Power and Cooperation

Thomas Schwartz[1]

[1] Department of Political Science, UCLA, Los Angeles, CA, USA

Correspondence: Thomas Schwartz, Department of Political Science, UCLA, Los Angeles, CA, USA. Tel: 1-310-825-1972. E-mail: tschwartz@polisci.ucla.edu

Abstract

The pure realist picture of world politics combines the demand for power with the problem of cooperation. In a more up-to-date, less austere variant, power is no unqualified asset: a state can profit from a unilateral loss of power, all else equal. And mutual cooperation is often achievable, thanks in great measure to the shadow of the future. My newer, richer variant enlarges and combines those two features of anarchy: sometimes a loss of power by a single state not only is advantageous to that state but enhances mutual cooperation, and when that happens the motives to cooperate—which turn out to be quite varied—only have to work in a one-sided way.

Keywords: asymmetry, cooperation, paradox, power, prisoners' dilemma

1. Introduction

Realism of the simplest sort depicts world politics as Hobbesian anarchy with a statist twist: it is an anarchy of leviathans, of person-like states rather than stateless persons. There, the want of common government makes sovereign states seek power to protect themselves from predators and then perchance to prey. Inequality makes the weak and then others seek power in numbers, friends to fight foes. But friends are fickle: little more than foes of foes, they can and do switch sides. And cooperation beyond shared enmity is even harder to achieve: although power rest in peaceful balance, Prisoner's Dilemmas abound, putting the fruits of collective action—of peaceful intercourse and mutual security even more than alarmed alignment—mostly out of reach.

A richer, more contemporary picture softens the demand for power and the problem of cooperation by overlaying these ingredients of interstate anarchy with two more. One is the paradox of power: sometimes a state profits from a loss of power, indeed a unilateral loss, all else being equal. The other is the shadow of the future: sometimes the prospect of indefinite repetition of the Prisoners' Dilemma makes present cooperation the best strategy for each of two states.

My own picture further enriches that one by expanding and combining those new ingredients. The most familiar power paradoxes show that a unilateral loss of power can help a state defeat an adversary. But in the versions discussed here, such a loss instead helps a state win over an ally and bring about mutual cooperation. That does not have to involve the shadow of the future, but it can, and then it is enough that only one of two states anticipate future interaction. The shadow of the future is but one of many formally similar motivations to cooperate, and thanks to the paradox of power, all of them can be effective even when one-sided. The overall result is conceptual and integrative: a simple structural condition for mutual cooperation subsumes quite a diversity of others, and it widens their scope and reveals their underlying similarity.

2. Cooperation Problems and Solutions

To cooperate with others is to help or benefit them in some way. A *problem of cooperation* arises when two or more actors can cooperate with each other to their mutual benefit but each finds it better *to defect*, or not cooperate, because cooperation is costly and the benefit to one would come solely or mostly from the cooperation of others. This sort of problem is epitomized by the folkloric Prisoners' Dilemma, or PD, one of the three games shown in Figure 2. In it, either player (1 or 2) is better off defecting (*D*), regardless of what the other does. So both players defect (*DD*). But both would have fared better had both cooperated (*CC*). Thus the problem.

Prisoners' Dilemma (PD)

2

	C	D
C	CC 2nd best for both	CD worst for 1 best for 2
D	DC best for 1 worst for 2	DD 2nd worst for both

Stag Hunt

2

	C	D
C	CC best for both	CD worst for 1 2nd best for 2
D	DC 2nd best for 1 worst for 2	DD 2nd worst for both

Asymmetric PD

2

	C	D
C	CC best for 1 2nd best for 2	CD worst for 1 best for 2
D	DC 2nd best for 1 worst for 2	DD 2nd worst for both

Figure 1. Cooperation games

It arose because both players were (let us say) *greedy*, preferring solo defection (*DC* for 1, *CD* for 2) to mutual cooperation (*CC*). But in the otherwise similar Stag Hunt (a.k.a. Assurance), both players are *generous*, preferring mutual cooperation above all. As a result, neither player has a *dominant option*, one that is better regardless of the other's choice: if player 1 cooperates then so should 2, but if 1 defects then 2 should too. As in Prisoners' Dilemma, therefore, *DD* is a *Nash equilibrium*, or pair of best responses to each other's choice. But now so is *CC*. Because both players prefer *CC* to *DD*—because, as we also say, *CC* would give both players greater *payoffs*—*CC* is the most reasonable prediction. Generosity has immunized 1 and 2 from the problem of cooperation.

Up to a point, the problem returns in Asymmetric Prisoners' Dilemma, where 1 is generous, 2 greedy. Whatever 1 does, 2 is better off defecting. But then 1 is better off defecting as well. As in ordinary PD, *DD* is the only Nash equilibrium.

All of this fits the conventional reading of game matrices, whereby players make their choices simultaneously. But what if they can *stagger* their choices? What if one player chooses first and the other sees that choice before choosing? In PD that makes no difference: the second chooser would choose *D* whatever the first chose. In Stag Hunt, staggering merely strengthens our expectation that both players will cooperate: after one did, the other would know it and do the same. In Asymmetric PD, staggering again makes no difference if generous 1 goes first: whatever 1 chose, greedy 2 would choose *D*, making *D* the only reasonable choice for 1. But if 2 can go first the cooperation problem is solved. If 2 chose *D* then so would 1, but if 2 chose *C* then 1 would too. Knowing that and preferring *CC* to *DD*, greedy 2 would first choose *C* and generous 1 would then reciprocate. The ball of cooperation now lies in the greedy player's court: it is he who can and must initiate cooperation.

