Don't Look Back in Anger: Cooperation Despite Conflicting Historical Narratives

YOSHIKO M. HERRERA   *University of Wisconsin–Madison, United States*
ANDREW H. KYDD   *University of Wisconsin–Madison, United States*

S tates in conflict often have divergent interpretations of the past. They blame each other for starting the conflict and view their own actions as justified retaliation, which makes them reluctant to cooperate. This phenomenon, while common in international relations, is not well understood by existing formal theories of cooperation. In the context of the Repeated Prisoner's Dilemma framework, we show that strategies that demand atonement for past misdeeds are outperformed by strategies that do not. The latter are able to get out of retaliatory cycles and return to cooperation more quickly when there are divergent perceptions of the past. We conclude with a case study of Chinese and U.S. responses to the Tiananmen protests of 1989. China and the United States strongly disagree about the cause of the Tiananmen uprising and the legitimacy of the Chinese response, but nevertheless returned to cooperation after a limited period of mutual punishment.

INTRODUCTION

States in conflict often have opposing interpretations of the past (Kacowicz 2005; Lind 2008; Ross 2005; Rotberg 2006), and these narratives can fuel further violence (Straus 2015). For instance, after the First World War, the question of "war guilt," or who bore responsibility for starting the war, was hotly disputed in the peace talks. The Allies insisted that Germany accept the blame, as reflected in Article 231 of the Versailles Treaty, and German guilt served as justification for Allied demands for reparations. This clause was deeply resented by Germany, and German politicians, and even historians, energetically promoted narratives of the war's origin that exonerated Germany (Herwig 1987; Lieber 2007). Eventually, the "war guilt lie" became an important motivating narrative for Hitler and the far right (Röhl 2015).

Generally, if each side thinks the other side started the conflict, then each side will view its own actions as justified retaliations against the aggression of the other. Compromise and cooperation will seem like backing down to a bully and letting aggression go unpunished. Any form of contrition or apology will be ruled out because each side thinks they are owed an apology, and have nothing to apologize for themselves. Thus, finding a way out of the conflict will be difficult, as each side clings to its own narrative of their own innocence and the other side's guilt.

Existing formal theories of cooperation (Axelrod 1984; Fearon 1998; Milgrom, North, and Weingast 1990; Oye 1986) are poorly equipped to analyze this problem because they assume, usually implicitly, that the history of the game is common knowledge between the players.[1] The basic Repeated Prisoner's Dilemma (RPD) model in which the strategy Tit for Tat (TFT) proved so successful at sustaining cooperation, assumes there is no uncertainty at all (Axelrod 1984). When "noise" is added to the model, it is often in the form of errors that translate intended cooperation into defection, or vice versa (Signorino 1996). These errors are recognized as such after they happen, so the history of the game is still essentially common knowledge. In this setting, a modified version of TFT called Contrite Tit for Tat (CTFT), which effectively "apologizes" for accidental defections by allowing the other side to punish it, is very successful at restoring cooperation. This provides an interesting theoretical connection to the literature on apology in international relations (Berger 2012; He 2010; 2011; Lind 2008). However, if CTFT thinks the other side started it by defecting first, it will not restore cooperation until the other side apologizes by accepting punishment. With conflicting narratives of the past, this apology may not be forthcoming, leaving the two sides mired in mutual defection.

To escape this trap, strategies need to be willing to restore cooperation to end a cycle of punishment, even if they are convinced that the other side started the conflict and has not apologized by cooperating. We consider a set of four strategies that have this characteristic to varying degrees. Two strategies have been studied in the literature. Probabilistic Tit for Tat

Corresponding author: Yoshiko M. Herrera [ORCID], Professor, Department of Political Science, University of Wisconsin–Madison, United States, yherrera@wisc.edu.

Andrew H. Kydd [ORCID], Professor, Department of Political Science, University of Wisconsin–Madison, United States, kydd@wisc.edu.

---

[1] Exceptions examine the effect of uncertainty about the underlying game and the players' payoffs or reputation (Aumann, Maschler, and Stearns 1995; Mailath and Samuelson 2006).

(PTFT), retaliates against perceived defections most of the time but forgives some defections at random (Molander 1985). Tit for Two Tats (TF2T) retaliates only if it thinks the other side has defected twice in a row. The remaining two strategies are novel. *Cooperate After Mutual Punishment* (CAMP) retaliates if it receives the "sucker's payoff" (*S*) but returns to cooperation after a round of mutual punishment (*P*). Finally, *Don't Look Back in Anger* (DLBA) retaliates if it thinks the other side has defected twice in a row (like TF2T), or if it receives the *S* payoff (like CAMP), unless (1) there have been two rounds of mutual punishment, or (2) it received the "temptation payoff" (*T*) followed by the *S* payoff. In other words, DLBA moves on and returns to cooperation when there has been enough punishment (two rounds) or when apparently imbalanced loss (*S*) was preceded by apparently imbalanced gain (*T*). These two exceptions enable DLBA to restore cooperation after it breaks down by exiting cycles of retaliation and mutual punishment. We perform an evolutionary computer tournament analysis of these four strategies and a few others which indicates that TF2T, CAMP, and DLBA form a successful ensemble and often emerge as the most successful and effective strategies for returning to cooperation when there are conflicting beliefs about the past.

We end with a discussion of how the analysis can contribute to understanding the return to cooperation in the face of persistent historical disagreements in the context of the 1989 Tiananmen uprising and United States (US)–China relations. We show that following the violence at Tiananmen Square, the US and China developed conflicting narratives of who was to blame, punished each other, but eventually returned to cooperation despite never coming to agreement about who acted wrongly. Thus, we argue that our analysis can both shed light on past events and potentially offer guidance on restoring and maintaining cooperation in the face of divergent narratives in the future.

CONFLICTING UNDERSTANDINGS OF THE PAST

The human capacity for knowledge about the world is limited. Even sincere efforts to understand the past in an unbiased way run up against incomplete evidence and the difficulty of grappling with complex and large scale events. The problem is compounded when we consider the political realm, where historical knowledge is scant, the incentives for objectivity are few, and the rewards for constructing biased narratives are potentially great. A number of scholars have argued that there is a connection between biased or self-serving narratives and conflict. To explain the puzzle of genocide, Straus (2015, 12) argues that "founding narratives" are a critical tool that elites use to garner support for genocidal violence, while "counter narratives" that focus on accommodation could also be important because they can deescalate situations. Rotberg (2006, vii) goes so far as to argue that, "Every conflict is justified by a narrative of grievance, accusation, and indignity. Conflicts

depend on narratives, and in some sense cannot exist without a detailed explanation of how and why the battles began, and how one side, and only one side, is in the right."

For instance, disputed narratives of the past are a central feature of the Israeli–Palestinian conflict. In a detailed analysis of such narratives, Kacowicz (2005, 343–4) concluded that "Each party blames totally and unconditionally the failure of the peace process upon the malign intentions of political destruction and annihilation of the other." Rather than converging over time, these narratives may actually continue to diverge, as Ross (2005) highlighted in his discussion of competing Israeli and Palestinian views.

In contemporary international affairs, the US and Russia strongly disagree about whether the US promised not to expand NATO at the end of the Cold War (Sarotte 2014; Shifrinson 2016), and therefore whether NATO expansion is a case of Western aggression. Russia justified its 2014 invasion of Crimea, and to some extent the 2022 full scale invasion of Ukraine, as a response to NATO aggression. Lukyanov (2016, 34), a prominent Russian journalist with insider access, wrote, "Moscow's operation in Crimea was a response to the EU's and NATO's persistent eastward expansion during the post-Cold War period." Russia's "loss" of Crimea in the Khrushchev era was "humiliating," and "the return of the peninsula righted that perceived historical wrong" (Lukyanov 2016, 35). In contrast, Ukrainians and Western leaders who support Ukraine reject this framing altogether, placing the blame squarely on Russia as the aggressor.

Note that the disagreement may not be about what physically happened but about how it is understood, or interpreted. In the case of World War I, the sequence of events leading up to it is not in serious dispute, the disagreement arises over the *meaning* of those events. From the allied perspective, Germany clearly started the war because its armies were the first to cross international frontiers when it invaded Belgium as part of the Schlieffen plan. From the German perspective, the Russian mobilization was the key event that started the war. When combined with the French–Russian alliance, it made it necessary for Germany to mobilize quickly and strike first, lest it be left helpless before the allied forces (Miller, Lynn-Jones, and Van Evera 1991). In the language of game theory, the dispute may not be over what happened, but whether a particular action should be coded as "cooperation" or "defection."

There are two additional reasons why we might expect historical narratives to diverge. First, particular narratives may be associated with identity groups. Adherence to the narrative may demonstrate loyalty to the ingroup (Ross 2001) and the narrative may be so important as to be constitutive of the social identity (Abdelal et al. 2009; Cruz 2000). Smith (1996) argues, "identification with a past is the key to creating the nation, because only by 'remembering the past' can a collective identity come into being. …One might almost say: no memory, no identity; no identity, no nation" (383). Moreover, Diesen and Keane (2017, 313) write,

"narratives are instrumental in constructing both national and regional identities." Social identity theory suggests that these narratives will reflect well on the ingroup and poorly on outgroups (McDermott 2009). The responsibility for negative outcomes is often externalized to others, so that one's own identity group is not blamed.

