Economics of Governance (2020) 21:215–236
https://doi.org/10.1007/s10101-020-00239-2

ORIGINAL PAPER

A unifying theory of positive and negative incentives in international relations: sanctions, rewards, regime types, and compliance

Byungwon Woo[1] · **Daniel Verdier**[2]

Received: 25 August 2019 / Accepted: 29 May 2020 / Published online: 15 June 2020

Abstract
Should democracies be rewarded and autocracies punished, or should it be the reverse? This is an important question for foreign policy makers who regularly find themselves wanting to alter the behavior of foreign governments favorable to their interests. Existing studies on economic sanctions and rewards provide an uneasy answer that sanctions are more effective toward democracies and rewards work better toward autocracies, suggesting democracies need to be punished while autocracies need to be rewarded. We revisit the issue of regime type and incentive form by building a game theoretical model focusing on domestic political dynamics in a Target country. When we distinguish betwen three types of reigmes lined up on an accountability continuum, the theoretical model yields the claim that sanctions and rewards work better with both extremes—democracies and dictatorships—than with the intermediate category of limited autocracy, for which only rewards work.

Keywords Sanctions · Rewards · Regime type · Compliance · Democracy · Dictatorship

1 Introduction

Should democracies be rewarded and autocracies punished, or should it be the reverse? This is an important question for foreign policy makers who regularly find themselves wanting to alter the behavior of foreign governments in a way that is

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10101-020-00239-2) contains supplementary material, which is available to authorized users.

✉ Byungwon Woo
bwwoo@yonsei.ac.kr

[1] Department of Political Science and International Studies, Yonsei University, Seoul, South Korea

[2] Department of Political Science, The Ohio State University, Columbus, OH, USA

favorable to their interests. For instance, for US presidents, how to curb Jewish settlements in the West Bank and to contain nuclear proliferation in Iran and North Korea are important questions. In a world where she could have her wish, there is little doubt that the citizen of an established democracy would have a preference for proffering positive incentives to democracies while reserving sanction threats for nondemocracies. But what feels right may not be what is most efficient. As we shall see, a large number of political scientists who have done research on sanction threats have concluded that such threats are more effective when used against democracies than nondemocracies. And although rewards have not received as much attention as sanctions, at least one study concludes that democracies are less sensitive to promises of rewards than nondemocracies. Against popular wish, current science claims that democracies should be punished and nondemocracies rewarded. Israel should be threatened with sanctions while Iran and North Korea should be engaged with positive incentives.

We revisit the issue of regime type and incentive form to conclude that neither the popular wish nor the scientific claim have it quite right. We argue, instead, that the question of the respective efficiency of sanctions and rewards is a non-linear function of regime type. Specifically, when we distinguish between three types of regimes lined up on an accountability continuum, governments in democracies are accountable, they are less so in limited autocracies, and from l east so to not at all in dictatorships. Given this trichotomy, we argue that sanctions and rewards work better with both extremes—democracies and dictatorships—than with the intermediate category of limited autocracy, for which only rewards work. In other words, we find a U-shaped relation between regime and effectiveness of incentive. It is intermediate regimes like Iran—regimes that are neither quite as democratic as Israel nor as absolutely autocratic as North Korea—that are the least responsive to sanctions.

We make our argument by formalizing the consequences that sanctions and rewards have on the domestic make-up of the targeted country. In doing so, we incorporate the rally round the flag effect, the fifth column effect, and the extortion effect in the model. Specifically, we model that sanctions elicit rallies around the flag among groups who expect to benefit from the sanction and thus oppose compliance; that rewards elicit fifth-column effects among those who expect to benefit from the reward and thus support compliance; and that rewards may elicit extortion, a preemptive investment in wrongdoing on the part of the target government with an aim to raise the ante. It is only by modeling sanctions, rewards, and their respective side effects in "a unifying theoretical model of positive and negative incentives" that we may be able to determine under what circumstances sanctions, rewards, or a mix of the two constitute the most appropriate incentive.

The theoretical model yields the claim that limited autocracies stand out from democracies and dictatorships by being less responsive to sanctions, and consequently by attracting more rewards. This is because limited autocracies uniquely reserve the ability to switch their supporting coalitions and thus can utilize the rally round the flag for their political benefits. The rally round the flag and fifth column effects can be felt in any type of regime, but we show that they only have significant consequences for rulers in limited autocracies that enjoy enough autonomy to strategically choose between compliance and defiance with a demand on the basis of

the external incentives that are threatened or promised. This condition eliminates democratic leaders from that list, for they do not enjoy enough autonomy from their support base. In dictatorial regimes where no support of any coalition is ever necessary for the government to stay in power, rallies and fifth column effects are also inconsequential.

In Sect. 2, we review existing findings on the matching of incentives to regime types and on the relative effectiveness of sanctions and rewards in general. Our goal is not to dismiss existing results but to bring them together in a unified theoretical framework. In Sect. 3, we present the model followed by detailed analysis of the model in Sect. 4. In Sect. 5, we summarize testable predictions from the theoretical model that are subjected to an empirical test in Sect. 6. We discuss real world implications of the main argument in Sect. 7 before concluding in Sect. 8.

2 Literature review

Works on the question on whether different regime types respond to sanctions differently divides into negative and positive incentives. Those on negative incentives display a remarkable degree of coherence. They build on the early work of Kaempfer and Lowenberg (1988), who argue that, in order to be effective, sanctions must threaten costs on groups and individuals who are in a position of power. Following Bueno de Mesquita et al's (1999) characterization of such groups as larger in democracies than in nondemocracies, the field argues that sanctions affect the two types of regime differently. Accountable to a broad base of support, leaders in democracies enjoy fewer opportunities to resist economic hardship than autocrats, who can often insulate themselves and their immediate supporters from retribution by the victims of sanctions.[1]

While the literature portrays democracies as more responsive to external pressure than nondemocracies when pressure takes the form of a sanction, the aid literature reverses this finding. That is, democracies are less sensitive to promises of aid than nondemocracies, because while democratic leaders must cater to the broad public, autocrats rely on rents distributed to a small coterie of cronies to establish their authority. This effectively make one dollar of aid have a bigger marginal impact in

[1] As a result, sanctions cause more antigovernment protests in democratic than in autocratic regimes (Allen 2008), lead authoritarian incumbents to increase repression to suppress dissent (Wood 2008; Peksen and Drury 2008), while making democracies more likely to comply than nondemocracies, with compliance measured by sanction duration (Bolks and Al-Sowayel 2000; Allen 2005). In the same vein, comprehensive trade sanctions are not as effective against autocrats than "smart" sanctions, which directly target the groups and individuals who are close to the leader (Brooks 2002; Lektzian and Souva 2007). Also, a change in leadership is unlikely to have any impact on sanction duration in democracies, where the newly elected leader must rely on the same median voter as did the incumbent. In contrast, leadership turnover in autocracy is more likely to elevate a different group to primacy, with corresponding impact on sanction duration (McGillivray and Stam 2004).One should note the existence of two works extending the democratic peace argument to economic sanctions (Lektzian and Souva 2003; Cox and Drury 2006). We do not include them here because they focus on the joint effect of regime type in Sender and Target, whereas we limit our investigation to the Target.

nondemocracy than in democracy (Lai and Morey 2006, see also Bueno de Mesquita and Smith 2007).[2]

Taken together, these two strands of literature point to a lack of equivalence between sanction threats and aid promises: sanction threats have a stronger impact on broad-based governments, whereas aid promises have a stronger impact on elitist governments. The common rationale, if any, seems to be that sanctions, having a greater marginal impact on masses than on elites, will be preferred by sanctioners targeting democracies, whereas aid, having a greater marginal impact on elites than on masses, will be preferred by sanctioners targeting nondemocracies.