If *C* and *D* are momentary acts, the fancied staggering may well be possible. But in the interesting cases, *C* and *D* are *long-term policies*—peace, free trade, resource sharing, etc., and their opposites. Then 2's choice of *C* could take forever to implement. Lest 1 wait that long to reciprocate, 2 can achieve mutual cooperation even in the short term if his long-term cooperative policy begins with an *effective commitment* to keep cooperating, not

to reverse course. And that is possible if 2 is in a position, not merely to begin cooperating, but to *shed the power* to defect. Though greedy, 2 would do just that if feasible because 2 prefers *CC* to the *DD* that would otherwise prevail. See how a paradox of power has spawned mutual cooperation: all else being equal, greedy 2 would benefit from a unilateral loss of the power to defect against generous 1, whose preferred strategy would then be to cooperate.

To apply this idea as broadly as possible, let us assume (what is implicit in many applications of games) that our game matrices are accurate and commonly known but not necessarily complete: both players have the indicated options and preferences, both know it, and both know that both know it, but either or both of them may have any number of further, cross-cutting options, in effect different ways to cooperate or defect—immediately or in five minutes, as fully as possible or less so, using a Cruise missile or a crossbow, while negotiating a trade agreement or not with some third player, and so on. In other words, *C* and *D* might be *categories* of more specific options: to cooperate is to choose one member of category *C*; to defect, one member of *D*. The difference between categories may be one of degree: every *C* option is *more* cooperative—more helpful to the other player—than any *D* option, but the line separating the two categories might be arbitrary. Besides being numerous, the members of either category may be quite complex. Each is a *strategy*, or complete plan of action. As such it prescribes a sequence of one or more *moves*, some possibly conditional on earlier moves by the strategist himself or the other player. We can retain the conventional assumption that players choose their strategies simultaneously, for those strategies may include staggered moves.

What if there are three or more players? Take the interesting case: some players are generous, others greedy, and every greedy player can initiate irreversible cooperation, by shedding power if need be. If the benefits of cooperation are diffuse and the generous players are willing to cooperate after many *but not all* of the greedy players have, then the greedy players face a cooperation problem among themselves: each has an incentive to free ride on the cooperation of others. But if the generous players offer great enough benefits from their own cooperation, and if they prefer not to cooperate until *all* the greedy players have, then every greedy player has an incentive to cooperate. Likewise if the generous players, besides offering great enough benefits, also can *selectively target* these benefits at those greedy players who have cooperated.

3. A Story

Here is an example of mutual cooperation in Asymmetric PD by way of staggered moves and lost power. A river runs from Upper Riparia through DamNation, which can dam it, creating a lake and with it more accessible water for itself and upstream Upper Riparia. The latter could then achieve a further increase, for itself alone, by digging irrigation canals. So DamNation has two options, to CONSTRUCT a dam (*C*) or not ($\bar{C}$), and Upper Riparia has two, to DIG canals (*D*) or not ($\bar{D}$). The possible outcomes are four: $C\bar{D}$, $\bar{C}D$, CD, and $\overline{CD}$.

Both states want more water, but dams and canals cost something. Without a lake, canals would carry too little water to be worth the cost. So Upper Riparia prefers CD to $\bar{C}D$ to $\overline{CD}$ to $C\bar{D}$. But with a lake, canals would lower the water level enough to make the dam too costly for DamNation. So DamNation prefers $C\bar{D}$ to $\overline{CD}$ to CD.

Let DamNation move first. If it chose *C* then Upper Riparia would choose *D*: the latter prefers CD to $C\bar{D}$. Otherwise Upper Riparia would choose $\bar{D}$: it prefers $\overline{CD}$ to $\bar{C}D$. So *C* would lead to CD, $\bar{C}$ to $\overline{CD}$. But DamNation prefers $\overline{CD}$ to CD. So it chooses $\bar{C}$ and Upper Riparia $\bar{D}$: the outcome is $\overline{CD}$, no new costs but no dammed water.

Add power to the story. Wisely or otherwise, DamNation can construct the dam and try to stop canal digging with a naval bombardment of the project site. But that is the only threat it has, and it is not good enough: Upper Riparia has shore batteries that can sink warships near the site.

Now subtract power. Upper Riparia destroys its batteries, losing the power to dig canals but (we may suppose) nothing else of value: like U.S. army posts on the Great Lakes, those batteries are relics of some conflict long passed. In response, DamNation dams and the outcome is $C\bar{D}$, better for both than old $\overline{CD}$. There is the paradox: Upper Riparia has gained from its loss of power, all else having stayed the same.