Second, interpretations of the past may be strategic, reflecting a pragmatic desire to avoid blame or justify aggression (Berger 2012; Finkel 2010; Glaeser 2005; Liang 2018). For instance, in the case of World War I, Germany was saddled with reparations justified by the war guilt clause. To undermine the narrative of war guilt was, therefore, to weaken the case for reparations and justify non-payment. Likewise, in order to protect its international image, the Turkish government strongly opposed the work of historians in researching the Armenian Genocide and repeatedly tried to block the U.S. Congress from recognizing it as such, although a resolution was eventually passed in 2019 (Suny 2009). In U.S. domestic politics, when President Trump lost the 2020 presidential election, he (falsely) claimed it was a result of widespread fraud. This narrative justified his effort to use violence to retain power in the January 6th attack on the U.S. Capitol.

If conflicting narratives cause conflict, a natural impulse is to encourage harmonization of historical understandings and apologies for agreed upon past misdeeds. For example, in the context of China–Taiwan relations and competing historical narratives, He (2010, 49) argues that "Cross-Strait reconciliation needs to begin with recognizing, rather than ignoring or covering up, this memory gap." More generally, He (2011, 1157) argues that "the harmonisation of national memories facilitates genuine reconciliation, while memory divergence resulting from national mythmaking hampers reconciliation." Other work on apologies in international relations supports this claim: Lind (2008, 159) argues that "denials of past aggression and atrocities fuel distrust and elevate threat perception." She documents the case of Japan following WWII, where Japan did not initially acknowledge wartime atrocities in Korea and China, and then later offered only lukewarm apologies, which were perceived as insincere, and this apparent lack of contrition contributed to distrust of Japan by Korea, China, and Australia.

However, efforts to agree upon the past, much less extract apologies for past crimes, may be inflammatory and lead to backlash (Rieff 2016). In Spain following the Spanish civil war (1936–39) and the subsequent Franco dictatorship (1939–75), the issue of who was to blame for past atrocities was so divisive that the parties agreed on an informal *pacto de olvido*, or pact of forgetting, which did not lead to reconciliation or shared understanding, but rather to a moratorium on discussion of the past (Shevel 2011).

Furthermore, some level of cooperation can often be restored in the absence of agreement or apology.[2] Lind argues that "International reconciliation is possible— even in the aftermath of horrendous crimes—with little or no contrition" (Lind 2008, 3). She shows that

although West Germany is widely held up as a textbook example of international contrition, the initial post-war response was "tepid" (Lind 2008, 102), and it was only in the 1960s under a left-leaning government that Bonn began to more thoroughly examine and accept responsibility for atrocities committed in WWII. Relations with France, however, had greatly improved by the late 1950s. Lustick (2006) makes a similar point in the case of German–Israeli relations. Adenauer's expression of repentance in 1951 was underwhelming and full of self-serving claims about the innocence of most Germans, but Germany and Israel nonetheless agreed on a reparations deal that helped rehabilitate Germany.

The question remains, how is cooperation best resumed in the face of diverging narratives about the past? We turn to the RPD to analyze this question.

THE REPEATED PRISONER'S DILEMMA

The RPD game is our starting point because it has been used to study cooperation in a wide variety of settings (Axelrod 1984; Fearon 1998; Milgrom, North, and Weingast 1990; Newton 2018; Oye 1986). In the Prisoner's Dilemma, there are joint gains to be had by cooperating, but each side has an incentive to exploit the other side. They can potentially overcome these incentives and cooperate provided that (a) the game is repeated, (b) they care about future payoffs, and (c) they are using strategies that give each other the proper incentives to cooperate. The importance of having the right strategy has led to extensive investigation of what strategies perform well under different conditions.

The payoff matrix for the game is illustrated in Table 1. There are two players with two choices.[3] Mutual cooperation yields R for each player, while mutual defection gives P. The temptation to exploit the other side is denoted T and the payoff for being exploited is S, and we posit the usual preference ordering for the Prisoner's Dilemma of $T > R > P > S$.[4] The unique Nash equilibrium of the game is (Defect, Defect), which is also the result of the selection of dominant strategies. Both sides could be made better off if they could switch to the (Cooperate, Cooperate) outcome, but each side has an incentive to defect if they

---

[2] A related phenomenon is "compartmentalization," in which states cooperate over some issues but not others. Compartmentalization can also sustain limited cooperation and is widely observed. However, it arises where there are multiple issues with varying degrees of shared interests, and the phenomenon we are interested in concerns a single issue area. Compartmentalization is perhaps understudied in international relations, but its opposite, issue linkage, has received attention (e.g., Poast 2012).

[3] The two-player assumption limits the application of the model to situations that are dyadic or can be reasonably decomposed into dyads, which is not always the case (Fordham and Poast 2016; Ruggie 1993).

[4] In addition, one usually assumes that $R > (T + S)/2$, so that the players prefer a steady cooperative relationship to alternating between exploiting the other side and being exploited.

TABLE 1. Prisoner's Dilemma

		Player 2	
		Cooperate	Defect
Player 1	Cooperate	R, R	S, T
	Defect	T, S	P, P

TABLE 2. Tit for Tat with an Implementation Error in Round 2

	Round				
	1	2	3	4	5 …
Player 1's choice	C	C	D	C	D…
Payoff	R	S	T	S	T …
Player 2's choice	C	D	C	D	C …
Payoff	R	T	S	T	S …

think the other side will cooperate, so this is not an equilibrium outcome.

In the RPD, the players repeat the game indefinitely.[5] In the RPD, there are many equilibria that support mutual cooperation under certain conditions (Abreu 1988). Perhaps the best known strategy is Tit for Tat (TFT), made famous by its success in computer tournaments staged by Axelrod (1984). TFT mandates that a player cooperate in the first round, and then from that point on do whatever the other side did in the previous round. On the equilibrium path, therefore, two players using TFT will cooperate forever. Axelrod (1984) identifies several qualities that make TFT perform well. Specifically, TFT is *nice*, or never the first to defect, *retaliatory*, in that it hits back for any defection, and *forgiving*, in that it will return to cooperation if the other side does.

For the bulk of his analyses, Axelrod initially assumed that there was no noise in the environment, so that each player knew perfectly what the other side had done and no one made mistakes. Scholars quickly began to consider the potential impact of relaxing this assumption. Downs, Rocke, and Siverson (1985, 139–40) introduced the distinction between what they called the problem of control and the problem of perception. The problem of control refers to the gap between intentions and actions, for instance, a state leader may order cooperation, but the bureaucracy may implement a defection by mistake. A problem of perception arises when there is a gap between how an act is perceived by each side. A leader may believe their state has cooperated, but the other side perceives it as a defection. We call the first type of problem an "implementation error" and the second type a "divergent perception."[6] Subsequent scholarship has analyzed both types of problems, but without always being conscious of the fact that solutions to one may not work for the other.

IMPLEMENTATION ERRORS

Mistakes happen, especially in combat or high-stress situations. For example, in the counterinsurgency campaigns in Iraq and Afghanistan, U.S. forces sometimes attacked noncombatant vehicles and homes, causing

unintended casualties. These accidental defections often led to retaliatory violence from the local population, either directly or through their increased support for insurgent forces (Shapiro and Condra 2012). Another example is the U.S. bombing of the Chinese embassy during the Kosovo war in 1999. Working from an outdated map, the US claims that it thought it was attacking a building associated with the Serbian regime, but in fact it hit the Chinese embassy, killing 3 and injuring 20. This caused a firestorm of protest in China and increased U.S.–Chinese tension for some time.

Implementation errors can be modeled by giving Nature a move after the two sides choose their strategies that map their strategy choices into a realized outcome. The final outcome of the round, once realized, is common knowledge between the players, so each player knows what happened. For instance, one can posit that there is a certain likelihood that cooperations become defections, and vice versa. "Negative noise" occurs when cooperations become defections, "positive noise" is when defections become cooperation, and neutral noise combines the two. Signorino (1996) argues that negative noise is more common in international settings, and it is certainly more problematic for sustaining cooperation, so we will focus on negative noise.

In the presence of negative implementation errors, TFT quickly gets bogged down in cycles of retaliation that then further break down into permanent defection (Downs, Rocke, and Siverson 1985; Molander 1985; Mueller 1987; Signorino 1996). The problem is illustrated in Table 2. Here, the players start off by cooperating, as TFT mandates. In the second round, they should cooperate again, since both sides cooperated to start with. But an implementation error occurs on player 2's part, so player 2 defects. In the next round, player 1 will retaliate by defecting and player 2 will cooperate. In round 4, player 2 will retaliate for player 1's prior defection, but player 1 will cooperate. This pattern of alternating cooperation and defection will continue until another negative implementation error occurs, after which both sides will defect forever.