Also relevant to our topic is the literature that bears on the relative merits and demerits of positive and negative incentives. A rally round the flag occurs whenever a sanction threat arouses a nationalist response from potential winners from the sanction, making compliance with the sanction threat more difficult. Conversely, a fifth-column effect occurs when it is those who expect to be hurt who mobilize in support of compliance (Galtung 1967; Selden 1999; Rowe 2001; Nincic 2005). Usually, the rally effect will prevail over the fifth-column effect, because gainers from sanctions prevail over losers for the same reasons that protectionists typically prevail over free traders (Selden 1999). Conversely, pro-compliance fifth columns are more likely to prevail in response to promises of reward, which, unlike sanction threats, are not coercive and trigger no rally (Galtung 1967; Baldwin 1971; Long 1996).

This does not mean that reward promises are better than sanction threats. Many of the same authors also point to a common limitation of positive incentives–their vulnerability to extortion. It is the idea that offering rewards to the targeted country for quitting wrongdoing will lead this country to engage in more wrongdoing in the hope of obtaining larger rewards (Baldwin 1971; Bernauer 1999, 167, Haas and O'Sullivan 2000). Absent a clear-cut frontrunner, a partial and analytically uneasy consensus seems to have jelled around the notion that sanction threats and promises of reward are most efficient when used simultaneously (Amini 1997; Dorussen and Mo 1999; Cortright and Lopez 2000; Haas and O'Sullivan 2000).

We integrate these various components | sanctions and rewards, the rally and fifth-column effects, and the possibility of extortion | in a unifying model featuring a Target country with two groups, one that gains from sanctions but loses from rewards, and another that loses from sanctions but gains from rewards. Each incentive, depending on the circumstances, provides the ruler with the opportunity to extend political tenure. Our model has several advantages over existing ones. First it is more comprehensive, combining the diverse literatures on sanctions and aid into a single model. second, it tries to delineate precise mechanisms through which external incentives are transformed into policy outcomes. Third, it yields a counterintuitive and testable implications. The next section describes the model.

[2] Note that the literature on the impact of aid on regime type (see Morrison 2007 for an overview) is of no concern to us here, as aid is their independent variable and regime type their dependent variable.

3 The model

3.1 Payoffs

We assume that Target and Sanctioner are competing for a good of total worth $Z \in [0, \infty)$. (Target is a "he" , Sanctioner a "she.") Target can be either a security type that value the good being competed or an exxtortionist type that do not value the good.

(a) Nature chooses the type of Target between a security and an extortionist.
(b) Target moves by claiming $z \in [0, Z]$. Target's claim z represents an investment in a behavior that is deemed delinquent by Sanctioner.[3] Security type values z, Extortionist type does not value z.
(c) Sanctioner offers the incentive package with two components: a reward t and a sanction s, each one zero or positive. Sanction s is bounded upward, $s \leq S$ so as to rule out the option of threatening Armageddon.[4]
(d) Target chooses to Comply, Defy. Complying means giving up z) and defying means keeping z). We assume that threats and promises are enforceable.[5]

Sanctioner's utility function is that of a unitary actor with no particular a-priori preference for reward or sanction. Sanctioner merely finds either kind of incentive costly to implement: a rise in aid hurts taxpayers, whereas a drop hurts international lobbies, while a rise in trade and investment hurts domestic producers, whereas a drop hurts exporters and multinationals. There is always a group of discontented producers who punishes the government. Formally, Sanctioner maximizes

$$U = Z - z - \xi_1 t - \xi_2 s, \tag{1}$$

with respect to sanction s and reward t. ξ_2 and ξ_1, both strictly positive, are the marginal costs of implementing the sanction and the reward respectively.

Within the target country, two coalitions compete on the basis of relative wealth initially set to $1 - p$ for the internationalist side, to p for the nationalist side, with $p \in [0, 1]$. Moreover, the internationalist coalition benefits from reward t (more trade, investment, or aid) but is hurt by sanction s (less trade, investment, or aid) whereas it is the opposite for the nationalist coalition. As a result, the

[3] z could, for instance, be the share of a territory of total size Z that Sanctioner considers to be hers.

[4] Promising the moon is not an option either, but this possibility is endogenously ruled out by Sanctioner's maximization.

[5] This assumption is potentially problematic as one could argue that Target may not believe that Sanctioner would act upon threats and promises costly to her if she had to. Nevertheless, we assume perfect credibility. Credibility does not result from the way the present game is played, but it does result from the way the larger, unmodelled game would be played. Sanctioner is engaged in subsequent sanction games, involving other targets one at a time. She has an interest in establishing a reputation as credible sanctioner and the only way of doing so is by delivering on the threats and promises that she makes to any target. This is a standard result in reputation games of imperfect information; see Kreps and Wilson (1981).

internationalist coalition earns $W^{Int}|_{z,s,t,C} = 1 - p + \delta_1 t$ if its government complies and $W^{Int}|_{z,s,t,D} = 1 - p - \delta_2 s$ if its government defies whereas the nationalist coalition earns $W^{Nat}|_{z,s,t,C} = p - \delta_1 t$ in the case of compliance and $W^{Nat}|_{z,s,t,D} = p + \delta_2 s$ in the case of defiance. δ_1 and δ_2 capture the propensity of the regime to respond respectively to a positive and negative incentive.

The target government's payoff function shows two components. First, it is a positive function of the aggregate wealth of the winning coalition $g\left(W^{i^*}\right)$, with $i^* = Int, Nat$ identifying the winning coalition. Since this function can take about any form as long as it is positive, we simply write that Target maximizes the aggregate wealth of its winning coalition: $g\left(W^{i^*}\right) = W^{i^*}$. Second, as mentioned earlier, Target benefits from investing in the delinquent behavior z, with an expected benefit of bz and at a cost of cz^2 (b is the marginal gain and c a component of the marginal cost) A generic payoff function for Target can be written as

$$V = W^{i^*} + bz - cz^2, \text{with } i^* \text{the winning coalition} \tag{2}$$

The first component of the payoff varies with the nature of the regime that we categorize into three. Democracy is defined as a political regime where leaders are politically accountable to people who elected her/him; Limited autocracy is conceptualized as a political regime with partial accountability and thus leaders in limited autocracies enjoy some freedom to switch domestic political base; Dictatorship is conceptualized as a political regime with no accountability at all.

In democracies, elections force leaders to credibly identify with a party or a movement. They change their policy orientation at the risk of tarnishing their reputation, typically losing the support of their party without for all that gaining that of the opposition party. In contrast, the absence of free elections in autocracies enables a leader to experiment with policies that, at the time of their adoption, may not be popular with her support base. An autocratic ruler who would expect these presently unpopular policies to eventually be successful and elicit support later from other groups in the polity would have a rationale to pursue these policies in the first place. At the very end of the democracy-autocracy spectrum stands the absolute autocrat (dictator for short), a type for whom domestic support and tenure-maximization are of no immediate concern. Although no dictator is ever absolute in that extreme sense, some come very close like Kim Jong Un in North Korea. And as they do, they should not place much weight on the expected side effects of external incentives such as the rally round the fag and the fifth column.

Consider first the case of the limited autocracy. Target's payoff is a positive function of the aggregate wealth of the coalition that supports it *ex post,* that is, the internationalist coalition in case of compliance and the nationalist coalition in case of defiance. Hence, compliance yields $V_{aut}|_{z,s,t,C} = W^{Int}|_{z,s,t,C} - cz^2$ or, after substitution, $1 - p + \delta_1 t - cz^2$. Conversely, defiance yields $V_{aut}|_{z,s,t,D} = W^{Nat}|_{z,s,t,D} + bz - cz^2$ or, after substitution, $p + \delta_2 s + bz - cz^2$. The non-investment payoff is $1 - p$ if the internationalists are dominant *ex ante,* before Target chooses z, and p otherwise.