Figure 2 portrays this story as a game in extensive form. Each decision nodeshows a state and its two options, the preferred one marked by an arrowhead. The part in a dashed box captures the original story, before I added the business about power. That part is Asymmetric PD: generous DamNation most prefers mutual cooperation ($C\bar{D}$), to which greedy Upper Riparia prefers solo defection (CD). The game is staggered, but with DamNation going first, blocking cooperation. A mere change of order would not promote cooperation because Upper

Riparia's momentary choice of nonirrigation would be reversible. It is double staggering that creates mutual cooperation: before DamNation constructs a dam, Upper Riparia has to shed the power to dig canals.
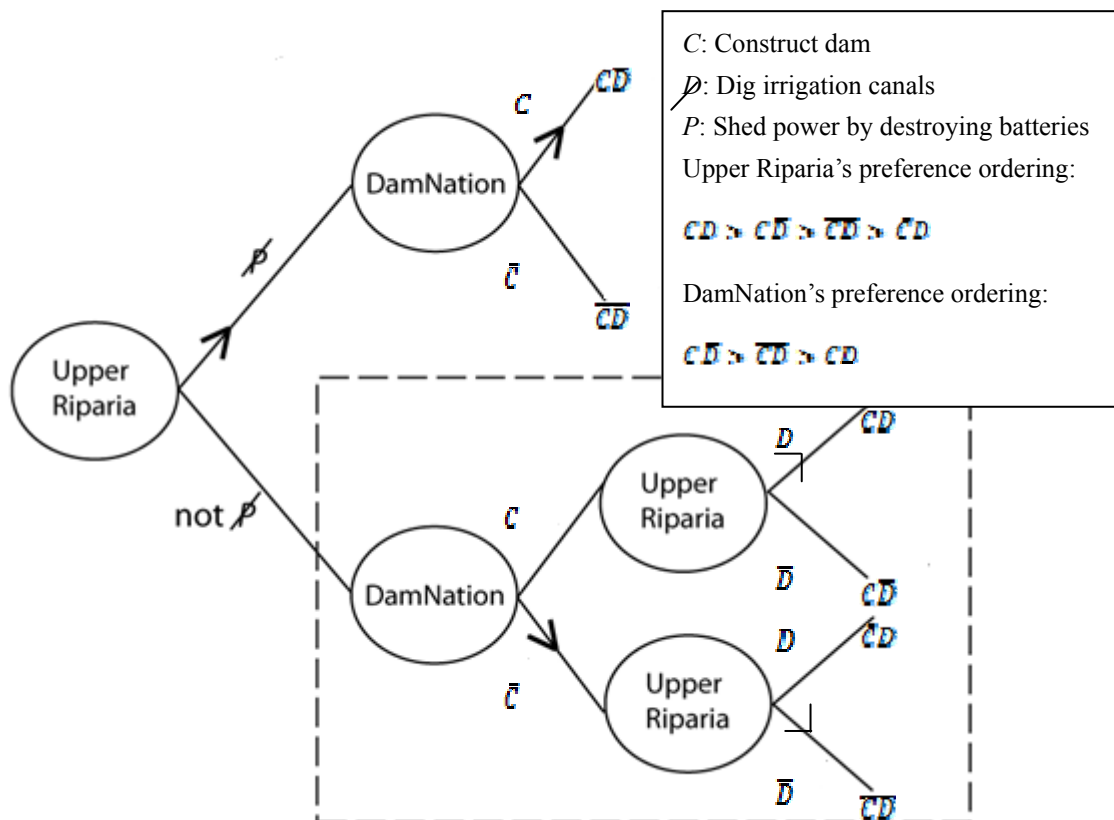

Figure 2. Extensive form of game

Power in geopolitics is like God in religion: the idea is elusive but hard to elude. Is it really power that Upper Riparia lost? Not if power is measured by success or its prospect, by the achievement or likely achievement of ultimate goals. Then Upper Riparia actually gained power. But that measure weakens power as an explanatory factor, or independent variable: we can no longer use it *to explain* success or failure, the achievement or failure to achieve ultimate goals. Worse, we must conclude, for example, that Mexico had greater power than Japan in WWII. What Upper Riparia lost was a military asset and with it an ability (Fr. *pouvoir*, the root of *power*) to scare off or defeat an adversary in armed combat. If that is not power, nothing is.

So a unilateral loss of power can be profitable, a gain costly. How surprising is that? Where Upper Riparia profited from losing the power to *hurt a benefactor*, Schelling (1980) celebrates cases in which one actor profits rather from losing the power to *accommodate an opponent*. You might scare off an enemy by burning a bridge or otherwise destroying a route of retreat, or by tossing the steering wheel out your car window in a game of chicken. You might also compel a bargaining adversary to accept your position by destroying your power to accept his, e.g., by having your principal or constituency or "domestic audience" block or rescind your authority to accept anything less. In those cases power shedding amounts to tying one's hands, or making an effective commitment. Of course it is obvious that effective commitments, such as enforceable contracts, can also secure cooperation. What I have added is asymmetry, or one-sided generosity (played against other-sided commitment).

4. Applications

Quite a variety of cases fit my pattern of cooperation in Asymmetric PD by way of staggered moves and the unilateral loss of power.

4.1 Surrender in War

If that does not strike you as a case of cooperation at all, it is because the victor does not seem especially "generous", the vanquished especially "greedy". But in my technical sense they are: pleased enough with the

prevailing disposition of forces, the victor most prefers that both sides stop shooting (*CC*); less pleased, the vanquished most prefers that the victor alone stop (*CD*). The greedy vanquished must move first by effectively surrendering, not only raising a white flag but placing his forces in a vulnerable position where they lay down their arms: he must shed his power to continue the fight.