Contrite Tit for Tat

One of the most elegant solutions to the problem of negative noise is Contrite Tit for Tat (CTFT) (Sugden 1986, 110). The key difference between CTFT and ordinary TFT is that whereas TFT retaliates against

---

[5] Analysts usually assume that the players discount future payoffs by a discount factor $\delta \in (0, 1)$.

[6] We use the term "divergent narratives" for real-world stories about blame for past wrongs that differ between parties to a conflict. We use the term "divergent perceptions" to indicate the game-theoretic representation of these divergent narratives, incidents in the game where an act of cooperation is perceived as a defection.

TABLE 3.   Contrite Tit for Tat with an Implementation Error in Round 2

	Round				
	1	2	3	4	5 …
Player 1's choice	C	C	D	C	C …
Payoff	*R*	*S*	*T*	*R*	*R* …
Standing	G	G	G	G	G …
Player 2's choice	C	D	C	C	C …
Payoff	*R*	*T*	*S*	*R*	*R* …
Standing	G	G	B	G	G …

any defection, CTFT refrains from retaliating if the other side is punishing it for defecting in the first place. That is, when CTFT accidentally defects it accepts the punishment in the next round and does not retaliate against the other side's retaliation. This keeps the retaliation short and promotes a quick return to cooperation.

CTFT makes use of the concept of good and bad standing. A player's standing is determined as follows. Both players start the game in good standing. Any player who cooperates will be in good standing in the next round. Defecting while only the other side is in bad standing is considered just punishment, and will keep a player in good standing in the next round. Defecting under any other circumstances results in being in bad standing in the next round. In terms of behavior, CTFT mandates cooperation unless only the other player is in bad standing, in which case it mandates defection.

CTFT's ability to short circuit cycles of retaliation caused by negative implementation errors is illustrated in Table 3. Here, we add a row for each player's standing, where G signifies good standing and B bad standing. The two players start the game in good standing and cooperate as the strategy mandates. As in the previous example, imagine an implementation error on the part of player 2 in round 2, so it defects where it should have cooperated. As a result, player 2 enters bad standing in round 3, while player 1 remains in good standing, because it cooperated in round 2. In round 3, player 2 cooperates, while player 1 punishes player 2 by defecting. This defection is permitted, however, because player 1 is in good standing and player 2 is in bad standing. Player 2's cooperation in round 3 restores it to good standing in the fourth round, so both sides are in good standing once more. If player 2 were playing TFT, they would defect because player 1 defected in round 3. However, with CTFT, player 1's third round defection is permitted, and cannot be retaliated against. Therefore, in the fourth round, both sides cooperate and the conflict is over.

CTFT has several good qualities and a successful track record in simulations. It forms a subgame perfect equilibrium with itself (unlike TFT) and is evolutionarily stable under implementation errors (Boyd 1989). Wu and Axelrod (1995) reprise Axelrod's tournament but add positive and negative implementation errors. They consider CTFT as well as a version of TFT that

only defects a certain percentage of the time (what we call PTFT). Both strategies handle noise well, but for higher levels of noise (over 1%), CTFT is superior and it prevails in evolutionary simulations. Finally, Signorino (1996), looking at environments with positive, mixed, and negative noise, shows CTFT fares very well against an array of other strategies, and comes to dominate the population in evolutionary simulations featuring negative noise. If by noise one means negative implementation errors, therefore, CTFT appears to be the best solution to the problem of noise in the RPD framework. As we will see below, the picture changes when one considers divergent perceptions.

DIVERGENT PERCEPTIONS

Consider a state embroiled in a civil war who suspects that the rebels are getting aid from a neighboring state (Schultz 2010). If the rebels launch a big attack, the state may think the neighbor must have helped them. The state cannot directly observe the level of aid that the rebels are getting from the neighbor. However, the more attacks the rebels launch, the more it looks like the neighbor is supporting them. But it may be the case that the neighbor is not actually supporting the rebels, and they are getting their support elsewhere. In that case, the state and the neighbor will have divergent perceptions of the neighbor's support for the rebels. The neighbor will know they did not support the rebels, but the state will think they did.[7]

In this example, there is a real level of support, known to the neighbor. However, the divergence of perception could be deeper in the sense that there is no "real" characterization of the strategy, or at least none that is agreed upon. The two sides may simply disagree over whether a certain act constitutes cooperation or defection, or whether it is "allowed" or not. For instance, when the Soviets and Cubans intervened in the Angolan civil war of the 1970s, was this a violation of the rules of superpower détente, and hence a defection? Or was it understood that superpower relations would be unaffected by regional conflicts, and so this kind of intervention was allowed? Dispute resolution mechanisms associated with agreements like the World Trade Organization serve to reduce this ambiguity by identifying legitimate defections that may be retaliated against. But in issue areas where there are not widely accepted norms of behavior or international regimes, what constitutes cooperation and defection may be subjective and contested (Krasner 1983).

Moreover, a perceived violation, or "cheating" is more injurious (the payoff is lower), even if the underlying action is the same. Individuals who believe the other side's conduct violates a rule or tacit understanding may be more aggrieved and be more prone to

---

[7] Axelrod briefly considers the problem of divergent perceptions (Axelrod 1984, 182–3). In a tournament with symmetrical noise, TFT once again comes out on top because the positive noise ends the retaliatory cycles caused by the negative noise.

TABLE 4. Contrite Tit for Tat with a Divergent Perception in Round 2

	Round				
	1	2	3	4	5 …
Player 1's choice	C	C	D	C	D …
Payoff	R	S	T	S	T …
Own standing	G	G	G	G	G …
Other standing	G	G	B	G	B …
Player 2's choice	C	C	C	D	C …
Payoff	R	R	S	T	S …
Own standing	G	G	G	G	G …
Other standing	G	G	G	B	G …

TABLE 5. Comparing Four Strategies: PTFT, TF2T, CAMP, and DLBA

Round		Strategy choices			
$t-2$	$t-1$	PTFT	TF2T	CAMP	DLBA
	T	C	C	C	C
	R	C	C	C	C
T	P	D/C*	C	C	C
R	P	D/C*	C	C	C
P	P	D/C*	D	C	C
S	P	D/C*	D	C	D
T	S	D/C*	C	D	C
R	S	D/C*	C	D	D
P	S	D/C*	D	D	D
S	S	D/C*	D	D	D

retaliate than if the conduct is injurious but within accepted bounds. A poker player who cheats will be treated differently from one who won the same amount of money fairly. Similarly, consider an arms control agreement that that each side will stop at 1,000 warheads. If one side builds up from 990 to 1,000, that may be viewed as acceptable by the other side. However, a buildup from 1,000 to 1,010 would be viewed as unacceptable, even though strategically almost identical.

To examine the problem of negative divergent perceptions, we consider a version of the game in which after the players choose their strategies, Nature determines how each side will perceive the move of the other side. We posit a probability ζ with which a cooperation by one side is perceived by the other side as defection. Defections are always perceived as such.

Since CTFT was able to cope with implementation errors, it is natural to wonder if it can cope with divergent perceptions as well. In fact, it breaks down quickly, in a manner similar to TFT under implementation errors. Consider the game history illustrated in Table 4. Here, we assume that players have access only to their half of the table, so player 1 sees the top half and player 2 sees the bottom half. The player's perceptions of each other's actions and standing can, therefore, differ. We have both sides begin the game as the strategy prescribes by cooperating. In the second round, however, assume that player 2's cooperation is perceived by player 1 as a defection, because player 1 receives an S payoff. In round 3, player 1, therefore, perceives player 2 to be in bad standing, but player 2, having cooperated, believes itself to be in good standing. Player 1 therefore defects, while player 2 cooperates. In round 4, since player 2 cooperated, player 1 now thinks player 2 is back to good standing, but player 2 thinks player 1's defection was unprovoked, and so thinks player 1 to be in bad standing. Therefore, player 2 will defect and player 1 will cooperate. In round 5, player 1 will think player 2's defection unjustified, and so it will view player 2 as returning to bad standing, while player 2 thinks player 1 has returned to good standing. Thus, player 1 defects and player 2 cooperates. Note that each player perceives itself to remain in good standing throughout the game, but sees the other side as alternating between good and bad standing. Each player perceives the other side's defections as unprovoked and its own defections to be justified punishments.

CTFT, therefore, breaks down in the face of divergent perceptions in the same way that TFT breaks down in the presence of implementation error. The concepts of good and bad standing require an allocation of blame for the initial defection. Players enter bad standing by being the first to defect. However, as we discussed above, actors are often reluctant to acknowledge their own defections as unprovoked, but easily see the other side's defections as unprovoked. Therefore, the required agreement on who is to blame for initial breakdown of cooperation is often absent. Put another way, CTFT readily offers apologies when it thinks it is to blame. However, it also demands apologies from the other side when it thinks the other side is to blame. If it does not get them, because the other side thinks it is innocent, it will not return to cooperation. CTFT is, therefore, a great solution to the problem of implementation errors, but it cannot cope with divergent perceptions.