In a democracy, the government values what its present winning coalition values. If the internationalist coalition is in power at the outset, that is, $p < 1/2$, complying yields $V^{Int}_{dem}|_{z,s,t} = 1 - p + \delta_1 t - cz^2$, while defying yields $1 - p - \delta_2 s + bz - cz^2$.

Table 1 Target Government's payoffs, V_r

Regime	Comply	Defy
Int. demo $\left(p < \frac{1}{2}\right)$	$V^{Int}_{dem}\vert_{z,s,t,C} = 1 - p + \delta_1 t - cz^2$	$V^{Int}_{dem}\vert_{z,s,t,D} = 1 - p - \delta_2 s + bz - cz^2$
Nat. demo $\left(p > \frac{1}{2}\right)$	$V^{Nat}_{dem}\vert_{z,s,t,C} = p - \delta_1 t - cz^2$	$V^{Nat}_{dem}\vert_{z,s,t,D} = p + \delta_2 s + bz - cz^2$
limited autocracy	$V_{aut}\vert_{z,s,t,C} = 1 - p + \delta_1 t - cz^2$	$V_{aut}\vert_{z,s,t,D} = p + \delta_2 s + bz - cz^2$
Dictatorship	$V_{dic}\vert_{z,s,t,C} = 1 + \delta_1 t - cz^2$	$V_{dic}\vert_{z,s,t,D} = 1 - \delta_2 s + bz - cz^2$
	No Investment (NI)	
Int. demo $\left(p < \frac{1}{2}\right)$	$V^{Int}_{dem}\vert_{NI} = 1 - p$	
Nat. demo $\left(p > \frac{1}{2}\right)$	$V^{Nat}_{dem}\vert_{NI} = p$	
Limited autocracy	$V_{aut}\vert_{NI} = \max(1 - p, p)$	
Dictatorship	$V_{dic}\vert_{NI} = 1$	

If it is the nationalist coalition that is in power, i.e., $p > 1/2$, complying yields $p - \delta_1 t - cz^2$, while defying yields $V^{Nat}_{dem}\vert_{z,s,t} = p + \delta_2 s + bz - cz^2$. The non-investment payoff is $1 - p$ or p depending on the winning coalition.

In a dictatorship, the dictator values the aggregate wealth of the country. Like a unitary actor, he benefits from an aid increase, $V_{dic}\vert_{z,t,s,C} = 1 + \delta_1 t - cz^2$, and loses from a cut, $V_{dic}\vert_{z,s,t,D} = 1 - \delta_2 s + bz - cz^2$. The no-investment payoff is 1.

Target government's payoffs are gathered in Table 1.

3.2 Target type and extortion

A positive incentive, argues the literature, invites extortion on the part of the target. Extortion implies that Target is investing z at cost cz^2 for no other reason than to extract a reward from Sanctioner. Extortion is made possible by the fact that Sanctioner is not aware of the actual purpose of the investment. Sanctioner's ignorance is modeled by positing two possible types of Target government, randomly drawn from the set $\Theta_T = \{0, b\}$ featuring two types, a "security" type "S" with marginal gain for the investment in the delinquent behavior b greater than zero and an "extortionist" type "E" with $b = 0$. The labels refer to the situation in which investment z enhances the security of one type but has no intrinsic value for the other type. Nature draws the security type with probability h and the extortionist type with probability $1 - h$. Target knows its type, but Sanctioner only knows the probability distribution.[6]

[6] Not all extortion models require incomplete information. The present one does because we vest all negotiating power in the Sanctioner, who makes a take-it-or-leave-it offer to Target. It would be unnecessary if Target was making the offer, as such is the case in the mafia and corruption models by Polinsky and Shavell (2001), Schlicht (1996), and Bueno de Mesquita and Hafer (2008).

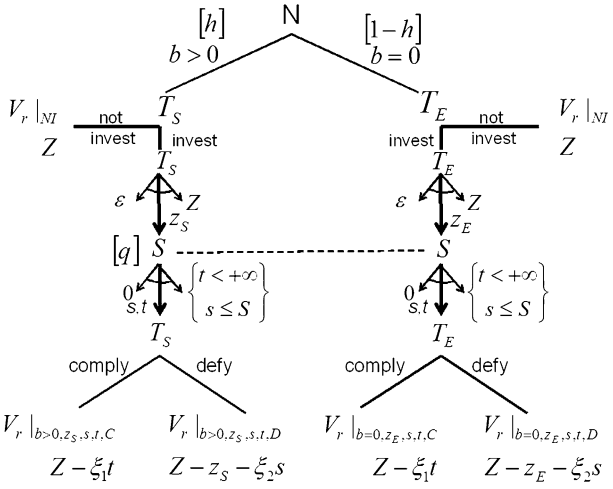

Fig. 1 Game Tree, with $r \in \left\{ {}^{I}_{dem}, {}^{N}_{dem}, aut, dic \right\}$

3.3 Tree, strategies, and equilibrium

We are now ready to provide a formal definition of the strategies and draw the tree (Fig. 1). A strategy for Sanctioner in this game is the mapping $\left\{\sigma_1 = (I, NI), z \in \zeta = R^+\right\} \rightarrow \left\{t \in T = R^+, s \in S = R^+\right\}$, specifying for an investment decision and each z value the values of t and s. A strategy for Target has two successive components. First, the mapping $\Theta_T \rightarrow \left\{\sigma_1, \zeta\right\}$ specifying for each type whether to invest or not, and, if the decision to invest is made, then the value of z. Second, the mapping $\Theta_T \times \left\{\sigma_1, \zeta\right\} \times \{T, S\} \rightarrow (c, d)$ specifying for each type, the decision to invest, each choice of z, and in response to all possible sanctioner's proposals, the decision whether to comply or defy.

We further deconstruct the target's choice of investment z into two steps: a first in which target chooses whether or not to invest in the delinquent behavior. If he chooses not to invest ($z = 0$), the game is over–Sanctioner cannot offer an incentive. Only if he goes ahead with the investment ($z \in [\varepsilon, Z]$), with ε and Z respectively the smallest observable and the largest possible investment in the delinquent behavior ($\varepsilon, Z > 0$), does he get to choose the actual value of z and can Sanctioner respond.

We denote Sanctioner's posterior belief about Target type by the conditional probability $q(b|z) = \Pr(b > 0|z)$; q is Sanctioner's updated belief, after having observed z, that target is of the security type. The equilibrium concept utilized is the Perfect Bayesian (PBE), which requires posterior beliefs to be calculated using Bayes' rule and each strategy to maximize expected utility given these beliefs and other players' strategies.

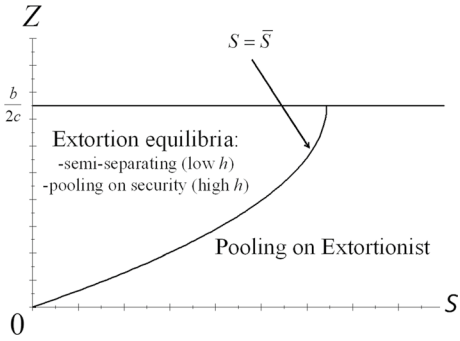

Fig. 2 Solution for free trade democracy

4 Analysis

We solve the game for each kind of regime, generating three propositions, which, together, show that limited autocracies are treated less harshly than either democracies or dictatorships. We start with the case of democracy with an internationalist winning coalition.

Proposition 1 (Internationalist Democracy equilibrium) *There are three PBEs:*

(1) *If* $S > \overline{S}$, *there is a pooling on the extortionist type's preference for not investing. Off the equilibrium path, Sanctioner offers the same incentives as in* (2); $q = h$.