4.2 Preventive War and Its Prevention

Scholars since Thucydides (1954) have pointed to cases where states apparently gained power only to suffer offsetting gains or armed attack, either by new enemies who feared future gains (Jervis, 1978; Morgenthau, 1967, 202f.; Niou & Ordeshook, 1987; Levy, 1987) or by old allies dissatisfied with their shares of past gains (Maoz, 1989). In those cases a unilateral gain was costly, so a unilateral loss would have been profitable, like surrender but in anticipation of war (or of an arms race). When one state does that, it no longer poses a threat to others, so those others cooperate with it by remaining at peace or preserving the balance of arms. Because that is their most preferred outcome, they are generous in my sense.

4.3 Surrender to a Leader

Similar to surrender in war is the acceptance of a Hobbesian sovereign by his subjects—the classical recipe for polity-wide cooperation—and more generally of a leader by followers absent other effective authority. According to Hobbes, those subjects must end the pre-social "war of all against all" by laying down their arms, by shedding their power to keep defecting. They need not expect especially fair or efficient rule. What makes the sovereign generous is that his rule would be better than the alternative. Because he is, by Hobbes's definition, the *de facto* sovereign or, more generally, the strongest candidate for that role, he is the focal point for coordinated obedience, so the alternative to his rule is no rule (Hardin, 1989, generalizes this argument to all effective constitutions). And because he already enjoys considerable power, he can reciprocate cooperation by selective targeting: each new adherent gains some protection not enjoyed by holdouts. How might he have acquired such power to begin with? Desperate for leadership, his adventurous first followers may have found him unusually leader-like and, by joining him, bet that he would eventually win the crown and reward early support. Publicity, growing power, the defeat of rivals, and ever greater benefits targeted at new adherents—these things would then interact and cascade to secure his sovereignty.

		C	2	D
1	C	R=3(+X), R=3(+X) Reward for mutual cooperation		S=1, T=4(-Y) Sucker's payoff and temptation to defect
	D	T= 4(-Y), S=1 Temptation to defect and Sucker's payoff		P=2, P=2 Punishment for mutual defection

1's payoffs precede 2's.

Figure 3. PD and modified payoffs

"He" could be a state or central government too, securing the adherence of other states or tribes or provinces. Athens played that role in the Attic League. The initially frail Continental Congress and Army played that role in the American Revolution. Prussia played it in the North German Federation and then the German Empire—as has every empire that secured the submission of vassal states.

4.4 Sanctions, or Selective Incentives

These are penalties and prizes targeted at potential defectors by some central authority or other third party. If effective they create an incentive to cooperate, and there is then no problem of cooperation, no PD. Ignoring X and Y, Figure 3 is PD with Axelrod's (1984) famous labels for payoffs. I have added simple, somewhat arbitrary payoff numbers (4 for best, 3 for second best, etc.). Now think of X and Y as (non-negative) selective incentives, X a prize and Y a penalty. They are effective only if $X+Y>T–R$ (=4-3); say X =Y=1. Precisely what that does is turn PD into Stag Hunt, encouraging mutual cooperation.

Sanctions can work imperfectly, however: cheaters and criminals exist, I have heard. But for mutual cooperation it is sometimes enough that they work halfway. Suppose they directly affect only player 1: *X* and *Y* are added only to his payoffs. The result is Asymmetric PD: 1 is generous, 2 greedy. That alone does not ensure mutual cooperation. But it encourages it if moves can be staggered so 2 goes first and, unless *C* and *D* are momentary acts, sheds the power to defect. Even one-sided selective incentives can bring about mutual cooperation in an initial PD. That happens when banks are compelled by law to treat debtors fairly and honestly, and the latter, less constrained by legal sanctions, shed the power to default by signing over assets as collateral. It happens too when a legally governed economic union bails out an insolvent member state or admits a new but less prosperous member, and the latter relinquishes some internal governance authority to outside monitors.

4.5 The Shadow of the Future, or the Effect of Repetition on Reputation and Reciprocity

In an indefinitely repeated PD, players who cooperate with cooperators but defect against defectors can secure each other's cooperation (Axelrod, 1984). If you and I are playing indefinitely repeated PD, there are countless strategies either of us might follow ("always cooperate", "always defect", "cooperate until the other defects and then defect forever", etc.), and countless pairs of them are Nash equilibria (including "always defect, always defect"). The most famous strategy, found by Axelrod to be especially advantageous against many other strategies, is TIT FOR TAT: cooperate on the first play, and after that do what the other player just did. If both players adopt TIT FOR TAT, both will cooperate forever and receive *R* at each play. Infinite total payoffs are impossible, of course, but we should discount future payoffs anyway, say by .9. Then the long-run payoff from *C* today is not 3 but 3+2.7+2.16+…, or 30, and the long-run payoff from *D* today is not 4 but 4+1.8 (=.9 ✗ 2)+1.62+…, or 22. At each play the game is that of Figure 3 with *X*=27 and *Y*=20. As in the selective-incentive case, those payoffs are great enough to turn greed into generosity, PD into Stag Hunt, *DD* into *CC*.

And as in that case, the shadow of the future can help secure mutual cooperation even when it directly affects only one player, making him alone generous. A player willing to cooperate with cooperators and defect against defectors builds a reputation that can help him secure the cooperation not only of past interactors but of new, similarly generous interactors (Axelrod & Keohane, 1985), and then of greedy players too, so long as the latter are in a position to initiate cooperation by shedding the power to defect. That happens between individuals when a notorious deadbeat gets a reputable tradesman to do a job by paying him in advance, shedding the power to welsh. It happens between states when a poor credit risk obtains a loan from a commercial state that has long cultivated and profited from a reputation for honesty by letting the latter control its customs houses.