DON'T LOOK BACK IN ANGER

Divergent perceptions pose a serious obstacle for cooperation, and yet they are not easily resolved, owing to both identity-related concerns and strategic incentives. How can cooperation be restored in such an environment? The problem is to cut short the cycles of blame and retaliation that arise from divergent perceptions. The basic solution is to find strategies that do not retaliate as much as TFT or CTFT, while still being retaliatory enough to deter exploitation by more predatory strategies. In particular, they need to be willing to end a cycle of punishment and restore cooperation, even if the other side has not apologized for starting the conflict.

We compare four strategies that have this characteristic in Table 5. The strategies choose actions based on the payoffs received in the previous two rounds, so the payoffs for rounds $t-1$ and $t-2$ are shown in the first two columns. The first two rows are the cases where the payoff in round $t-1$ was good (either T or R). The next

TABLE 6. Don't Look Back in Anger with a Divergent Perception in Round 2

	Round				
	1	2	3	4	5 …
Player 1's choice	C	C	D	C	C …
Payoff	R	S	T	S	R …
Player 2's choice	C	C	C	D	C …
Payoff	R	R	S	T	R …

four rows cover the cases where the payoff in round $t-1$ was P, for mutual punishment, and the payoff in round $t-2$ was T, R, P, or S. The last four rows are where the immediately preceding payoff was S. The Strategy Choice columns contain the choices mandated by each strategy given the payoff history.

Probabilistic Tit for Tat (PTFT), the first column among the Strategy Choices, only cares about the immediately preceding round ($t-1$). It starts by cooperating and then if it receives R or T (indicating that the other side cooperated), cooperates again. If it receives P or S, however, it defects with a certain probability (D/C* signifies the probabilistic defection). PTFT is, therefore, randomly more forgiving or generous than TFT (and is often called forgiving or generous TFT). Molander (1985) and Mueller (1987) show the benefits of this strategy under implementation errors, as do Wu and Axelrod (1995), although in their study CTFT does better. Bendor (1987) and Bendor, Kramer, and Stout (1991) look at a continuous strategy version with hidden actions, which produces divergent perceptions, and find that a version of PTFT which gives more than the player receives is quite successful in that environment as well. PTFT is, therefore, a leading candidate for handling both kinds of noise.

Next, we have Tit for Two Tats (TF2T), which only retaliates after two perceived defections rather than one. TF2T, therefore, retaliates when it receives payoffs in the previous two rounds of P, P, or P, S, or S, P, or S, S. This strategy is remarkably successful in some tournament environments, despite the fact that it is easily preyed on by Alternate (ALTR), which alternates between cooperation and defection and so never triggers retaliation from TF2T. TF2T is, however, very robust against highly pessimistic strategies that defect continuously, since TF2T will quickly give up on these strategies and respond in kind.

The next two strategies have not been previously studied, to the best of our knowledge. Cooperate after Mutual Punishment (CAMP) only retaliates if it gets the S payoff in the previous round, indicating that the other side has unilaterally exploited it, but not if it gets the P payoff, resulting from mutual defection. In effect, CAMP goes back to cooperation if there has been a round of mutual punishment, in the hopes that the other side will follow suit. CAMP will avoid being taken advantage of by Alternate, but will not fare well against strategies like All Defect (ALLD) and Grim Trigger (GRIM), that are unwilling to return to cooperation

after mutual defection. CAMP will also fall prey to cycles of alternating retaliation, unless the other side's cooperation or further noise gets it out of the rut.

Finally, Don't Look Back in Anger (DLBA) is a modified combination of TF2T and CAMP. Like TF2T, it retaliates if it thinks the other side has defected twice in a row. Like CAMP, it retaliates if it receives the S payoff. However, it makes two exceptions to these rules. First, if it has received the P payoff twice in a row, DLBA cooperates to try to break out of the cycle of mutual defection. Second, if it received the T payoff and then the S payoff, it also cooperates, to try to escape another round of alternating defection and cooperation. These two exceptions make DLBA able to quickly return to cooperation with itself or other strategies that respond well to cooperative gestures.

To illustrate the advantage of DLBA under divergent perceptions, consider Table 6. As before, both sides cooperate in round 1. In the second round, player 2 cooperates, but player 1 receives the S payoff, so there has been a divergence in perception. In round 3, player 1 will defect because it received the S payoff, but player 2 will cooperate because it received the R payoff. In round 4, player 1 will cooperate because it received the T payoff and player 2 will defect because it received the S payoff. At this point, however, player 1 will have received the T followed by the S payoff, so it will realize that a conflict cycle has happened. Following DLBA, it will, therefore, cooperate in round 5, as will player 2, because it got the T payoff in the previous round. This ends the conflict.

DLBA breaks out of conflict cycles caused by divergent perceptions, restoring cooperation between basically cooperative strategies. PTFT, TF2T, and CAMP can also accomplish this feat in their own ways. They would, therefore, seem to be appropriate candidates for study with computer tournaments that models such an environment.

A TOURNAMENT WITH TWO TYPES OF NOISE

The tournament pits the strategies against each other in a round-robin format. The payoffs were $T = 6$, $R = 5$, $P = 2$, and $S = 1$. This produces a sizable gap between mutual cooperation and mutual defection, which favors strategies that are able to return to cooperation. Each strategy plays each other strategy 50 times in games lasting 200 rounds and the average over the scores is recorded. These scores then form the basis of an evolutionary analysis over 1,000 generations, where each strategy receives a weighted score in each generation based on the current prevalence of the other strategies. This implies that strategies that do well against popular strategies will tend to do well, but strategies that do well only against rare strategies will not. Each strategy grows or shrinks from one generation to the next in proportion to how well it does in the current generation. A strategy's "fitness" depends on how far behind the leading strategy it is. The leader in any round, and any strategies with scores close to the leader's, will

TABLE 7.   The Round-Robin Scores of the Tournament with a 10% Likelihood of Implementation Error and Divergent Perception

	CTFT	PTFT	TF2T	CAMP	DLBA	GRIM	BTFT	ALTR	ALLD
CTFT	2.17	2.44	2.74	3.43	3.87	2.04	2.08	3.28	1.99
PTFT	2.38	2.52	3.28	3.39	3.86	1.99	2.23	3.30	1.95
TF2T	2.59	2.98	3.53	3.80	3.89	2.04	2.28	2.93	1.99
CAMP	3.20	3.09	4.26	3.46	3.98	1.57	2.85	3.19	1.52
DLBA	3.55	3.36	4.19	3.71	3.89	1.72	3.14	3.08	1.67
GRIM	2.06	2.19	2.07	3.71	3.17	2.03	2.05	3.70	2.00
BTFT	2.08	2.34	2.39	3.37	3.81	2.04	2.06	3.34	1.99
ALTR	3.07	3.02	4.28	3.36	3.82	1.55	2.88	3.24	1.53
ALLD	2.03	2.20	2.03	3.72	3.18	2.02	2.03	3.71	2.00

grow, while those further behind will diminish.[8] The environment featured both implementation errors and divergent perceptions. We varied the probability of both implementation errors and divergent perceptions, so we could see which strategy did best under each particular combination of parameters.

We compared nine strategies. CTFT is included because of its success in previous tournaments featuring implementation error.[9] We also include five other "nice" strategies which never intentionally defect first: PTFT,[10] TF2T, CAMP, DLBA, and GRIM, which defects forever in response to any defection. We also include three predatory strategies to introduce a downside to being overly forgiving. We examine Bad Tit for Tat (BTFT), which mostly plays TFT but sometimes defects at random to see if it can get away with it,[11] Alternate (ALTR) which alternates cooperation and defection, and ALLD, which defects in every round.

Table 7 shows the scores for each strategy against the others, when there is a 10% chance of both implementation errors and divergent perceptions. The entries in the table are the score the row strategy obtained when playing against the column strategy. Reading down a column, therefore, one can find what strategy did best against the column strategy. Reading across a row, one can find out what strategies the row strategy does well against. The scores are normalized to the scale of the payoffs in the normal form game, ranging between 1 and 6. The higher the score, the better the row strategy did against the column strategy. The highest scores were obtained by ALTR against TF2T (4.28) and by CAMP against TF2T (4.26), while the lowest was CAMP against ALLD (1.52) and ALTR against ALLD (1.53). This makes sense in that alternation works pretty well against forgiving strategies, and poorly against predatory ones.