(2) *If* $0 < S < \overline{S}$ *and* $h > \widehat{q}$, *there is a pooling on the security type's preference for investing* $z_S^* = \begin{cases} \frac{b}{2c} \text{ if } Z > \frac{b}{2c} \\ Z \text{ if } Z < \frac{b}{2c} \end{cases}$, *with Sanctioner offering* $t_S^* = \frac{bz_S{}^* - \delta_2 S}{\delta_1}$ *and* $s^* = S$, *while Target complies*; $q = h$.

(3) *If* $0 < S < \overline{S}$ *and* $h < \widehat{q}$, *there is a semi-separating in which the security type invests* z_S^* *while the extortionist type mimics him with probability* $g^* = \frac{h}{1-h} \frac{z_S^*(\delta_1 - b\xi_1) + s^*(\delta_1\xi_2 + \delta_2\xi_1)}{\xi_1(bz_S^* - \delta_2 s^*)}$ *and does not invest with probability* $1 - g^*$. *Upon seeing an investment, Sanctioner offers* t_S^* and $s^*(= S)$ *with probability* $r^* = \frac{cz_S{}^{*2}}{bz_S^* - \delta_2 s^*}$ *but* $t_E^* = 0$ *and* s^* *with probability* $1 - r^*$. *The security type complies in response to* t_S^* *but defies in response to* t_E^*; *the extortionist type always complies.* $q = \widehat{q} \equiv \xi_1 \frac{t_S{}^*}{z_S^* + \xi_2 s^*}$.

With $\overline{S} = \frac{bz_S{}^* - cz_S^{*2}}{\delta_2}$. *All results assume* $\delta_1 > b\xi_1$.

The proof of proposition 1 is offered in the Supporting Material and graphed in Fig. 2 along parameters S and Z on the horizontal and vertical axes respectively. We extensively develop the intuition behind the results, because the same reasoning underlies all the other proofs.

The target that is of the security type (thereafter, Security) is willing to invest in the delinquent behavior provided that Sanctioner does not threaten to implement a tough sanction. It is in Sanctioner's interest, however, to threaten the highest

sanction possible because sanction threats, unlike promises of reward, are costless, for they need not be acted upon in case of success. We thus expect Sanctioner to avail herself of the full extent of sanction available, that is, $s^* = S$. Now, S may not be enough of an incentive to elicit Target's compliance, forcing Sanctioner in such a case to supplement the sanction threat with a reward promise t^*. Sanction and reward, therefore, complement each other, with t^* being a reverse function of S.

Given that Sanctioner is always threatening the full value of S to elicit compliance, there will be values of S that are so high that Security will find himself better off not investing in the first place (when his compliance payoff is inferior to his do-nothing payoff). The area in which this obtains lays right of the curve drawn in Fig. 2 as $S = \overline{S}$.

Alternatively, one might easily imagine that when the maximum sanction allowed is too low, it would be Sanctioner who would rather do nothing. There are parametric specifications, indeed, for which there exists a second delimiting S-curve, this time close to the origin, below which Sanctioner remains inactive; it is just that the specification of marginals that we opted for in Fig. 2 and throughout the paper—we posited $\delta_2 > b\xi_1$—places this case out of the positive range.

To fully characterize the equilibria graphed in Fig. 2, we need to determine the extortionist type's (thereafter Extortionist) best reply to both Sanctioner's and Security's moves. Note, first, that, were the game one of complete information, Extortionist would never invest because, short of a positive incentive, which Sanctioner, in such a case, would have no reason to offer, not investing would always be better than investing. But incomplete information gives Extortionist the option within the interval $\left[0, \overline{S}\right]$ of improving upon his reservation value by mimicking Security. Hence, there exists a *pooling on Security equilibrium* in that interval. Sanctioner gets both types to comply by mixing bribe and sanction. An extortion rent is paid to Extortionist, because the bribe is calculated to buy Security's compliance, a compliance that is expensive because Security values the investment (his b is positive). Extortionist, who has no value for the investment (his b is zero) and would thus comply in exchange for nothing, extorts the same bribe.

The scheme works fine provided that Sanctioner believes that she is facing a security type with a high probability (h is high). However, if Nature failed to stack the deck with enough security types (h is low), then Sanctioner has a cheaper alternative at her disposal, one that could wreak havoc with Extortionist's plan: she could tailor the incentive, not to the costly security type, but to the cheaper extortionist type, giving just enough for Extortionist to comply, thereby canceling the rent component of the incentive. Note that it would make business sense for Sanctioner to act like this because the cheaper incentive would more than offset the occasional cases of defiance suffered from the rare security types. But for Extortionist, Sanctioner's counterstrategy would mean the end of extortion.

Unless, of course, Extortionist is smart enough to make his presence scarcer than Nature did initially. This is the essence of a semi-separating strategy. Basically, Extortionist randomizes the decision to invest, investing with probability g and not investing with probability $1 - g$, in such a way that, upon getting the opportunity to play, Sanctioner believes that she is facing a security type with a high probability,

irrespective of the low initial draw of h. By pumping back up Sanctioner's posterior belief q to a level that is high enough, Extortionist is able to bootstrap his payoff to that of the munificent pooling equilibrium, with two caveats. First, he receives the pooling payoff only g of the time. Second, the semi-separating equilibrium cannot hold unless Extortionist, himself, is indifferent between investing and not investing, for if he found investing better, then he would invest with certainty, unraveling his own randomization strategy. To help him commit to his mixed strategy, Sanctioner, in turn, must randomize between offering the high incentive that it takes to get Security to comply and the low incentive that suffices to get Extortionist to comply so as to bring the latter's payoff down enough to make him indifferent between investing and not investing.

Left of cutpoint $\overline{S}$, consequently, we have a pooling on Security equilibrium if Nature chose a high frequency of such types in the first place and a semi-separating equilibrium otherwise. Both equilibria involve extortion, in the sense that the state of uncertainty in which Sanctioner finds herself forces her to be more generous than if she knew the identity of her protagonist. The difference between the two equilibria is that, in the pooling equilibrium, Sanctioner is paying a rent (she pays a higher transfer on average than it would take to elicit compliance) and Extortionist cashing it. In the semi-separating equilibrium, Sanctioner is still paying a rent, but Extortionist is not making a real profit, for the transfer just covers his reservation value for not investing.

The case of democracy with a protectionist coalition yields starkly different results in terms of the kinds of incentives that are offered by Sanctioner:

Proposition 2 (Nationalist Democracy equilibrium) *There is one separating PBE in which the security type invests and defies and the extortionist type does not invest, while Sanctioner in either case offers no incentives.*

This is a case where incentives are inefficient, Security invests in the delinquent behavior, Extortionist does not, and Sanctioner does nothing.

The third regime, dictatorship, is very similar to the first case of democracy, in which internationalists are dominant. This similarity is easily read off Target's payoffs in Table 1, where the only difference is the first component of each payoff, $1 - p$ for democratic government, 1 for the dictator. As a result, the solution is very similar, with the minor exception of Target's payoffs. Proposition 1 and Fig. 2 apply equally well to free trade democracy and dictatorship.[7]

We last solve the game for the limited autocracy regime. The results significantly differ from all preceding ones. The size of the supporting coalition, which, in democracy, is exogenously given and, in dictatorship, irrelevant, changes endogenously in limited autocracy. The initial size of each coalition thus plays a central role in the results of Proposition 3.

[7] The formal proof in included in the Supporting Material.