4.6 Trade and Constitutional Empowerment

Not every cooperation problem or power paradox involves force. For trading states, to cooperate is to lower or forego barriers to imports, and to defect is to protect domestic producers from foreign competition. A common textbook illustration of plain PD consists of two states who both adopt trade barriers, greedily following "beggar thy neighbor" strategies, although two-way trade would have been better for both. Possibly they can achieve mutual cooperation by exploiting the "shadow of the future": if both begin cooperating but one later switches to defection, the other can quickly retaliate. But not if the latter found retaliation too costly after its consumer-voters had become addicted to cheap imports and euthanized its less competitive industries.

In *Asymmetric* PD based on trade, generous 1 prefers mutual free trade above all, and greedy 2 prefers solo protection. To initiate cooperation, 2 must not only unlock its door to 1's imports but throw away the key. How? Start with a prior question: How can 2 be sure that 1 is generous? Possibly by observing 1's history of trade relations. But if that is too short or otherwise inconclusive, 2 can read 1's constitution and election code and ask if, in matters of trade, it clearly makes consumers sovereign. That is something 2 might try itself, by institutional tinkering to empower a broad electorate of consumers over any narrow oligarchy of producers. Not that democracies always favor free trade. But as Chang, Kayser, Linzer, and Rogowski (2010) have shown, it is not democracy as such that gives effective voice to consumer interests. It is certain forms of voting and representation, notably single-member-plurality systems (in contrast chiefly with list PR). Nor is electoral reform the only way to shed protectionist power. Another is to constitute a politically independent trade court charged with banning and rescinding protectionist rules, an in-house WTO.

One might object that if a constitutional change replaces greedy with generous rulers, the immediate result is as much a newly generous state as it is a greedy one shorn of certain powers—and indeed the former is how I describe a similar change in the next section. The final outcome is still *CC*, however, now got by way of Stag Hunt rather than Asymmetric PD. It is interesting that the two cases are hard to distinguish. The reason, I believe, is that constitutions effectively limit power less through "parchment barriers to the encroaching spirit of power"

(James Madison, *Federalist* 47) than by empowering opposing preferences, or interests: "Ambition must be made to counteract ambition" (Madison, *Federalist* 51).

4.7 Kindliness

In the story of Section 2, DamNation was generous only because its dam would unavoidably create a spillover benefit (forgive me) for Upper Riparia. But some states, like some individuals, are avoidably generous: they like helping others (or certain others). Anyone who has tried to simulate PD in the classroom must be struck by the difficulty of getting students to follow the assigned payoffs and defect. PD experiments are heralded for showing that indefinite repetition (or a high enough probability of continuation) boosts cooperation, as expected (Selten & Stoecker, 1986; Dal Bo, 2005), but an unheralded finding of those same experiments is that cooperation by no means disappears in finitely repeated or even single-play PD. The obvious explanation (and the one always voiced by my uncooperatively cooperative students) is that some players are disinclined to hurt others. That is captured again by Figure 3, with X now interpreted as the gratification of acting kindly and $-Y$ as the discomfiture of not doing so. This phenomenon, too, extends to the asymmetric case if greedy 2 can move first by irreversibly initiating cooperation. In geopolitics that would happen if an impoverished rogue state destroyed a power asset that it had used to threaten others, and some kindly and wealthy states then helped it feed its people.