Considering specific strategies, in the top row, CTFT did relatively badly against itself, better against PTFT, TF2T, CAMP, and DLBA, and badly against the predatory strategies except for ALTR. Basically, CTFT only does well when another strategy gets it out of retaliatory cycles. The next four strategies, PTFT, TF2T, CAMP, and DLBA, do well against each other, usually getting over 3 and sometimes even over 4. DLBA does the best against itself of any strategy (looking down the diagonal), and other strategies also do well against it (see the DLBA column) because of its forgiving nature. However, DLBA does poorly against GRIM and ALLD because it keeps trying to restore cooperation with these strategies which cannot be induced to do so, and they take advantage. By contrast, TF2T keeps these strategies in their place by defecting constantly, but scores for all are relatively low, around 2, in these pairings. ALTR does very well against TF2T, but less well against CAMP and DLBA, which do not forgive its alternating defections as much as TF2T. The bottom line is that the four strategies we focus on, PTFT, TF2T, CAMP, and DLBA, all do relatively well against themselves and each other, and so if there is a stable predominant strategy, or set of strategies, it is likely to be found among these four.

We can see how these scores influence the evolutionary sequence in Figure 1. Here again, we have a 10% chance of both implementation error and divergent perceptions. The first five strategies are shown in black, and the last four in gray. DLBA is the clear winner, but CAMP and TF2T are a significant share of the population to the end. The three of them together are able to keep the predatory strategies at bay, only ALTR enjoys a brief upward trajectory before declining to a low level. This pattern is representative of many cases, usually DLBA, TF2T, and CAMP are the main strategies in the population.

Finally, we may wonder how the results would differ for different probabilities of implementation error and divergent perception. We address this in Table 8. The rows show the likelihood of an implementation error, from 0% to 15%, and the columns show the likelihood of a divergent perception, also from 0% to 15%. Listed in the cells of the table is the strategy that emerged as most numerous after 1,000 rounds of evolution. DLBA emerges as the most numerous strategy under most

---

[8] The worst-performing strategy's percentage of the total declines at a rate of 3%, before renormalization. A floor of 0.1% is imposed for the population percentage of the worst-performing strategy, so no strategy goes completely extinct.

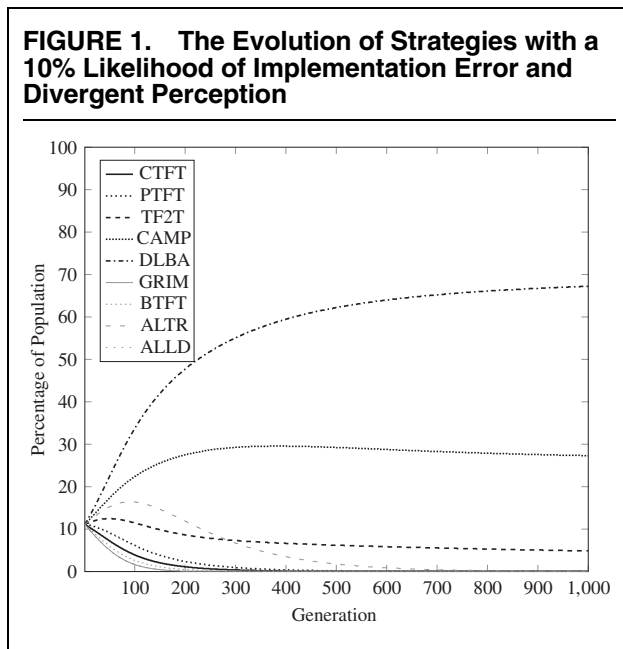[9] TFT is not, because of its known inability to cope with negative noise.

[10] PTFT forgives 5% of the time.

[11] BTFT cheats 5% of the time. This strategy is sometimes called "tester."

FIGURE 1. The Evolution of Strategies with a 10% Likelihood of Implementation Error and Divergent Perception

conditions, the exception being a region of low implementation error when TF2T wins out, and a region of high implementation error and divergent perceptions, when CAMP sometimes wins. However, examining the evolutionary trajectories in these cases (available at the Dataverse; see Herrera and Kydd 2023) indicates that PTFT, CAMP, and DLBA remain in the population, but TF2T emerges on top in this region.

To sum up, when noise takes the form of divergent preferences as well as implementation errors, strategies that retaliate for every offense, like TFT, or demand that the other side be the first to cooperate after a defection, like CTFT, are outperformed by less implacable strategies, like PTFT, TF2T, CAMP, and DLBA. Each in its own way is more hesitant to retaliate, carving out exceptions or retaliating with less than certainty. Coping with divergent preferences, therefore, requires a somewhat more pragmatic approach that nonetheless retaliates enough to deal with predation from the more aggressive strategies.

US–CHINA RELATIONS AFTER TIANANMEN

The relationship between game-theoretic modeling and empirical work is a perennial subject of debate. Advocates of "positive political theory" argue that game theory provides a norm-free positivist theory of political behavior, suitable for deriving hypotheses that can be empirically tested (Amadae and Bueno de Mesquita 1999; Morton 1999). This has led to a large literature linking formal models and statistical data analysis, promoted by the Empirical Implications of Theoretical Models program. Critics point out that a hard distinction between normative and positive theory is untenable; social choice theory, for instance, is devoid of empirical implications and firmly wedded to a normative commitment to democracy (Knight and Johnson 2015).

Our position is that game-theoretic modeling has both a normative and a positive component. Game theory can be helpful for answering normative questions like "what is the best course of action to achieve certain goals?" or "are there more efficient outcomes that could be obtained by structuring the situation in a certain way?" It is also able to offer predictions about behavior, if we are willing to assume that the individuals we study are more or less rational and the situations they face resemble the games sufficiently—two conditions that apply to all theories, not just formal models. In our case, we can make the normative claim that because strategies like Don't Look Back in Anger and Cooperate after Mutual Punishment are successful in RPD games featuring implementation errors and divergent perceptions, they may be good options for policymakers thinking about how to return to cooperation in real-world settings with persistent disagreements about the past. We can also make the positive empirical prediction that some actors who implicitly or explicitly realize that fact may put such strategies into effect, and we can look for evidence to support that claim.

Moreover, we see great benefit in linking game-theoretic work with qualitative research, with the goal of using formal modeling to clarify incentives and strategic choices in historical episodes (Bates et al. 1998; Goemans and Spaniel 2016; Kydd 2005; Lorentzen, Fravel, and Paine 2017). In this regard, we hope this article sheds light on a class of empirical cases, namely where there are divergent perceptions of the past and unresolved disagreement about who is to blame, and where parties nevertheless resume cooperation. In the case study that follows, we provide an empirical example of the Don't Look Back in Anger strategy in a context of divergent perceptions, with the goals of providing insight into the incentives for and consequences of the strategies at work in US–China relations.

China's relations with its neighbors and with the US are fertile ground for investigating conflicting narratives. For instance, China's claims to the South China Sea rest on historical narratives that are sharply disputed by surrounding states (Lim 2016) and opposing narratives of past territorial rights and competing accusations of which side is threatening versus peaceful are used to justify military maneuvers in the region (Weissmann 2019). We will focus on the 1989 Tiananmen Square incident and its aftermath in China's foreign relations, which we suggest contains divergent narratives of an important event, a clear punishment phase in the form of sanctions, and then a return to cooperation without contrition or agreement on what happened in the past by either party.[12] The US, we argue, pursued a strategy much like DLBA, inflicting a

---

[12] The distinction between narratives and interests may seem blurry in the real world. At a game-theoretic level, interests are the payoffs in the game matrix and narratives are the beliefs about what the other player did in the previous round. In this case, China's interests are its preferences over Taiwan, the South China Sea, and so forth. The narrative we focus on is how China understands the causes of Tiananmen.

TABLE 8. The Most Popular Strategies after 1,000 Generation

IE%	The probability of a divergent perception (%)															
	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	TF2T	TF2T	TF2T	TF2T	TF2T	TF2T	DLBA	TF2T	DLBA	TF2T	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
1	DLBA	TF2T	TF2T	TF2T	TF2T	TF2T	DLBA	TF2T	TF2T	TF2T	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
2	DLBA	DLBA	TF2T	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
3	DLBA	DLBA	DLBA	TF2T	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
4	DLBA	DLBA	DLBA	TF2T	DLBA	DLBA	TF2T	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
5	DLBA	DLBA	DLBA	TF2T	TF2T	TF2T	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
6	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
7	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
8	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
9	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
10	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
11	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
12	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
13	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	CAMP
14	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA
15	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	DLBA	CAMP

short punishment phase followed by a return to cooperation.

In the spring of 1989, Mikhail Gorbachev's policies of perestroika and glasnost were in full swing and the future of the Communist Party of the USSR was being actively debated both in the USSR and outside. In Eastern Europe, "Round Table Talks" between the Polish Communist Party and the Solidarity protest movement were underway. The longtime head of the Communist Party in Hungary, János Kádár, had been removed in 1988 and in March 1989 thousands of people had come out to call for democracy in Budapest. By early April, the Round Table talks in Poland had produced an agreement that there would be some free elections and the creation of the office of the Presidency to counter the Communist-led parliament (Service 2015). It was in this context that China found itself in the midst of massive protests in April 1989 at Tiananmen square.

