

Arbitration in Anarchy

John Ferejohn (NYU)

Tom Ginsburg (U Chicago)

Richard Holden (U NSW)

February 6, 2020

1 Introduction

There has in recent years been a resurgence of literature on arbitration in anarchy, that is, dispute resolution even in the absence of state power (Benson 1989; Hadfield and Weingast 2012, 2013). This literature has produced a good deal of insight into primitive legal systems, showing us how coordination between disputants can be facilitated even without coercive force (McAdams 2015). It draws on game theoretic models in which third party punishment arises through repeated interactions, typically in a world of strong and weak types. But this literature has left several questions unanswered: how do strong and weak types emerge in the first place? Why do we observe variation in the willingness of parties to submit to arbitration and comply with decisions? And why is it in the interest of a third party to provide the dispute resolution service? These are the questions we tackle in this paper, with a unified framework that explains arbitration in anarchy.

We start with our own model of how neutral dispute resolution can arise endogenously. In the absence of state enforcement, arbitration can only be effective if three conditions are met. First, the arbitration must be incentive-compatible *ex ante*; normally an arbitrator cannot be appointed unless the disputing parties make a request, and cannot succeed unless the parties participate in the process. Thus, an arbitrator always needs some consent from the parties. Second, the arbitration must be *ex post* incentive-compatible: the parties must accept the settlement in the sense of conforming to its terms. This requires the arbitrator to craft a self-enforcing settlement which must take account of the

three problems, focusing primarily on the problems of incentive compatibility. The intuition is that dispute resolution saves the costs of fighting, and we show why it is often incentive-compatible for the loser to comply, as well as the dispute resolver to expend resources resolving the dispute. This is consistent with prior literature although our model is novel.

We then go on to make two additional novel contributions. First, we show how a diversity of types of players can arise endogenously, rather than simply allowing nature to distribute the types. Second, we show that arbitration is not always the equilibrium outcome, and that under some circumstances, the presence of arbitration as an option might actually induce more conflict, because of its incentive effects on the decision of whether or not invest in strength.

Our framework helps understand a wide variety of situations in which arbitration is found in the absence of centralized government. We provide several examples. Indeed, in the last section we speculate how dispute resolution of this type may lead to the accumulation of power over time. We thus provide a novel theory of the state, distinct from the standard accounts that focus on war and taxes. Law, it turns out, can make states too, though in most observed circumstances it is complementary to power, rather than a pure substitute.

2 Background

In a series of papers, Hadfield and Weingast (2010, 2012, 2013) advance our understanding of early law by focusing on the distinct aspects of law as a dispute resolution mechanism. They emphasize that law is characterized by a normative classification system, with a central institution to classify behavior as being in or out of a particular rule. Their framework emphasizes decentralized collective punishment, which is coordinated by a central institution. They further draw on the work of Lon Fuller (1964) to elaborate the attributes of successful resolution of the coordination problem. Hadfield and Weingast have contributed a good deal to our understanding of these issues, but their focus is slightly different from our own. They take the set of disputes in a society as given, and do not focus on the questions of where disparate types come from, who pays the classification institution, or how that institution is incentivized to resolve disputes. Our model covers some of the same territory as theirs but grapples with these additional issues, along with others.

To motivate the discussion, we begin with a famous story, drawn from the Shield of Achilles, as described in the Iliad. Achilles shield features several scenes of urban life

in times of peace and war. One of the scenes of peace concerns dispute resolution:

And the people massed, streaming into the marketplace where a quarrel had broken out and two men struggled over the blood-price for a kinsman just murdered. One declaimed in public, vowing payment in full – the other spurned him, he would not take a thing – so both men pressed for a judge to cut the knot. The crowd cheered on both, they took both sides, but heralds held them back as the city elders sat on polished stone benches, forming the sacred circle, grasping in hand the staffs of clear-voiced heralds, and each leapt to his feet to plead his case in turn. Two bars of solid gold shone on the ground before them, a prize for the judge who'd speak the straightest verdict."