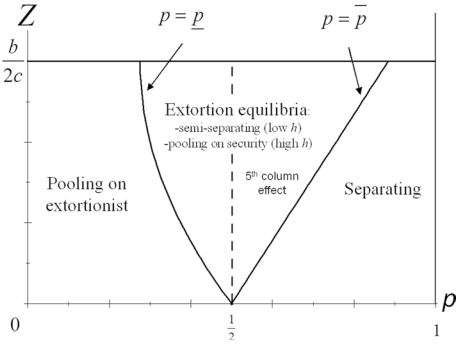

Fig. 3 Solution for limited autocracy

Proposition 3 (Limited Autocracy Equilibrium) *There are four perfect-Bayesian-Nash equilibria:*

(1) *If* $p < \underline{p}$, *there is a pooling on the extortionist type's preference for not investing. Off the equilibrium path, Sanctioner offers* $s = t = 0$; $q = h$.

(2) *If* $\underline{p} < p < \overline{p}$ *and* $h > \widehat{h}$, *there is a pooling on the security type's preference for investing* $z_S^* = \begin{cases} \frac{b}{2c} & \text{if } Z > \frac{b}{2c} \\ Z & \text{if } Z < \frac{b}{2c} \end{cases}$, *with Sanctioner offering* $t_S^* = \frac{2p+bz-1}{\delta_1}$ *and* $s^* = 0$, *while Target complies;* $q = h$.

(3) *If* $\underline{p} < p < \overline{p}$ *and* $h \leq \widehat{h}$, *there is a semi-separating in which the security type invests* z_S^* *while the extortionist type mimics him with probability* $g^* = \begin{cases} h\frac{\delta_1 z_S^* - \xi_1(2p+bz_S^*-1)}{(1-h)\xi_1 bz_S^*} & \text{if } \frac{p}{2} \geq 1 \\ h\frac{\delta_1 z_S^* - \xi_1(2p+bz_S^*-1)}{(1-h)\xi_1(p+bz_S^*-1)} & \text{if } \frac{p}{2} < 1 \end{cases}$ *and does not invest with probability* $1 - g^*$. *Upon seeing an investment, Sanctioner offers* t_S^* *and* $s^*(= 0)$ *with probability* $r^* = \begin{cases} \frac{cz_S^*}{b} & \text{if } \frac{p}{2} \geq 1 \\ \frac{1-2p+cz_S^{*2}}{bz_S^*} & \text{if } \frac{p}{2} < 1 \end{cases}$ *but* $t_E^* = \begin{cases} \frac{2p-1}{\delta_1} & \text{if } \frac{p}{2} \geq 1 \\ 0 & \text{if } \frac{p}{2} < 1 \end{cases}$ *and* s^* *with probability* $1 - r^*$. *The security type complies in response to* t_S^* *but defies in response to* t_E^*; *the extortionist type always complies.* $q = \widehat{h}$.

(4) *If* $p > \overline{p}$, *there is a separating equilibrium in which the security type invests* z_S^*, *while the extortionist type does not invest. Upon observing the investment, Sanctioner offers nothing and the security type defies;* $q = 1$.

With $\underline{p} = \frac{1}{2}\left(1 - bz_S^* + cz_S^{*2}\right)$, $\overline{p} = \frac{1}{2\xi_1}\left(\xi_1 + z_S^*\delta_1 - bz_S^*\xi_1\right)$, *and* $\widehat{h} = \frac{\xi_1(t_S^* - t_E^*)}{z_S^* - \xi_1 t_E^*}$. *All results assume* $\delta_1 > b\xi_1$.

Proposition 3 is proven in the Supporting Material and graphed in Fig. 3. The horizontal axis represents the initial strength p of the protectionist coalition, while the vertical axis represents the maximum investment Z. The reason for changing the parameter on the horizontal axis is that sanctions play no role in limited autocracy because a sanction threat may risk a perverse rally round the

flag. The initial size of the respective coalitions, in contrast, plays a determinant role in the results.

On the very left, the internationalists are a robust majority: p is low (below the lower cutpoint $\underline{p}$). Target government, irrespective of type, is already siding with the internationalists and will keep doing so in the future. There is no point in making the objectionable investment in the first place. Target invests nothing and the game is over. This is a case where Sanctioner relies on a majority that is favorable to maintaining open trade relations between the two economies to do her bidding. There is no reward and thus no extortion.

On the opposite side of the spectrum, the nationalists are a robust majority: p is high (above the upper cutpoint $\overline{p}$). Sanctioner cannot profitably engineer a shift to the free trade coalition by offering a carrot, for it would be too costly, costlier than doing nothing. As a result, the two types go their separate routes. Security, who is wired to benefit from the investment in the delinquent behavior, does invest and, absent any incentive from Sanctioner, then defies. Extortionist, who, in contrast to Security, has no use for the investment in the first place other than to extract a rent from Sanctioner, anticipating that no incentive will be proffered, shuns from investing.

Squeezed between these two cutpoints are the extortion equilibria, in which Security steadfastly invests in the delinquent behavior while Extortionist mimics, systematically or randomly, Security's investment, hoping to fool Sanctioner into buying him out of that investment for the same reward than Sanctioner is paying to Security. The extortion equilibria are twofold: a pooling on Security equilibrium for high values of h and a semi-separating equilibrium for lower values of h, according to a logic that is identical to that developed in Proposition 1 and need not be repeated here.

The two extortion equilibria feature a remarkable case of fifth-column effect: this effect occurs in the range where the protectionists are in power ex ante ($p > \frac{1}{2}$) and for values of p below the second cutpoint ($p < \overline{p}$). In this area, by means of a positive transfer, Sanctioner is able to engineer within the domestic politics of the target a power realignment away from the nationalist coalition toward the internationalist coalition, with the latter being supportive of compliance with Sanctioner's demand. This equilibrium makes a powerful case in theory for a pure engagement policy, even though the existence of a rent makes the engagement policy second-best as far as Sanctioner is concerned.

The starkest result, however, is that sanction threats are not part of any equilibrium solution, even though a sanction would, in other circumstances, make it unnecessary for Sanctioner to pay a rent. In other circumstances, a Sanctioner that is willing to threaten a sanction would typically be able to implement a screening strategy, by which he would lure Extortionist into complying while forcing Security to defy. In our game, the rally round the flag interferes with the freedom to punish, with the consequence that even such a screening strategy, with or without rent, provides Sanctioner with no optimal course of action.

Sanctions are never used, either because they are large enough to cause a rally effect, thereby causing defiance, an outcome that hurts Sanctioner, or because they are small enough not to cause a rally effect, but come at a cost nevertheless, which

Sanctioner would rather not pay. The cost is twofold: direct ($-\xi_1 s$), like any other incentive, and indirect as well, in the form of a higher compliance transfer[8]. The indirect cost reflects the idea that a sanction in this particular case is not the functional substitute of a reward but in effect cancels out the effect of the reward, thereby calling for a higher reward to extract the same level of compliance. In sum, large sanctions encourage rather than deter defiance, whereas small sanctions do nothing except drain Sanctioner's budget. Only bribes are used, because they reinforce the fifth-column coalition (internationalists) whose interests are aligned with Sanctioner's interest in extracting compliance.

The no-sanction result is robust to any kind of variation in the two marginals δ_1 and δ_2, which measure the propensity of the regime to respond respectively to a positive and negative incentive by means of a coalition realignment, as long as these marginals are greater than zero. The results are also robust to variations in ξ_1 and ξ_2, the respective marginal costs of the positive and negative incentives for Sanctioner. The fifth-column effect, in contrast, exists only if $\delta_1 > b\xi_1$.