On April 15, the pro-reform Communist general secretary Hu Yaobang, who had earlier been forced to resign in 1987, died of a heart attack. The next day there was a small commemoration for him and calls for the Chinese government to reconsider his legacy. A week later, the day before his funeral, the third major democracy protest in less than a decade occurred when more than 100,000 students marched on Tiananmen Square, despite the square being officially closed (Calhoun 1989). On April 26, 1989, the official newspaper of the Chinese Communist Party (CCP), *The People's Daily*, issued a front-page editorial titled "It is necessary to take a clear-cut stand against disturbances." However, rather than quelling the unrest, the article spurred further protests by students, and by May there was a student-led hunger strike, calls for democracy, and protests in more than four hundred cities (Brown 2021). There was a great deal of negotiation internally among Chinese Communist Party elites, and with student leaders, but protests continued even during Mikhail Gorbachev's visit in mid-May.

Eventually, however, the regime decided to crack down. Martial law was declared on May 20 and by early June the leadership of the CCP was preparing to take decisive action. PLA troops amassed around Tiananmen on the evening of June 3, and in the early hours of June 4 the Army cleared the square with infantry troops, gunfire, and tanks. The Chinese government initially reported the death toll at 241, and 7,000 wounded, but this estimate was questioned by many other sources which estimated the death toll to be higher, ranging from several hundreds to several thousands (Brook 1998, 151–69). The lone man confronting a column of tanks the next day, "Tank Man," became an iconic picture in the West, and a symbol of resistance to dictatorship.

The CCP leadership's narrative of the Tiananmen Square "incident" is that it was an attack on China and the Chinese system by a very small group of domestic opponents inspired and supported by the West. In terms of the model, the CCP leadership perceived Tiananmen as an unprovoked defection by the West, and a very serious one. Documents that subsequently

made their way out of China uncover the deliberations leading up to the crackdown (Nathan 2001). They are worth quoting at length, because they provide such a clear picture of Chinese thinking and how different it is from Western perceptions.

In the debate over imposing martial law, Bo Yibo, a senior leader who initially supported reforms, said,

> The whole imperialist Western world wants to make socialist countries leave the socialist road and become satellites in the system of international monopoly capitalism. The people with ulterior motives who are behind this student movement have support from the United States and Europe and from the KMT [Kuomintang] reactionaries in Taiwan. Members of the overseas Chinese Alliance for Democracy, which we have declared to be an illegal and reactionary organization, not only voice support for the student movement but openly admit that they advise the students and even plan how to reenter China and meddle directly. …So you see, it was no accident that the student movement turned into turmoil. (Nathan 2001, 24–5)

In the final debate before clearing the square, the hardline Premier Li Peng argued,

> When the turmoil began employees of the U.S. embassy started to collect intelligence aggressively. Some of them are CIA agents. Almost every day, and especially at night, they would go and loiter at Tiananmen or at schools such as Peking University and Beijing Normal. They have frequent contact with leaders of the AFS [Autonomous Federation of Students] and give them advice. The Chinese Alliance for Democracy, which has directly meddled in this turmoil, is a tool the United States uses against China. This scum of our nation, based in New York, has collaborated with the pro-KMT Chinese Benevolent Association to set up a so-called Committee to Support the Chinese Democracy Movement. They also gave money to leaders of the AFS. As soon as the turmoil started, KMT intelligence agencies in Taiwan and other hostile forces outside China rushed to send in agents disguised as visitors, tourists, businessmen, and so on. They have tried to intervene directly to expand the so-called democracy movement into an all-out "movement against communism and tyranny." There is evidence that KMT agents from Taiwan have participated in the turmoil in Beijing, Shanghai, Fujian, and elsewhere. …It is becoming increasingly clear that the turmoil has been generated by a coalition of foreign and domestic reactionary forces and that their goals are to overthrow the Communist Party and to subvert the socialist system. (Nathan 2001, 31)

Deng Xiaoping, still the most powerful CCP leader, agreed,

> The causes of this incident have to do with the global context. The Western world, especially the United States, has thrown its entire propaganda machine into agitation work and has given a lot of encouragement and assistance to the so-called democrats or opposition in China—people who in fact are the scum of the Chinese nation. This is the root of the chaotic situation we face today. (Nathan 2001, 32)

Two days after the square was cleared, he reflected,

> In the future, whenever it might be necessary, we will use severe measures to stamp out the first signs of turmoil as soon as they appear. This will show that we won't put up with foreign interference and will protect our national sovereignty. (Nathan 2001, 43)

Two weeks later, Deng convened the Politburo and other key officials in order to unify the party elite in support of the use of force in Tiananmen. There was unchallenged agreement that the West was to blame for the Tiananmen uprising. Xu Xiangqian, a retired military officer argued that the disturbances were the result of an alliance between domestic and foreign forces who wanted to overthrow the regime and reduce China to "vassalage" to the west. Peng Zhen agreed and argued that the domestic reactionaries and their Western supporters wanted to establish "capitalist dictatorship." Song Renqiong recalled John Foster Dulles' prediction that the third or fourth post-revolutionary generation would restore capitalism and urged his colleagues not to let this prophecy come to pass. Vice President Wang Zhen recalled previous western anti-communist interventions, starting with the Bolshevik revolution, and argued that the West was now trying to achieve the same goal "the easy way" (Nathan 2019).

The Chinese leadership's understanding of Tiananmen, in sum, is that it was an unprovoked attack by the West, led by the US, with the aim of overthrowing the Chinese regime and reducing China to a condition of servitude (Sarotte 2012). This narrative was probably sincerely held by many in the leadership, and for others it had obvious strategic advantages in terms of remaining in power and rallying domestic support.

The American perspective on Tiananmen was, of course, quite different. The U.S. narrative held that the student movement was a benign domestic development to bring democracy and other universal international values to China. The Chinese crackdown was, therefore, an unprovoked attack on those values that should be strongly condemned and punished. The idea that the US "started it" was not even entertained in public statements. On June 5, the morning after the use of force to clear the square, U.S. President George H. W. Bush gave a press conference in which he condemned the violence by the PLA and announced a series of punitive measures. Bush portrayed the Chinese students as human rights advocates worthy of international support, oblivious to the emerging Chinese narrative which deemed that goal and that support to be a mortal threat.

At the same time, Bush wanted to maintain ties and not let the relationship be derailed by the crackdown or the U.S. response. Congressional leaders were more critical of China and advocated harsher measures. The punishment debate was, therefore, marked by a certain tension between the President and Congress (Skidmore and Gates 1997). In the same speech on June 5, President Bush ordered the suspension of all U.S. government and commercial weapons sales to China and a suspension of

US–China military visits. He also called for allowing Chinese students in the US to extend their stays and a reassessment of the bilateral relationship including suspension of all foreign aid, export licenses, and loans and grants in international financial institutions to China. Shortly thereafter, however, Bush sent his national security adviser, Brent Scowcroft, to China on a secret trip in order to limit the damage.[13]

The U.S. Congress followed up with further sanctions in three bills: the 1989 International Development and Finance Act, which banned the U.S. Export-Import Bank from providing loans, insurance, credits, and other financing to China; a 1990 appropriations act that prohibited funding and export licenses for U.S. satellites on Chinese-built launch vehicles; and the Foreign Relations Authorization Act, 1990 and 1991, which continued President Bush's suspension of Overseas Private Investment Corporation financing for China as well as suspension of satellites and defense-related export licenses, and the U.S. Trade and Development Agency's export licenses and financing of crime-related technology to China (Rennack 2016).

The punishment phase, however, turned out to be short. The U.S. president was only legally allowed to end these sanctions if he reported to Congress that China had met various criteria for political reform, including human rights standards. In the years following Tiananmen, there was a heated debate as to whether China's "Most Favored Nation" (MFN) trade status should be renewed or revoked.[14] President Bill Clinton went so far as to sign an Executive Order (on May 28, 1993), supported by Senate Majority Leader George Mitchell and Representative Nancy Pelosi, to link China's MFN status to its human rights performance (Li 2014). Rather than conceding, China made a show of arresting human rights activists shortly after a contentious visit by U.S. Secretary of State Warren Christoper in 1994. Eventually, as with the other sanctions, the US gave up and in May 1994 Clinton announced that the US would delink China's human rights performance from MFN renewal. In 2000, Clinton welcomed China into the World Trade Organization.