The scene describes early dispute resolution, which in pre-state societies typically involves the crowd or the community as a whole. But it also identifies a figure called a judge, in this case a non-professional one, who tries to resolve the dispute with knowledge. The Achilles story suggests that there is value in epistemic quality. The ability to predict the best solution to a conflict is to be rewarded, as it saves the resources involved in conflict.

In this case the crowd is a kind of constraint on the parties, disincentivizing the loser from ignoring the decision, and so complements the judge in his decision. In other circumstances, a crowd might have a different effect. Without a designated process of resolving the dispute, the crowd might incentivize fighting. E.E. Schattschneider started one of his little books by describing a fistfight between two men in front of a crowd of people who were cheering and booing the participants. He argued that this situation was unstable because once one of the parties started to get the worst of it, he had a strong motive to appeal to the third party, the crowd, for help. What was important about the crowd is that it had the brute power to determine the outcome. Schattschneider argued that bilateral conflict is inherently unstable in this sense always conferring on the weaker party an incentive to try to widen the conflict in order to try to reverse an unfavorable balance of power. The presence of the crowd to break up the fight might actually incentivize fighting. We build on these stories to address the following questions: what determines the level of fighting in the absence of government? Under what circumstances might arbitrators be effective without power? And why are they motivated to provide their service?

3 The Models

In this section we present a two formal models of conflict resolution. The first model allows assumes that there is no possibility of arbitration and all conflicts lead to fighting. The second model allows for the possibility of mutually agreed upon arbitration to prevent fighting.

By comparing these two models and their equilibria, we are aiming to answer two questions: first, how do endogenous distributions of power within these societies change with and without the possibility of arbitration; and second, how does the probability of a fight ensuing change with and without the possibility of conflict.

3.1 Without arbitration

In this section we will analyze a game of conflict without the possibility of arbitration. Specifically, there are N ex ante symmetric players who will compete pairwise to win a prize with value b . Each player starts as "weak" but can pay to (with certainty) become "strong". These strength levels determine the relative costs and rewards of conflict.

The game proceeds as follows:

1. The players simultaneously choose whether or not to pay x to become strong.
2. The players choose whether to play "F(ight)" or "DF (Don't Fight)".
3. Players are randomly matched against each other, without full information on their opponents type. If both players in a match have chosen F then a conflict occurs, else if one or no players chose to fight the prize is awarded or divided without a fight.

When a conflict occurs strong players incur a (small) fighting cost of c_l whereas weak players incur a (large) fighting cost of c_h . To model this relative strength we assume $c_h > c_l$. When two players of equal strength fight, they each win $b/2$. When a strong player fights a weak player, the strong player wins the entire prize of b and the weak player wins nothing. When only one player chooses to fight while the other chooses not to fight, the fighting player wins all of b regardless of types without paying the fighting cost. If neither player fights then neither player wins nor pays anything.

We search for equilibria of the following form:

- At the first stage, all players play the mixed strategy "pay x and become strong with probability $q \in (0, 1)$, and remain weak with the complementary probability $1 - q$.

- At the second stage, all strong players choose to fight with certainty, whereas weak players play fight with probability $r \in (0, 1)$ and not fight with the complementary probability $1 - r$.

When x , the cost of becoming strong, is sufficiently small, in particular when $x < \frac{b}{2} - c_h$, if we fix the strategies in stage two as described above, then at stage 1 the game is a prisoner's dilemma with the strictly dominant strategy being to always become strong. Hence we rule this case out by assuming $x > \frac{b}{2} - c_h$ to allow for an endogenous distribution of power that is non-trivial.

Subject to this assumption, and others to ensure that the parameters that define the mixed strategies are indeed probabilities between 0 and 1, we obtain the following equilibrium characterization result.

Proposition 1. *Under mild regularity conditions, there exists a unique equilibrium of the form described above, defined by the mixing probabilities*

$$q_{NA}^* = \frac{2(bB - Bx - 2c_l x)}{bB}, \quad r_{NA}^* = \frac{2b(b + 2c_h - 2c_l) - 4x(b + c_h)}{b(b + 2c_h - 2c_l) - 2x(b + 2c_h)}$$

where $B = b + 2c_h - 2c_l$.