5 Predictions

The model predicts a differential use of rewards and sanctions across regime types. We focus on the extortion equilibria, the only ones in which Sanctioner confronts Target with a set of incentives. These equilibria exist in the context of dictatorship, limited autocracy, and democracy controlled by an internationalist majority, all but democracy controlled by a nationalist majority where Sanctioner offers no incentives. Focusing on the first three regimes, limited autocracy stands apart from dictatorship and internationalist democracy in two ways. First, only a positive incentive is offered in limited autocracy, whereas sanctions are also threatened in the other two regimes. Second, the positive incentive is higher in limited autocracy ($t^*_{aut} = \frac{2p+bz^*_S-1}{\delta_1}$) than in the other two regimes $\left(t^*_{dic,dem} = \frac{bz^*_S-\delta_2 S}{\delta_1}\right)$; this is easily seen by setting the value of p in t^*_{aut} to its average value of one half. These two differences are captured in claims 1 and 2.

Claim 1 *Limited autocracies are offered high positive incentives only; dictatorships and democracies are offered moderate positive and negative incentives.*

Claim 2 *All countries comply in response to positive incentives, yet (1) dictatorship and democracy comply in response to modest positive incentives, whereas (2) limited autocracies comply in response to high positive incentives only.*

Both claims point in the same direction: limited autocracies are treated less harshly than democracies or dictatorships whenever they are offered incentives so they quit misbehaving.

[8] Raising sanction s by one unit means having to raise the compliance transfer t by δ_2/δ_1.

Although in equilibrium Sanctioner should not threaten a sanction toward a limited autocracy, in practice they do (Hufbauer 2007). This discrepancy between model and reality suggests that the model is leaving out important aspects of reality. Missing is the modeling of domestic politics for the sanctioner similar to that prescribed for the target. If the sanctioning country, like the target, had a preference for sanctioning, results would look strikingly different. For instance, Kaempfer and Lowenberg (1988) model the sanctioner as composed of two groups, one of them protectionist and lobbying for sanctions, while Baldwin (1971, 34) stresses voters' general dislike to reward criminal action : "When the North Koreans seized the Pueblo, it was "unthinkable" that President Johnson should offer to buy it back." There is no doubt that a more realistic model would have to incorporate domestic political constraints on the Sender's side of the kind these authors refer to. Nevertheless, our model is not devoid of interest for all that, but it allows us to offer the following corollary based on Target's behavior off the equilibrium path:

Claim 3 *If a sanction is imposed, dictatorships and democracies should comply, whereas limited autocracies should not.*

6 Empirical analysis: compliance to economic sanctions

We offer no empirical tests of the first two claims, leaving this stage to future research. Although there exists no dataset comprising both sanction threats and reward promises on which to test the full model, the fact that our argument regarding the effect of sanctions proved to be analytically robust to the introduction of positive incentives suggests that we could test the argument on a sanction-only dataset. We thus put the third claim to empirical test.

The empirical analysis examines how the three types of regimes respond to economic sanctions. If a sanction is imposed, we claim that dictatorships and democracies should comply, whereas limited autocracies should not. This is slightly different from the consensus in the literature, which views democracies as more responsive than dictatorships and limited autocracies lumped together (Bolks and Al-Sowayel 2000; Allen 2005; Hufbauer 2007).

We test our predictions on the Economic Sanctions Reconsidered dataset by Hufbauer (2007). We only consider unilateral sanctions by the U.S., dismissing sanctions involving multiple states and international organizations because the model assumes a unitary sender and we wish to remain consistent with the first test. This also allows to hold target related variables relatively constant.

We derive two versions of the dependent variable from the compliance *Score* in the Hufbauer (2007) dataset. Score is coded on a 16-point scale and created by multiplying two 4-point scale variables–the result of the episode and the contribution of the economic sanction to that result. Thus, when an economic sanction decisively brings a successful outcome, Score is equal to 16. In the very opposite case, it is equal to unity. We create two different dependent variables using Score. First, we create *Logged Score* by taking natural log of Score to cancel out the multiplying effect, which exaggerates the higher end of the scale. Second, we

Table 2 Success of economic sanctions: logit and regression models

Variable	Success	Log of score
Democracy	2.492**	.466 **
	(.906)	(.233)
Dictatorship	1.273**	.344**
	(.736)	(.185)
Prior relations	.211	.086
	(.612)	(.179)
Sender cost	− .433	− .142
	(.632)	(.132)
Tradelink	.009	.010
	(.029)	(.008)
Stability	− 1.221	− .191
	(.628)	(.130)
Constant	.273	1.718**
	(1.723)	(.493)
N	71	71
$R^2(Pseudo)$	.23	.20
$Prob. > \chi^2/F$	.007	.007

Standard errors in parentheses

*$p < .10$, **$p < .05$

create *Success*, a binary variable coding any value above nine as success and any below nine as failure.

The main independent variable is our trichotomous regime type variable. We expect that a democratic or a dictatorial regime are positively related to sanction success while an autocratic regime reduces the likelihood of sanction success.

We include as controls the variables that have been found most relevant by the literature. The first is the sanction cost, found by Dashti-Gibson et al. (1997) and Hufbauer (2007) to increase the likelihood of sanction success. The dataset offers two cost variables, one measuring the economic cost to the sender, the other that to the target. We only include the *Sender Cost* in the analysis, because available data for the cost to the target are clustered around zero, with a few observations significantly larger than zero, thus raising the risk of letting a few outliers drive the results. The Sender Cost ranges from 1 (little effect on sender) to 3 (major loss to sender). An additional variable, *Prior Relations*, controls for whether the two countries are friends or foes. Drezner (1999) argues that sanctions work better against friends than foes. The variable ranges from 1 (antagonistic) to 3 (cordial). Van Bergeijk (n.d.) and Hufbauer (2007) found that trade interdependence between sender and target increases the likelihood of success. We include the *Tradelink* variable in our analysis. Last, we control for economic health and political stability of a target. Dashti-Gibson et al. (1997) found that when a target is economically healthy and stable, sanctions are less likely to be successful. The *Stability* variable ranges from 1 (distress) to 3 (strong and stable).

A look at the summary statistics (not provided) indicate that only 28 percent of 71 U.S. unilateral sanctions are coded as successful. Of all cases, 41 percent are against dictatorships, 21 percent against democracies, and the residual, 38 percent, against limited autocracies.

Estimation results are reported in Table 2. We estimate two different models according to the dependent variable of use: a logit model on Success and an OLS model on Logged Score.

The logit model exhibits positive and statistically significant coefficients for Democracy and Dictatorship, 2.492 and 1.273 respectively. Since the excluded regime type among the three regimes is limited autocracy, the results can be interpreted as saying that sanctions against democracies and dictatorships are more likely to succeed than against limited autocracies. The same pattern is observable in the OLS model.

To get a sense of the substantive impact of regime type, we use the results of the logit model to calculate the predicted probability of sanction success while holding other variables at their respective means. When a target is limited autocracy, the predicted probability of sanction success is a mere 9 percent; when dictatorship, it significantly increases to 25 percent; when democracy, it further increases to 50 percent.

The results confirm existing findings that sanctions are more likely to work against democracies than non-democracies (Bolks and Al-Sowayel 2000; Allen 2005; Hufbauer 2007). Furthermore, the results conform with our prediction that dictatorships should behave like democracies more than limited autocracies. The discrepancy between our result and the standard result in the literature reflects the fact that the literature combines two distinct types of regimes (autocracy and dictatorship) together. If one only includes Democracy in the model (thus examining the effect of democracy vs. lumped up nondemocracies), one gets the literature's positive and statistically significant coefficient for the Democracy variable.

Among the control variables, none is statistically significant across specifications. Only Stability is statistically significant with the predicted negative sign in the logit model, suggesting that when a target is economically healthy and politically stable, the sanction is less likely to be successful. The coefficients for Prior Relations and Tradelink exhibit the correct signs but are not statistically significant.