4.8 Fidelity

Cooperation in Asymmetric PD can come from the value of *fidelity* (or "honor" as celebrated by O'Neill, 2001; or "trust" as celebrated by Kydd, 2007). Start with ordinary PD, and suppose that 1 and 2 try to escape *DD* by promising to cooperate. So far we have no reason to expect cooperation: either player would do better to break his promise, whatever the other did. But suppose that player 1 prizes fidelity: as a matter of pride in principle, he prefers keeping his word, so long as 2 cooperates, to cheating 2. If 2 is likewise faithful and they both know all this, they will certainly cooperate. But even if 2 is not faithful, they will still cooperate provided 2 can irreversibly initiate cooperation, shedding power if need be, as in Figure 4. There, although 1 is greedy at first, he is faithful all along. As a result, his conditional promise to cooperate would change his preferences, or payoffs, to *make* him generous (in the subgame following that promise), turning plain into Asymmetric PD, staggered in both cases. Or if it seems too paradoxical that a player can prefer to change his own preferences, we can say that preferences are fixed but 1 prefers PCC to PDC and also PDC to PCC. Mutual fidelity can secure mutual cooperation, of course, but so can one-way fidelity: in staggered Asymmetric PD it is enough that one player is faithful. That happens when a famously faithful belligerent secures an enemy's surrender by promising generous terms.
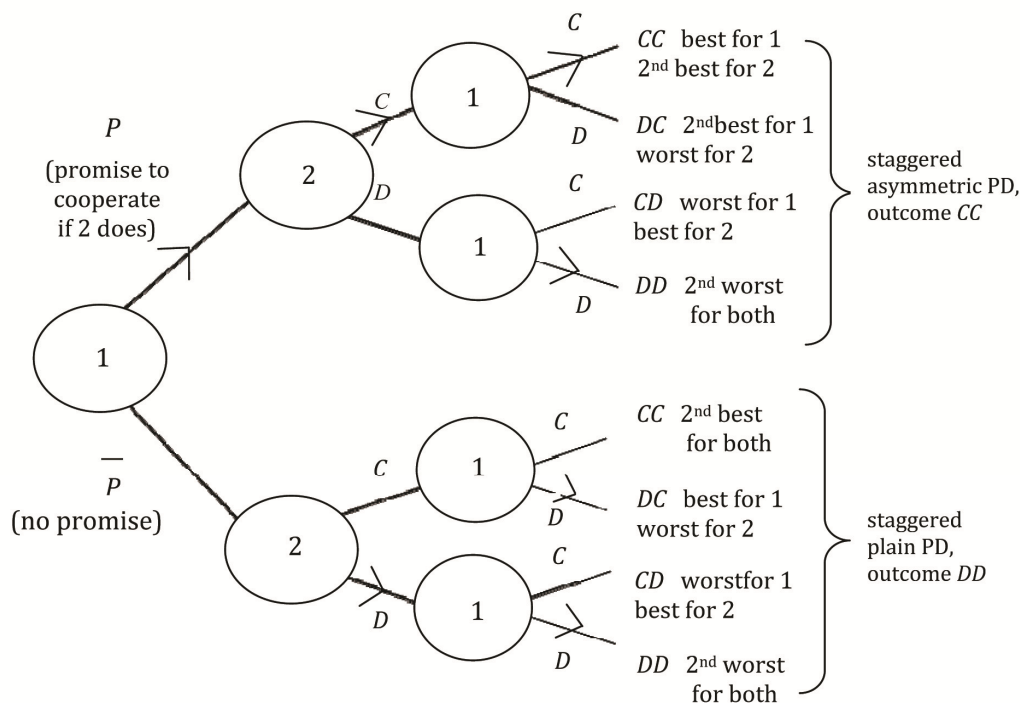
Figure 4. The value of fidelity for player 1

4.9 Predation, Resistance, and the End of the Cold War

Perhaps the most obvious sort of Asymmetric PD in world politics pits a greedy *predator* against a generous *resister*. The predator seeks conquest and spoliation. Even if not an intended prey, the resister sees its interests threatened or its principles flouted and is willing to block or limit the predator's aggrandizement. Like attempted predation, resistance is costly, and both states would fare better from mutual cooperation—no predation and no resistance. The predator might be France under Louis XIV or Napoleon, or Germany under William II or Hitler; the resistor, England/Britain. By Elizabeth's time, the latter sought no territory in Europe but saw territorial aggrandizement as a threat to its coast and commerce. To end or avoid war with England/Britain, the predator had to withdraw or surrender its invading or occupying army, the instrument of its power to prey.

When common knowledge is not there to begin with, the resister must create it, and that may require patience and skill and a heavy investment. Obviously it should make its willingness to resist, its preference for two-way defection to solo cooperation, as evident as possible to the predator. Likewise its generosity, its preference for two-way cooperation to defection by itself alone. The predator might or might not be able to shed power. But if it can then the resister may have to do something more to achieve two-way cooperation: bear the cost of making two-way defection as costly as possible to the predator.

My favorite Asymmetric PD of this sort is the Cold War in the days of Ronald Reagan and Mikhail Gorbachev (whereof *v.*esp. Skinner, 2008). *C* and *D* were broad policy options, wide categories of long-term strategies, and they differed in degree: *CC* would involve a greater degree of cooperation than had prevailed, but some cooperation had of course taken place already under the *DD* of the Cold War (Trachtenberg, 1999, Part II). The Soviet Union was player 2, a predator, greedy for empire. The United States was player 1, a resister, generous in the sense of not seeking more territory but keen to reduce the risk of war and the cost of arms and to engage in commercial, scientific, cultural, and other mutually advantageous exchanges with a less predatory Soviet Union. The arms race by itself might be portrayed as plain PD, but it was a byproduct of the main issue, Soviet imperialism. Soviet President Gorbachev was willing to cooperate, he said, but U.S. President Reagan insisted that Gorbachev initiate irreversible cooperation. This required that Gorbachev shed the power to defect. He did that by scrapping his chief asset of imperial power: the Red Army of occupation was effectively withdrawn from Eastern Europe. Thus ended the Cold War.

U.S. generosity was evident. The Soviet Union, not the United States, had seized territory as WWII was ending. The United States emerged from that war, as it had from WWI, the world's top economic and military power, but as in the aftermath of WWI, U.S. leaders and their constituents were disinclined to build a territorial empire and instead rapidly demobilized, changing course only to resist the predatory threat of the Soviet Union and especially its North Korean vassal. Neither had the United States any reason to threaten Russia's security by relaxing NATO's embrace or its own velvet-gloved occupation of Germany, whose prosperity and protection left it with no incentive to reclaim its status as a fully sovereign power.

Even so, Reagan had his work cut out. All relevant actors did not necessarily see the payoffs of the four outcomes. The complexity and temporal lengths of the *C* and *D* policies, the variety of beliefs and preferences within either superpower, and the smog of deception and distrust that always obscures grand-strategy options—these things inevitably led some influential Soviet actors to underestimate the costs of mutual defection and maybe the benefits of mutual cooperation. Like the British and French, they underestimated the burden of empire. As a result, it behooved Reagan to make those benefits and costs as salient as possible by making the U.S. commitment to *CC* as credible as possible but also by driving up the cost of *DD* as much as possible. The latter entailed a massive increase in defense spending, the deployment of U.S. nuclear missiles in Europe, the resurrection of anti-missile defense, and a manifest willingness to oppose Soviet imperialism on the ground in such peripheral theaters as Nicaragua, Angola, and Afghanistan. Many Americans found Reagan's *DD* too costly. So did sufficiently many Russians.