Although the punishment phase came to an end, the Chinese and U.S. narratives over what happened and who is to blame remain diametrically opposed. Chinese propaganda has continued to reflect many of the themes put forward by the hard-liners in the summer of 1989, namely that the use of force was justified in order to defend the stability and sovereignty of China against external subversion. According to Nathan, "one can draw a direct line connecting the ideas and sentiments expressed at the June 1989 Politburo meeting to the hard-line approach to reform and dissent that President Xi Jinping is following today" (Nathan 2019,

81).[15] Similarly, the views of the US on Tiananmen have remained remarkably consistent over the last three decades. In June 2020, the bipartisan and bicameral Congressional–Executive Commission on China issued a statement on the Tiananmen anniversary saying "On this day we remember the courage and sacrifice of the students, workers, and others who were peacefully protesting in the streets of Beijing and over 400 other cities to call for democracy, human rights, and an end to corruption. Sadly, the Chinese Communist Party dispersed these peaceful protesters by using military force in Tiananmen Square, crushing their peaceful demands for rights and reform."[16]

Returning once more to the framework of the model, according to CTFT, a return to cooperation would occur only if the guilty party took responsibility and submitted to unilateral punishment in order to get back into "good standing." In contrast, more forgiving strategies like DLBA and CAMP require no such acknowledgement. For these strategies, it does not matter whether disagreement over the past is a function of identity concerns, strategic incentives to promote a particular view, or a combination of both—and it seems that there is indeed a mixture of sincere and insincere posturing in the statements about Tiananmen by both the US and China. The important thing is that in this case China and the US have never agreed on whether the uprising was a foreign plot or whether force was justified to put it down. No one apologized or acknowledged error, and yet the punishment (i.e., the sanctions and other punitive measures that were imposed by the US and EU)[17] were ended, heralding a return to cooperation.

Although we lack space to consider a range of alternative explanations, one alternative framework worth considering is the literature on stochastic shocks to the cost of cooperation. Downs and Rocke (1995) and Rosendorff and Milner (2001) develop models of trade cooperation where the two sides face variable domestic pressures that sometimes make cooperation too costly. The question is how to sustain cooperation as best as may be despite these shocks. Downs and Rocke, like us, focus on short punishment phases, and Milner and Rosendorff model "escape clauses" whereby states pay a cost and get an excused defection when they need one. This framework may appear to fit the China–Tiananmen case well, since the student movement was unanticipated and the cost of not crushing it was perceived as high by the Chinese leadership. However, China never conceded that its actions constituted defection of any kind, and certainly never accepted any penalty for invoking an escape clause. Instead, each side developed deeply held narratives of the other side's guilt, but returned to cooperation anyway.

---

[13] https://www.nytimes.com/1989/12/19/world/2-us-officials-went-to-beijing-secretly-in-july.html.
[14] Prior to China's entry into the World Trade Organization in 2001, China's MFN status was subject to annual review in accordance with the 1974 Jackson–Vanik amendment and frequently opposed by human rights groups and import competing trade associations.

[15] For a Chinese perspective on this question, see Chen (2003).
[16] https://www.cecc.gov/media-center/press-releases/statement-on-the-31st-anniversary-of-the-violent-repression-of-the.
[17] The European perspective was very similar to the US. In 1989, the European Economic Community condemned the Chinese government's response to the protesters, and the EU has continued over the years to commemorate the Tiananmen events.

CONCLUSION

Seemingly intractable disagreements over interpretations of the past are common in cases of conflict. These disagreements inhibit cooperation by obfuscating blame and impeding contrition, and they undermine strategies for achieving and maintaining cooperation in the RPD models that depend on common interpretations of the past. Strategies that are more forgiving, like Don't Look Back in Anger, Cooperate after Mutual Punishment, and Tit for Two Tats, provide a possible way to cope with such situations where there are deep disagreements over culpability. These strategies, unlike Contrite Tit for Tat, do not demand apologies from other states to return to cooperation. They may instead provide a model for dealing with conflicts in which identities or interests preclude common understandings of past events.

An interesting question for further research is whether the level of cooperation that can be achieved by such strategies under such conditions is less profound than that which can be achieved when states can agree on blame. A common observation in the German–Japanese comparison is that because Germany acknowledged its responsibility for the war much more fully than the Japanese did, this allowed for a much deeper level of cooperation among European states. Germany is now fully integrated with its neighboring states in myriad economic, security, and cultural relations. Japan certainly cooperates with its neighbors on economic and other matters, but the level of reconciliation is nowhere near as strong and the level of cooperation is arguably much shallower.

However, it may still be counter-productive to press for acknowledgements of guilt when national narratives (Li 2014) or other strategic or material incentives of the two sides preclude such admissions (Berger 2012). In these cases, pressing a side to admit guilt may simply deepen the conflicting narratives and lead away from cooperation. Li (2014) argues in the case of China, where there are conflict-avoidant cultural norms at play, explicitly pressing China to apologize will have the opposite effect and for that reason the sanctioning approach by the US has been a failure. To the extent that audience costs are at work, overt calls for apologies may similarly backfire, though for different reasons.

Although we have applied the analysis mainly to cases of states in the context of international relations, the analysis could also apply to individuals and groups or other situations in which there is a need to return to cooperation in the context of unresolved disagreement over the past. The recent upswing in partisan polarization suggests that narratives within democratic countries, in particular the US, are growing more and more divergent, and the role of impartial adjudicating institutions is increasingly under strain (Iyengar et al. 2019).

DATA AVAILABILITY STATEMENT

The computer program that produced the findings of this study and the complete findings are openly available at the American Political Science Review Dataverse: https://doi.org/10.7910/DVN/XY9TOC.

CONFLICT OF INTEREST

The authors declare no ethical issues or conflicts of interest in this research.

ETHICAL STANDARDS

The authors affirm this research did not involve human subjects.

REFERENCES

Abdelal, Rawi, Yoshiko Herrera, Alastair Iain Johnston, and Rose McDermott. 2009. *Measuring Identity: A Guide for Social Scientists*. Cambridge: Cambridge University Press.

Abreu, Dilip. 1988. "On the Theory of Infinitely Repeated Games with Discounting." *Econometrica* 56 (2): 383–96.

Amadae, S. M., and Bruce Bueno de Mesquita. 1999. "The Rochester School: The Origins of Positive Political Theory." *Annual Review of Political Science* 2: 269–95.

Aumann, Robert J., Michael B. Maschler, and Richard E. Stearns. 1995. *Repeated Games with Incomplete Information*. Cambridge, MA: MIT Press.

Axelrod, Robert. 1984. *The Evolution of Cooperation*. New York: Basic Books.

Bates, Robert H., Avner Grief, Margaret Levi, Jean-Laurent Rosenthal, and Barry R. Weingast. 1998. *Analytic Narratives*. Princeton, NJ: Princeton University Press.

Bendor, Jonathan. 1987. "In Good Times and Bad: Reciprocity in an Uncertain World." *American Journal of Political Science* 31 (3): 531–58.

Bendor, Jonathan, Roderick M. Kramer, and Suzanne Stout. 1991. "When in Doubt … Cooperation in a Noisy Prisoner's Dilemma." *Journal of Conflict Resolution* 35 (4): 691–719.

Berger, Thomas U. 2012. *War, Guilt and World Politics After World War II*. Cambridge: Cambridge University Press.

Boyd, Robert. 1989. "Mistakes Allow Evolutionary Stability in the Repeated Prisoner's Dilemma Game." *Journal of Theoretical Biology* 136 (1): 47–56.

Brook, Timothy. 1998. *Quelling the People: The Military Suppression of the Beijing Democracy Movement*. Redwood City, CA: Stanford University Press.

Brown, Jeremy. 2021. *June Fourth: The Tiananmen Protests and Beijing Massacre of 1989*. Cambridge: Cambridge University Press.

Calhoun, Craig. 1989. "Revolution and Repression in Tiananmen Square." *Society* 26 (6): 21–38.

Chen, Youwei. 2003. "China's Foreign Policy Making as Seen through Tiananmen." *Journal of Contemporary China* 12 (37): 715–38.

Cruz, Consuelo. 2000. "Identity and Persuasion: How Nations Remember Their Pasts and Make Their Futures." *World Politics* 52 (3): 275–312.

Diesen, Glenn, and Conor Keane. 2017. "The Two-Tiered Division of Ukraine: Historical Narratives in Nation-Building and Region-Building." *Journal of Balkan and Near Eastern Studies* 19 (3): 313–29.

Downs, George W., and David M. Rocke. 1995. *Optimal Imperfection: Domestic Uncertainty and Institutions in International Relations*. Princeton, NJ: Princeton University Press.

Downs, George W., David M. Rocke, and Randolph M. Siverson. 1985. "Arms Races and Cooperation." *World Politics* 38 (1): 118–46.

Fearon, James D. 1998. "Bargaining, Enforcement and International Cooperation." *International Organization* 52 (2): 269–305.

Finkel, Evgeny. 2010. "In Search of Lost Genocide: Historical Policy and International Politics in Post-1989 Eastern Europe." *Global Society* 24 (1): 51–70.

Fordham, Benjamin, and Paul Poast. 2016. "All Alliances Are Multilateral: Rethinking Alliance Formation." *Journal of Conflict Resolution* 60 (5): 840–65.

Glaeser, Edward L. 2005. "The Political Economy of Hatred." *Quarterly Journal of Economics* 120 (1): 45–86.

Goemans, Hein, and William Spaniel. 2016. "Multimethod Research: A Case for Formal Theory." *Security Studies* 25 (1): 25–33.