The assumptions required for this equilibrium to be sensible, and the proof of the proposition, are in the appendix. Given these assumptions, the comparative statics of these equilibrium mixing probabilities are relatively straightforward and intuitive. They are summarized in the following proposition, the proof of which is also in the appendix.

Proposition 2. *The equilibrium probability with which each player becomes strong, q_{NA}^* is:*

- Decreasing in the cost of becoming strong x
- Increasing in the cost of fighting for weak players c_h
- Decreasing in the cost of fighting for strong players c_l .

On the other hand, the equilibrium probability with which each weak player chooses to fight, r_{NA}^ is:*

- Increasing in the cost of becoming strong x
- Decreasing in the cost of fighting for weak players c_h
- Increasing in the cost of fighting for strong players c_l .

These comparative statics results are intuitive. It says that there will be fewer strong players in equilibrium when: getting strong is cheaper, the fighting cost if one stays weak is higher, or the fighting cost if one becomes strong is lower. Conversely, weak players will fight exactly when they expect the probability of meeting a strong opponent is low, and as such the comparative statics for the probability with which weak players fight are exactly in the opposite direction to the proportion of strong players in the game.

3.2 With arbitration

Now, in order to isolate the effects of introducing an arbitration mechanism with the potential to avoid conflict, we study the following game, in which an arbitrator is assumed to have epistemic power to discern the players' types. The sequence of play in this extended game is as follows.

1. Players choose whether to pay x and become strong.
2. Players choose whether to contest for the prize or not.
3. If both players contest, players choose whether to agree to arbitration or not.
4. If at least one player does not agree to arbitration, players choose whether to fight or not.

Similarly to before, we will look for an equilibrium with a non-trivial endogenous distribution of power, and non zero probabilities of both arbitration and fighting by both types.

Specifically, ignoring stage 2, we are looking to construct an equilibrium of the following form:

- At stage 1, everyone plays the same mixed strategy 'get strong (at cost x) with probability q .
- At stage 3 strong types choose to arbitrate with probability w and weak types choose to arbitrate with probability y .
- At stage 4, if it occurs:
 - Strong types always fight.
 - Weak types who played NA in stage 3 have learned nothing about q , and so they mix and play fight with probability r .

- Weak types who played A in stage 3 but nevertheless found themselves in stage 4, update their posterior to $Q > q$ thus never fight.

The intuition is (we expect) weak types to arbitrate more than strong types in equilibrium $y^* > w^*$ (since they are hurt more by fighting) and if weak types attempt to arbitrate but do not get agreement then they update upwards their belief about their opponent being strong, since it is the strong types who elect not to arbitrate more often.

Indeed, a unique equilibrium of the form above exists, subject to regularity conditions. The structure of this equilibrium is given in the following proposition.

Proposition 3. *Under mild regularity conditions, there exists a unique equilibrium of the form described above, defined by the mixing probabilities*

$$\begin{aligned} w_A^* &= \frac{b(b-2k)(2x(c_l(b+2c_h)-kB) - B(b(c_l-k) + 2c_lk))}{2(b+2c_h)(bc_l - k(b+2c_l))(b(c_l+k) - 2k(c_l+x))} \\ q_A^* &= \frac{b(c_l+k) - 2k(c_l+x)}{bc_l} \\ r_A^* &= \frac{1}{2}b \left(\frac{1}{c_l} - \frac{1}{k} \right) + 1 \\ y_A^* &= \frac{bB(b(c_l-k) + 2c_lk) - 2bx(c_l(b+2c_h) - kB)}{(b+2c_h)(bc_l - k(b+2c_l))(b - 2(c_l+x))} \end{aligned}$$

where $B = b + 2c_h - 2c_l$.

The comparative statics in this case are far less clear and typically indeterminate in sign.