5. Lessons

How to explain cooperation in the face of an initial or apparent PD? The prominent explanations are selective incentives and the shadow of the future, although kindliness is so obvious and common that it goes without saying. All three start with original, self-interested, or short-term payoffs that together constitute a PD, then call attention to added payoffs (the *X* and -*Y* of Figure 3) that turn PD into Stag Hunt by turning the greedy preferences of both players into generous ones. My own explanation subsumes these and adds four things. One is further sources of generosity: a player can be generous without selective incentives, indefinite repetition, or kindliness. In our examples, DamNation owed its generosity to thirst, the free-trade state to consumer demand for cheap imports, the victorious belligerent to success at arms, the promise keeper to pride in fidelity, and so on. Three more additions are asymmetry, staggering, and the unilateral loss of power. Among other things they extend the effects of selective incentives, the shadow of the future, and kindliness from the symmetric to the asymmetric case.

Because states can profit from their own generosity, some greedy ones may wish to become generous. How? It is not a matter of becoming especially pacific or enamored of others or reliant on the "soft power" of diplomacy, tolerance, and charity. On the contrary, a generous state facing a greedy state may have to be quite ruthless in raising and maintaining the cost to the latter of *DD*. Neither is it enough to combine carrots with sticks, promises with threats. They are effective only if believed, and the most effective source of belief is truth: generosity is not a promise to cooperatebut a preference for mutual cooperation, persuasive if commonly known. Sometimes generosity requires no special effort to change: it is a byproduct of narrow national goals, as in DamNation's demand for more water, the consumer state's preference for free trade, and the traditional resister's want of appetite for spoliation. But there are at least two things a greedy state might do to cultivate greater generosity in itself. Neither is easy.

One is institutional fine tuning: put the decision whether to cooperate in the hands of constitutional or other institutional actors chosen for their generous preferences. Consumer-voters played that role in the trade story. Taxpayer-voters and citizen-soldiers might play a similar role in cases of armed conflict. So might sectoral representatives in cases of common-pool resource sharing. Notably it is sometimes easier for a state than an individual person to become generous and convince others that it is: political constitutions are more malleable and more transparent than personal ones. Care is required. On the one hand, the generosity of those institutional actors must be robust: they must be immune from nationalist demaguery or checked by others who are. On the other hand, besides being generous, they must be willing, or checked by others who are willing, ruthlessly to impose the cost of *DD* when necessary: on trade issues they must resist addiction to cheap imports, and in armed conflict they must avoid timidity, soft-heartedness, and gullibility. Even then generosity requires more than the empowerment of voters, judges, or others who have generous preferences. It requires the right procedures to elicit those preferences and combine them into one choice of policy. Recall that it is only certain electoral systems that give effective voice to consumer over producer interests. Obviously those procedures have to be politically sticky, hard to disestablish. But for that they do not always have to be written into a constitution.

Electoral formulas and judicial powers are typically codified in statutes, sticky because harder to rescind than to reject to begin with.

The other way for a state to cultivate generosity is to plant and nurture, within its own borders, a norm or ideal of *fidelity* backed by honor and pride. Fidelity lets a faithful actor turn an initially greedy preference into a generous one by the simple act of promising, of giving one's word: faithful players can push a button to change their own payoffs and turn PD into Asymmetric PD or Stag Hunt. Danger lies in the temptation to feign fidelity. A state that tries it may gain a momentary advantage but suffer ever after. That very fact can discourage deception, of course. But so can public efforts to teach and preach fidelity, to celebrate and reward it, and to sanction violations by officials, perhaps through special tribunals, certainly by adding public humiliation to any punishment. An advantage of cultivating fidelity rather than generosity as such is that generosity is issue specific but fidelity can cover any number of unexpected future problems of cooperation. Another advantage is that it leaves a state free to decide when and toward whom to be generous.

It is common sense that effective commitments to do costly things—to repay a loan, for example—can secure mutual cooperation. But it is a common caveat of game theory that commitments to do costly things are not effective unless explicitly allowed by the "rules of the game" (payoffs included). Allowed how? We have examined two ways. One is to let players unilaterally shed power, or tie their own hands. They can then commit to do a costly thing by shedding the power not to do it, by pruning the rest of the game of that option, as in Figure 2. The other way is to let players give their word and include some whose word is good. The latter can then commit to do a costly thing—costly in prospect, that is—simply by saying so, thereby shedding their initial greed, or changing payoffs so the once-costly thing now gratifies their taste for word keeping, as in Figure 3. Institutional change straddles these modes of commitment. Let a state transfer certain sovereign powers from a greedy organ of government to a generous one. If we equate the first organ with the state itself then the transfer is a shedding of power, as in Figure 2. If instead we see both organs as voices of the state but at different times, then the state has changed its preferences, or payoffs, as in Figure 3 but with P now interpreted as "institutional change" rather than "promise".

