ideational disagreements.<sup>11</sup> Second, after parties decide to cooperate, they may discover political costs to backing out of agreements. This is seen in the example of the Anti-Ballistic Missile Treaty (ABM), which I discuss below. Third, anti-cooperation sentiment may depend on mistrust. If cooperation that begins from misperception leads to the building of trust, then it may change policymakers’ concerns about cooperation. Only one of these arguments is featured in this book, although the others are worth exploring.

Ward raises several related criticisms about durability as well. If cooperation is founded on misperceptions, why are they durable? Ward points out—rightly—that the evidence in the book describes only some of the above-mentioned pathways to durability. The book was not intended to do more. The broad themes in the introduction are stated more generally than the specific empirical evidence in the middle of the book. Here I want to mention two pathways that strike me as particularly important to unpack.

One pathway is similar to audience costs. Two states may have a misperception that leads to cooperation; once the misperception is discovered, they may stick to the agreement for fear of the domestic political consequences of renegeing. Ward dismisses this idea because audience costs are weakly supported by some studies. Although weakly supported in general, détente is a case where they mattered. Henry Kissinger, President Richard Nixon’s National Security Adviser, feared the domestic political consequences of backing out of ABM limitations once the Soviets agreed to them. If the Soviets went public, it would undermine domestic political support. This is a version of the audience cost argument, where going public with a commitment (in favor of ABMs) makes renegeing on the commitment difficult. This is not identical to audience costs (where the literature focuses on crises), but part of the logic is the same.<sup>12</sup>

Another pathway that Ward dismisses outright is what game theorists call off-path play. In off-path play, we learn new information when we face a situation in which others behave in a way contrary to our expectations. If an event arises (a war) and a state we thought was an ally does not act in accordance with our expectations (by not cooperating), then we infer that our perception of the ally was wrong. Often, however, no event arises through which we might learn. If there is no off-path play and actors do not experiment, there are no opportunities for actors to correct their misperceptions. In the above example, if a war does not occur, then we will never know whether our friends are really allies or not.

A sustained interest in off-path play would lead to deep insights about the nature of international political culture. We develop norms and institutions to handle situations that often arise; there is, however, much less agreement about what to do further down the road or in low-probability situations. If true, then there is less

---

<sup>11</sup> Jennifer Mitzen, *Power in Concert: The Nineteenth-Century Origins of Global Governance* (Chicago: University of Chicago Press, 2013).

<sup>12</sup> James Fearon, “Domestic Political Audiences and the Escalation of International Disputes,” *The American Political Science Review* 88:3 (1994): 577-592.

## H-Diplo/ISSF Roundtable 9-13

intersubjectivity than many would suggest. These misperceptions, however, are not revealed because they are off-the-equilibrium path; they are never revealed unless actors go searching for off-path differences through experimentation. I find this idea very interesting; it suggests that international order depends more on us behaving predictably than on understanding why agents behave as they do. Some of these ideas are briefly explored in chapter one as examples of incompletely theorized agreements, drawing on Cass Sunstein's work (35-36).<sup>13</sup>

### *On the Nixon Summit*

Ward has some concerns about the analysis of Nixon's decision making around the summit. To set the stage, chapter three of *Constructive Illusions* describes the process by which the United States and the Soviet Union reached the ABM Treaty. It looks at three mutual misperceptions that elites in Washington and Moscow had about one another's positions. Two relate to the substance of arms control.

Ward's concern is the third episode, the decision about the summit and its relationship to the Vietnam War. In 1972, Nixon faced a dilemma. He had an upcoming summit with the Soviets in Moscow in which the ABM Treaty (among other agreements) would be signed. But, before the meeting, North Vietnam launched an offensive against South Vietnam. Nixon feared the domestic political consequences of meeting with Soviet leader Leonid Brezhnev in the middle of a major escalation. At first, Nixon planned to cancel his meeting with Brezhnev; but then, advisor John Haldeman arrived with polling data showing that the summit was popular. Nixon decided, in the end, not to cancel. Why? Chapter three shows that Nixon agreed not to cancel because of a misperception. He believed Brezhnev also had to cancel because they could not meet during an escalation of the war. If Brezhnev canceled, then Nixon would reap some political rewards for being willing to go, while avoiding the political costs of actually going. As events transpired, Brezhnev did not do what was expected, and he did not cancel. As time passed, Nixon realized that he no longer needed to cancel because the situation was improving.

Ward asks why Nixon did not change his mind once he learned Brezhnev planned to attend the summit. Should not Nixon have canceled when he learned Brezhnev would not cancel?

Ward is right. Nixon changed his mind. When Nixon chose not to cancel, believing Brezhnev would (around May 2, 1972), he escalated the war and blockaded North Vietnam. Nixon quickly came to believe that the offensive was not as dire as he had earlier thought, and he believed that the use of force was working (around May 11-12). As a result, events transpired that would again make it politically possible to meet. The basic logic is straightforward. Going to Moscow after successfully standing tough in Vietnam was different than the outcome Nixon had earlier feared, that he would be "in Moscow at a time when the North Vietnamese are rampaging through the streets of Hue" (104). This does not affect the conclusion I reach, which is that the original misperception (on May 2) contributed to cooperation because the summit would have been canceled before the President discovered that the offensive would be successful. The book should have finished this part of the story.

### *Future Work*

---

<sup>13</sup> Cass R. Sunstein, "Incompletely Theorized Agreements," *Harvard Law Review* 108:7 (1995): 1733-1772.

## H-Diplo/ISSF Roundtable 9-13

I want to thank Chowdhury and Ward for their helpful reviews. They have forced me to rethink some core issues within the book and led me to two ideas for future studies that can help trace these ideas in broader contexts. The first idea, prompted by Chowdhury, is to trace out more scope conditions for the argument. Under what conditions are agreements durable after misperceptions are revealed? The second idea, described by Ward, is a more explicit statement about what this means for constructivist theory. I suspect that too many of our theories of international order depend on analogies to domestic political cultures; they do not make sense of the specificity of international politics. Unpacking this is a useful future avenue for research. These reviews have therefore been enormously helpful in clarifying my own thinking about these questions, and they will prompt further work.