7 Discussion: regime type, sanctions, rewards

A key recommendation of the sanction literature is that sanction threats and promises of reward be used simultaneously. This recommendation is usually supported by formal models of moral hazard. In a moral hazard situation, a principal writes a contract with an agent conditioning payment on the execution of a task. Because the amount of effort that is required to accomplish the task is unknown in advance and the actual effort expended by the agent is not directly observable by the principal, the principal ends up paying a rent to the agent; the contract is not efficient. Unless the principal is allowed to modify the incentive by adding a negative one to the promised payment in the form: "I pay you if you deliver, but you owe me

if you don't." Mixing sanctions with rewards make the resulting contract efficient (Laffont and Martimort 2002: 147). No rents need be transferred to agents, only payments covering compliance costs. Sanctions of that sort are typically illegal in contracts between an employer and an employee, but they are commonplace in relations between states. Does this mean that contracting in a world of anarchy is more efficient than contracting in a well-ordered domestic labor market?

Anyone who is familiar with the sanctions literature will scoff at this suggestion, for sanctions most often do not work (Hufbauer et al. 1990; Doxey 1996; Morgan and Schwebach 1997; Dashti-Gibson et al. 1997; Pape 1997; Drury 1998). One of the key reasons that have been underlined by the literature is the rally-round-the-flag effect, the fact that the threat of a sanction, let alone its implementation, arouses a defiant response within the target's government or population, making compliance with the sanctioner's demand more difficult (Galtung 1967). How much more difficult though? Is the rally round the flag a mere reputational nuisance that every sanctioner must endure with forbearance but that has no real impact on the target response? or is it a valid reason to disqualify the use of sanctions altogether? The answer to this question, as we show above, depends on the regime type of the target.

Our argument builds on a simplified polity inspired from that assumed by Bueno de Mesquita et al. (1999). The selectorate is made of two potential winning coalitions: one is trade, investment, or aid oriented, while the other is introverted. As in standard trade models, we distinguish between, on one side, export-oriented constituencies (merchants, exporters, services) and, on the other side, protectionist sectors (home-based industry, agriculture) (Galtung 1967; Rowe 2001; Selden 1999; Nincic 2005). Like trade, aid and investment may also split society into two coalitions, one made up of constituencies that thrive on aid (military, urban populations) or foreign investment (export-oriented zones) and, over time, become dependent on foreign capital for their well-being, and another coalition who is chaffing under the political dominance of the former group and opposing both the donor's aid and demands (Tokdemir 2017). While money is fungible, bilateral aid may not always be. The leaders who receive it are usually not free to redistribute it as they want but must abide by conditions laid out by donor governments, usually amounting to establishing a special trade or investment relationship between donor and beneficiary. Nevertheless, it is true that aid recipients can often divert the aid and use it for purposes other than those prescribed by the donor, making aid less of a political divider than trade. We refer to the two coalitions in the target country as "internationalist" and "nationalist" for short.

What is the impact of an external incentive, negative or positive, on the balance of power between the two coalitions? Leaving aside the rally-round-the-flag effect, an embargo of trade or aid threatens to undercut the relative wealth and power of the internationalist sectors while enriching and empowering the nationalist sectors. In contrast, more trade or aid promises to tilt the balance in favor of the internationalists. The additional effect of a rally round the flag in response to an embargo is to skew the balance in favor of the nationalist side.

How does this strengthening of the nationalist side affect the incumbent government's response to the sanctioner's demands? Intuitively, it should make compliance more difficult. We argue that it actually depends on the type of regime. If the regime

is democratic, it makes no difference, irrespective of which coalition is in power. If the internationalist coalition is in power, first, whenever faced with a sanction threat, their leader will comply to avoid the imposition of a sanction that would weaken the coalition and endanger his or her tenure. There is no reason to believe a priori that the nationalist surge caused by the sanction threat would make a dent in the internationalist coalition. Second, if the nationalist coalition is in power, whenever faced with a threat, their leader will invite the imposition of the sanction because it would strengthen the relative control that her supporting coalition enjoy over the economy. Therefore, a rally round the flag has no significant impact on the leaders's policy orientation in a democracy.

In contrast, if the regime is nondemocratic (without being dictatorial), then the threat of a sanction could actually lead to a realignment in coalition and policy orientation. Of course, this is not going to be the case if the nationalists are in power, for the imposition of the sanction will strengthen that coalition and secure their leader's tenure. But this could be the case if the internationalists were in power when the sanction threat is aired. In this situation, the leader would have two options: remain true to her internationalist coalition and comply, or betray the internationalists and defy. The latter option is tempting in view of the weakening of the internationalist coalition that an implementation of the sanction portends.

A nondemocratic leader can afford the luxury of jumping coalition but a democratic leader cannot. This proposition indirectly comes out of Bueno de Mesquita et al. (1999)'s selectorate model, on which we built our theoretical model presented above. They argue that coalitions in autocracies are loyal to their leaders because the smallness of the winning coalition in relation to the selectorate make them redundant. Conversely, in a democracy, they argue, the fact that the winning coalition is a larger subset of the selectorate makes the coalition less loyal. Because loyalty is a zero-sum game between leaders and constituents in the sense that "your loyalty to me relieves me from my loyalty to you", it follows that leaders are less loyal to their winning coalition in autocracies than they are in democracies.

Logically implacable, the argument is intuitive too: imagine an American president switching partisan allegiance in the course of his first term; decried by his former party while mocked by the other party, his chances of reelection would be nil.

Jumping coalition in autocracies, however, is not uncommon. A famous historical example is Prussian chancellor Bismarck's historical switch in the mid 1870s from a free trade policy that was supported in Parliament by a coalition of Liberals along with Socialists and Catholics to a policy of protection that had the support of the Conservative junkers and industrialists.

A more recent instance of similar reversal occurred in 2005 Iran, when Supreme Leader Ayatollah Khameini decided to terminate the reformist experiment under president Mohammad Khatami and throw his support behind the antiwestern government of Mahmoud Ahmadinejad. Although Bismarck was not reacting to a sanction threat–only to a world recession–Khameini was. Four rounds of UN sanctions subsequently helped consolidate the regime realignment from the cosmopolitan reformists to the nationalist hard-liners. The sanctions provided the new government with the means and rationale to build up the political and economic power of a para-military organization–the Revolutionary Guards–an organization which, today,

controls the country's strategic missile forces, with ties to companies in oil, construction, telecommunications, and weapons manufacture as well as black market enterprises smuggling embargoed products, alcohol and nuclear fuel in particular.[9]

We argue that rallies round the flag should not affect the incumbent's policy orientation in democracies but might in nondemocracies. We also consider a third type of regime located at the very end of the democracy-autocracy spectrum: the absolute autocrat (dictator for short), a type for whom support and tenure-maximization are of no immediate concern. Although no dictator is ever absolute in that extreme sense, some come very close, such as Kim Jong-un in North Korea. And as they do, they should not place much weight on the expected side-effects of sanctions, such as the rally round the flag. Instead, a dictator should privilege two dimensions: first, the sanctioner's actual demand per se–what costs and benefits are; second, unlike leaders who depend on coalitional rivalry to stay in power, dictators have an interest in maximizing the welfare of all groups in society, for they stand in the position of residual claimant of their subjects' output–they have a monopoly over property rights (Findlay 1990; Barzel 2000).

Again, we do not mean to say that rallies never occur in dictatorships, they do, only that, whenever they occur, they merely play into the hands of the leader, with no impact on the policy orientation of the regime.

Therefore, we agree with the literature that a sanctioner's sanction threat will in most cases provoke a rally round the flag, but we believe that such rally will blunt the impact of the sanction threat in only one type of regime: limited autocracies. It should have no negative impact at the two extremes of the accountability/loyalty continuum: democracies and dictatorships.