Lest we read too much into these lessons, let me add two qualifications. First, the simplicity of Figures 1-3 should not hide the challenge to would-be cooperators. Remember that C and D are categories, possibly large and diverse, of strategies, possibly complex and loaded with contingencies. Remember too that those pictures assume common knowledge, but a challenge facing some would-be cooperators is to *create* common knowledge, to get all players to see the same strategies and payoffs and to see that all players see them. In Asymmetric PD it may be especially hard for 1 to convince 2 of its generosity. Those pictures also treat states as single actors, but a challenge to some within-state actors (some individual components of the leviathan) is to get others to see and prefer and act as one.

Second, to the extent that my words are read prescriptively, their intended import is merely instrumental. It is how to cooperate and achieve shared or complimentary goals, but those goals can be anything, good or bad. Game outcomes were rated best, second best, etc., according to players' preferences, but preferences need bear no particular connection to their subjects' welfare (Schwartz, 1982), much less that of any wider community. Individuals might cooperate with each other by fulfilling a suicide pact, robbing a bank, lynching a member of a hated race, or—need I add?—refusing to betray each other by confessing to such a crime. Likewise states: Nazi Germany and Soviet Russia cooperated for two years to conquer and despoil much of Europe. In world politics as in industrial organization, a more cooperative world could well be a worse world.

It has become fashionable since Axelrod's (1984) brilliant book to explain how cooperation might evolve, within the human race or particular communities of men or beasts or the world "community" of sovereign states. The underlying idea is that, in an endlessly repeated PD, a strategy of cooperating or defecting, conditional on the acts of one's fellow players, conduces to survival (of the community or species that adopts it or of the strategy itself) to the extent that it secures mutual cooperation. Although I would insist that mutual cooperation *need* have no survival value (it depends on cooperators' goals, which could be anything, e.g., mutual suicide), let us take a world and a class of cases in which cooperation does have survival value. There, if complex, hard-to-implement game strategies can evolve, as Axelrod and his followers contend, consider how much easier it might be for simple preferences to evolve, for interaction between states to select generous ones, or better, a general preference for fidelity. The value of such a preference is that it spawns cooperation with greedy as well as generous fellow actors, and it does that without requiring subordination to some central authority or endless repetition of some PD whose stakes are low enough that occasional unrequited cooperation becomes tolerable.

6. Conclusion

The classical solutions to cooperation problems all start with PDs composed of narrow, short-term, or self-interested payoffs, then alter them so as to turn greed into generosity, PD into Stag Hunt. But it is enough to make one of two players, or several of many, generous, so long as the remaining greedy players can irreversibly initiate cooperation by shedding the power to defect. Because generous players are thus in a position to attract the cooperation of greedy as well as generous interactors, generosity can be an asset worth cultivating. Because fidelity enables the faithful to choose when and toward whom to be generous, it is an even greater asset, one that states can try to cultivate by tying it to public and personal honor.

References

Axelrod, R. (1984). *The evolution of cooperation*. New York: Basic Books.

Axelrod, R., & Keohane, R. (1985). Achieving cooperation under anarchy: Strategies and institutions. *World Politics*, *38*, 226-254. https://doi.org/10.2307/2010357

Chang, E. C. C., Kayser, M. A., Linzer, D., & Rogowski, R. (2010). *Electoral systems and the balance of consumer and producer power*. Princeton: Princeton University Press. https://doi.org/10.1017/CBO9780511761423

Dal Bo, P. (2005). The shadow of the future: Evidence from infinitely repeated games. *American Economic Review*, *95*, 1591-1604. https://doi.org/10.1257/000282805775014434

Hardin, R. (1989). Why a constitution? In B. Grofman, & D. Wittman (Eds.), *The federalist papers and the new institutionalism* (pp. 100-128). New York: Agathon Press.

Jervis, R. (1978). Cooperation under the security dilemma. *World Politics*, *20*, 167-187. https://doi.org/10.2307/2009958

Kydd, A. H. (2007). *Trust and mistrust in international relations*. Princeton: Princeton University Press.

Levy, J. S. (1987). Declining power and the preventive motivation for war. *World Politics*, *40*, 167-187. https://doi.org/10.2307/2010195

Maoz, Z. (1989). Power, capabilities, and paradoxical conflict outcomes. *World Politics*, *61*, 239-266. https://doi.org/10.2307/2010410

Morgenthau, H. J. (1967). *Politics among nations* (4th ed.). New York: Knopf.

Niou, E. M. S., & Ordeshook, P. C. (1987). Preventive war and the balance of power. *Journal of Conflict Resolution*, *31*, 387-419. https://doi.org/10.1177/0022002787031003001

O'Neill, B. (2001). *Honor symbols and war*. Ann Arbor: University of Michigan Press.

Schwartz, T. (1982). Human welfare: What it is not. In H. Miller, & W. Williams (Eds.), *The limits of utilitarianism* (pp. 195-206). Minneapolis: University of Minnesota Press.

Selten, R., & Stoecker, R. (1985). End behavior in sequences of finite prisoner's dilemma supergames. *Journal of economic behavior and organization*, *7*, 40-70.

Skinner, K. K. (2008). An alternative conception of mutual cooperation. In K. K. Skinner (Ed.), *Turning points in ending the Cold War* (pp. 93-110). Stanford, CA: Hoover Institution Press.

Trachtenberg, M. (1999). *A constructed peace*. Princeton: Princeton University Press.

Copyrights