He, Yinan. 2010. "Competing Narratives, Identity Politics and Cross-Strait Reconciliation." *Asian Perspective* 34 (4): 45–83.

He, Yinan. 2011. "Comparing Post-War (West) German-Polish and Sino-Japanese Reconciliation: A Bridge Too Far." *Europe-Asia Studies* 63 (7): 1157–94.

Herrera, Yoshiko M., and Andrew H. Kydd. 2023. "Replication Data for: Don't Look Back in Anger: Cooperation Despite Conflicting Historical Narratives." Harvard Dataverse. https://doi.org/10.7910/DVN/XY9TOC.

Herwig, Holger H. 1987. "Clio Deceived: Patriotic Self-Censorship after the Great War." *International Security* 12 (2): 5–44.

Iyengar, Shanto, Yphtach Lelkes, Matthew Levendusky, Neil Malhotra, and Sean J. Westwood. 2019. "The Origins and Consequences of Affective Polarization in the United States." *Annual Review of Political Science* 22: 129–46.

Kacowicz, Arie M. 2005. "Rashomon in the Middle East: Clashing Narratives, Images, and Frames in the Israeli–Palestinian Conflict." *Cooperation and Conflict* 40 (3): 343–60.

Knight, Jack, and James Johnson. 2015. "On Attempts to Gerrymander 'Positive' and 'Normative' Political Theory: Six Theses." *Good Society* 24 (1): 30–48.

Krasner, Stephen D., ed. 1983. *International Regimes*. Ithaca, NY: Cornell University Press.

Kydd, Andrew H. 2005. *Trust and Mistrust in International Relations*. Princeton, NJ: Princeton University Press.

Li, Yitan. 2014. "US Economic Sanctions against China: A Cultural Explanation of Sanction Effectiveness." *Asian Perspective* 38 (2): 311–35.

Liang, Ce. 2018. "The Rise of China as a Constructed Narrative: Southeast Asia's Response to Asia's Power Shift." *Pacific Review* 31 (3): 279–97.

Lieber, Keir A. 2007. "The New History of World War I and What It Means for International Relations Theory." *International Security* 32 (2): 155–91.

Lim, Kheng Swe. 2016. "China's Nationalist Narrative of the South China Sea: A Preliminary Analysis." In *Power Politics in Asia's Contested Waters*, eds. Enrico Fels andd Truong-Minh Vu, 159–72. Berlin: Springer.

Lind, Jennifer. 2008. *Sorry States: Apologies in International Relations*. Ithaca, NY: Cornell University Press.

Lorentzen, Peter, M. Taylor Fravel, and Jack Paine. 2017. "Qualitative Investigation of Theoretical Models: The Value of Process Tracing." *Journal of Theoretical Politics* 29 (3): 467–91.

Lukyanov, Fyodor. 2016. "Putin's Foreign Policy: The Quest to Restore Russia's Rightful Place." *Foreign Affairs* 95 (3): 30–7.

Lustick, Ian S. 2006. "Negotiating Truth: The Holocaust, *Lehavdil*, and *Al-Nakba*." *Journal of International Affairs* 60 (1): 51–77.

Mailath, George J., and Larry Samuelson. 2006. *Repeated Games and Reputations: Long-Run Relationships*. Oxford: Oxford University Press.

McDermott, Rose. 2009. "Psychological Approaches to Identity: Experimentation and Application." In *Measuring Identity: A Guide for Social Scientists*, eds. Rawi Abdelal, Yoshiko Herrera, Alastair Iain Johnston, and Rose McDermott, 345–67. Cambridge: Cambridge University Press.

Milgrom, Paul R., Douglass C. North, and Barry R. Weingast. 1990. "The Role of Institutions in the Revival of Trade: The Medieval Law Merchant, Private Judges and Champaign Fairs." *Economics and Politics* 2 (1): 1–23.

Miller, Steven E., Sean M. Lynn-Jones, and Stephen Van Evera (Eds.). 1991. *Military Strategy and the Origins of the First World War. International Security Readers*. Princeton, NJ: Princeton University Press.

Molander, Per. 1985. "The Optimal Level of Generosity in a Selfish Uncertain Environment." *Journal of Conflict Resolution* 29 (4): 611–18.

Morton, Rebecca. 1999. *Methods and Models: A Guide to the Empirical Analysis of Formal Models in Political Science*. Cambridge: Cambridge University Press.

Mueller, Ulrich. 1987. "Optimal Retaliation for Optimal Cooperation." *Journal of Conflict Resolution* 31 (4): 692–724.

Nathan, Andrew J. 2001. "The Tiananmen Papers." *Foreign Affairs* 80 (1): 2–49.

Nathan, Andrew J. 2019. "The New Tiananmen Papers: Inside the Secret Meeting That Changed China." *Foreign Affairs* 98: 80–91.

Newton, Jonathan. 2018. "Evolutionary Game Theory: A Renaissance." *Games* 9 (31): 1–67.

Oye, Kenneth A., ed. 1986. *Cooperation under Anarchy*. Princeton, NJ: Princeton University Press.

Poast, Paul. 2012. "Does Issue Linkage Work? Evidence from European Alliance Negotiations, 1860–1945." *International Organization* 66 (2): 277–310.

Rennack, Dianne E. 2016. "China: Economic Sanctions." Congressional Research Service Report, R44605, August 22. https://crsreports.congress.gov.

Rieff, David. 2016. *In Praise of Forgetting: Historical Memory and Its Ironies*. New Haven, CT: Yale University Press.

Röhl, John C. G. 2015. "Goodbye to All That (Again)? The Fischer Thesis, the New Revisionism and the Meaning of the First World War." *International Affairs* 91 (1): 153–66.

Rosendorff, B. Peter, and Helen Milner. 2001. "The Optimal Design of International Trade Institutions: Uncertainty and Escape." *International Organization* 55 (4): 829–58.

Ross, Dennis. 2005. *The Missing Peace: The Inside Story of the Fight for Middle East Peace*. New York: Farrar, Straus and Giroux.

Ross, Marc Howard. 2001. "Psychocultural Interpretations and Dramas: Identity Dynamics in Ethnic Conflict." *Political Psychology* 22 (1): 157–78.

Rotberg, Robert I. 2006. *Israeli and Palestinian Narratives of Conflict: History's Double Helix*. Bloomington: Indiana University Press.

Ruggie, John Gerard, ed. 1993. *Multilateralism Matters: The Theory and Praxis of an Institutional Form*. New York: Columbia University Press.

Sarotte, Mary Elise. 2012. "China's Fear of Contagion: Tiananmen Square and the Power of the European Example." *International Security* 37 (2): 156–82.

Sarotte, Mary Elise. 2014. *1989: The Struggle to Create Post-Cold War Europe—Updated Edition*. Princeton, NJ: Princeton University Press.

Schultz, Kenneth A. 2010. "The Enforcement Problem in Coercive Bargaining: Interstate Conflict over Rebel Support in Civil Wars." *International Organization* 64 (2): 281–312.

Service, Robert. 2015. *The End of the Cold War: 1985–1991*. New York: PublicAffairs.

Shapiro, Jacob N., and Luke Condra. 2012. "Who Takes the Blame? The Strategic Effects of Collateral Damage." *American Journal of Political Science* 56 (1): 167–87.

Shevel, Oxana. 2011. "The Politics of Memory in a Divided Society: A Comparison of Post-Franco Spain and Post-Soviet Ukraine." *Slavic Review* 70 (1): 137–64.

Shifrinson, Joshua R. Itzkowitz. 2016. "Deal or No Deal? The End of the Cold War and the US Offer to Limit NATO Expansion." *International Security* 40 (4): 7–44.

Signorino, Curtis S. 1996. "Simulating International Cooperation under Uncertainty." *Journal of Conflict Resolution* 40 (1): 152–205.

Skidmore, David, and William Gates. 1997. "After Tiananmen: The Struggle over U.S. Policy toward China in the Bush Administration." *Presidential Studies Quarterly* 27 (3): 514–39.

Smith, Anthony D. 1996. "Memory and Modernity: Reflections on Ernest Gellner's Theory of Nationalism." *Nations and Nationalism* 2 (3): 371–88.

Straus, Scott. 2015. *Making and Unmaking Nations: War, Leadership and Genocide in Modern Africa*. Ithaca, NY: Cornell University Press.

Sugden, Robert. 1986. *The Economics of Rights, Cooperation and Welfare*. Hoboken, NJ: Blackwell.

Suny, Ronald Grigor. 2009. "Truth in Telling: Reconciling Realities in the Genocide of the Ottoman Armenians." *American Historical Review* 114 (4): 930–46.

Weissmann, Mikael. 2019. "Understanding Power (Shift) in East Asia: The Sino-US Narrative Battle about Leadership in the South China Sea." *Asian Perspective* 43 (2): 223–48.

Wu, Jianzhong, and Robert Axelrod. 1995. "How to Cope with Noise in the Iterated Prisoner's Dilemma." *Journal of Conflict Resolution* 39 (1): 183–89.