8 Conclusion

What incentives work best against what regimes? The literature on regime type and incentive format has it that sanctions work better against accountable types of government whereas rewards work better when directed to unaccountable ones. This is because unaccountability allows autocrats to brush off the pain and dissent caused by the sanction while accountability implies that democratic leaders have little use for a mere handful of bribes.

We modify this clear-cut dichotomy, arguing instead that autocratic regimes with a measure of accountability that falls in between accountable democracies and downright unaccountable dictatorships, have the opportunity to ride external incentives to their advantage, engineering a rally round the flag in response to a sanction or a fifth column in response to a reward. This makes them sui generis and

[9] While the election of Hassan Rouhani to the presidency in 2013 led to a deal with the Obama administration and its European partners that resolved the nuclear standoff, it has not translated into any form of retrenchment from Iran's involvement in regional conflicts. Decades of sanctions and isolation have not led to a reduction in Iran's support of Hezbollah in Lebanon, the Shiite militias in Iraq, the Houthis in Yemen, the Taliban in Afghanistan, and Assad in Syria.

actually less responsive to sanctions than the other two types of regime. The reason is that in both democracies and dictatorships, rewards and sanctions are functional substitutes: the greater the sanction can the sanctioner inflict, the lower the reward needs the sanctioner afford. In contrast, in limited autocracies, where sanctions risk backfiring, rewards and sanctions work at cross purposes: the higher the sanction is, the higher the reward needs to be to cancel out the risk of rally round the flag that is caused by the sanction.

Therefore, it is intermediate regimes like Iran, regimes that are neither quite democratic like Israel nor absolutely autocratic like North Korea, that are the least responsive to sanctions. Only the promise of large rewards is likely to work, putting Washington in a difficult policy position because promising a large sum of rewards toward a country like Iran is bound to be politically unpalatable.

The non-linear relation between regime type and compliance questions the rationale that has been given in the literature to account for the different responses offered to external incentives by democracies and non-democracies. The currently-held rationale that rulers in democratic regimes enjoy less freedom of maneuver than rulers in non-democratic regimes may be necessary to explain why democracies are strongly responsive to incentives but is insufficient to explain why dictators in particular are equally responsive to the same incentives. Key to the non-linear result is the idea that dictators need not worry about tenure but, instead, can enjoy the advantages that come with the status of residual claimant to their subjects' output, a status that makes them responsive to the overall welfare of society.

References

Allen SH (2005) The determinants of economic sanctions success and failure. Int Interact 31(2):117–138
Allen SH (2008) Political institutions and constrained response to economic sanctions. Foreign Policy Anal 4:255–74
Amini GM (1997) A larger role for positive sanctions in cases of compellence. Working Paper No. 12. Center for International Relations, University of California at LA
Baldwin D (1971) The power of positive sanctions. World Polit 24(1):19–38
Barzel Y (2000) Property rights and the evolution of the state. Econ Govern 1(1):25–51
Bernauer T (1999) Positive incentives in nuclear proliferation and beyond. In: Bernauer T, Ruloff D (eds) The politics of positive incentives in arms control. University of South Carolina Press, Columbia, pp 157–191
Bolks SM, Al-Sowayel D (2000) How long do economic sanctions last? Examining the sanctioning process through duration. Polit Res Q 53(2):241–65
Brooks R (2002) Sanctions and regime type: what works, and when? Secur Stud 11(4):1–50
Bueno de Mesquita B, Morrow J, Siverson R, Smith A (1999) An institutional explanation of the democratic peace. Am Polit Sci Rev 93(4):791–807
Bueno de Mesquita B, Smith A (2007) Foreign aid and policy concessions. J Conflict Resolut 51(2):251–84
Bueno de Mesquita E, Hafer C (2008) Public protection or private extortion? Econ Polit 20(1):1–32
Cortright D, Lopez GA (2000) The sanctions decade: assessing UN strategies in the 1990s. Lynne Rienner, Boulder
Cox DG, Drury AC (2006) Democratic sanctions: connecting the democratic peace and economic sanctions. J Peace Res 43(6):709–722
Dashti-Gibson J, Davis P, Radcliff B (1997) On the determinants of the success of economic sanctions: an empirical analysis. Am J Polit Sci 41:608–618

Dorussen H, Mo J (1999) Sanctions and incentives. Paper presented at the 1999 annual meeting of the American Political Science Association, Atlanta, GA, 2–5 September, 1999
Doxey M (1996) International sanctions in contemporary perspective, 2d edn. St. Martin's Press, New York
Drury AC (1998) Revisiting economic sanctions reconsidered. J Peace Res 35(4):497–509
Findlay R (1990) The new political economy: its explanatory power for LDCs. Econ Polit 2(2):193–221
Galtung J (1967) On the effects of international economic sanctions: with examples from the case of Rhodesia. World Polit 19(3):378–416
Haas RN, O'Sullivan ML (2000) Terms of engagement: alternatives to punitive policies. Survival 42(2):113–135
Hufbauer G, Schott J, Elliott KA (1990) Economic sanctions reconsidered: history and current policy, 2nd edn. Institute for International Economics, Washington DC
Hufbauer CG et al (2007) Economic sanctions reconsidered, 3rd edn. Peterson Institute for International Economics, Washington DC
Kaempfer WH, Lowenberg AD (1988) The theory of international economic sanctions: a public choice perspective. Am Econ Rev 78(4):786–793
Kreps DM, Wilson R (1981) Reputation and imperfect information. J Econ Theory 27:253–79
Lai B, Morey DS (2006) Impact of regime type on the influence of U.S. Foreign Aid. Foreign Policy Anal 2:385–404
Laffont J-J, Martimort D (2002) The theory of incentives: the principal-agent model. Princeton University Press, Princeton
Lektzian D, Souva M (2003) The economic peace between democracies: economic sanctions and domestic institutions. J Peace Res 40(6):641–60
Lektzian D, Souva M (2007) An institutional theory of sanction onset and success. J Conflict Resolut 51(6):848–71
Long WJ (1996) Trade and technology incentives and bilateral cooperation. Int Stud Quart 40:77–106
McGillivray F, Stam AC (2004) Political institutions, coercive diplomacy, and the duration of economic sanctions. J Conflict Resolut 48(2):154–72
Morgan TC, Schwebach VL (1997) Fools suffer gladly: the use of economic sanctions in international crises. Int Stud Quart 41:27–50
Morrison KM (2007) Natural resources, aid, and democratization: a best-case scenario. Public Choice 131:365–86
Nincic M (2005) Renegade regimes: confronting deviant behavior in world politics. Columbia, New York
Pape RA Jr (1997) Why economic sanctions do not work. Int Secur 22(2):90–136
Peksen D, Drury AC (2008) Coercive or corrosive: the negative impact of economic sanctions on democracy. Working paper
Polinsky AM, Shavell S (2001) Corruption and optimal law enforcement. J Public Econ 81:1–24
Rowe DM (2001) Manipulating the market: economic sanctions, institutional change, and the politics of white rhodesia. The University of Michigan Press, Ann Arbor
Schlicht E (1996) Exploiting the coase mechanism: the extortion problem. Kyklos 49(3):319–330
Selden Z (1999) Economic sanctions as Instruments of American Foreign Policy. Praeger, Westport, Conn
Tokdemir E (2017) Winning hearts & minds (!): The dilemma of foreign aid in anti-Americanism. J Peace Res 54(6):819–32
Verdier D, Woo B (2011) Why rewards are better than sanctions. Econ Polit 23(2):220–238
Wood RM (2008) A hand upon the throat of the Nation': economic sanctions and state repression, 1976–2001. Int Stud Quart 52:489–513