3.3 Model comparison

Given the sharp characterization of equilibrium play in these two games: one with and one without the possibility of arbitration, we are now able to offer a more complete analysis of how arbitration affects the endogenous strength distribution and the probability of fighting in some senses.

We wish to know whether the equilibrium q from the first model without arbitration,

$$q_{NA} = \frac{2(bB - Bx - 2c_lx)}{bB}$$

is larger or smaller than the q from the second model:

$$q_A = \frac{b(c_l+k) - 2k(c_l+x)}{bc_l}.$$

We have the following,

Proposition 4. *Under the assumptions of propositions 1 and 3, we have $q_{NA}^* < q_A^*$. That is, the endogenous power distribution when arbitration is a possibility features more powerful types than when arbitration is not a possibility.*

The intuition for this is relatively clear: arbitration makes contestation cheaper for those who would otherwise fight, and the effect of that at the first (investment) stage is to induce more players to invest in strength in order to gain an advantage in arbitration.

Next, we wish to compare the probabilities of fighting occurring under both scenarios. The first sensible comparison is a like-for-like comparison of the equilibrium values of r in both models. This parameter represents the probability that a weak type, who believe that their potential opponents are strong with probability q , will choose to fight. This comparison allows a direct comparison of how actions are chosen in a part of the game tree common to both models. We have the following:

Proposition 5. *Under the assumptions of propositions 1 and 3, we have $r_{NA}^* > r_A^*$. That is, holding beliefs constant, weak types are more likely to fight without arbitration than with arbitration.*

To understand this, note that when it comes to fighting the situation under arbitration features one more eventuality than without arbitration. Without arbitration, all weak types mix and all strong types fight. On the other hand, with arbitration, in addition to strong types always fighting and the weak types who mix, some weak types (those who have attempted to arbitrate but been rebuffed) update their posteriors of the proportion of strong types in the population upward, and then with certainty choose not to fight.

And so, for a weak type with unchanged beliefs, there are two incentives in opposite directions that may induce more or less fighting. On the one hand, here there exist potential weak type opponents who will surrender without a fight, making fighting more attractive. On the other, we have seen above that there are more strong types in this power distributions under arbitration. Evidently, the second effect dominates the second, such that the net effect of introducing arbitration is that these mixing weak types fight more often as they are afraid of the increased chance of meeting a strong type.

Finally, we would ideally like to compare the ex ante probability of a fight occurring in both models. This is more complex than a simple comparison of equilibrium parameters owing to the fact that the contingencies that lead to fighting differ between models.

In particular, when there is no possibility of arbitration, a fight occurs in three ways: either two strong people meet, or a strong person meets a weak person who is choosing to fight, or two weak people meet and both choose to fight. Thus the overall probability of a fight breaking out is

$$Prob_{NA}(Fight) = q_{NA}^2 + q_{NA}(1 - q_{NA})r_{NA} + (1 - q_{NA})^2r_{NA}^2.$$

On the other hand, when there is the possibility of arbitration, a fight can still break out either between two strong players, or two weak players, or one of each. To be specific, either both are strong and at least one chooses not to arbitrate, or one is strong and one is weak and the weak player chooses not to arbitrate but then to fight, or both are weak and both choose not to arbitrate and then both choose to fight. Thus the overall probability of a fight breaking out is

$$Prob_A(Fight) = q_A^2(1 - w_A) + q_A(1 - q_A)(1 - y_A)r_A + (1 - q_A)^2(1 - y_A)^2r_A^2.$$

Unfortunately, no tractable sufficient or necessary conditions are attainable that simplify the comparison of these two total probabilities to any substantial effect. For instance, one might attempt a comparison of each of term in each expression that corresponds to the ex-ante probability a fight breaks out between two strong types. Whereas there are more strong types under arbitration than without, not all strong types fight, given there is a chance they resolve their dispute in arbitration. Similarly, the comparisons between the ex-ante probability of a fight breaking out between a strong and a weak type, or two weak types have factors pushing in both directions that cannot cleanly be resolved.

4 Applications

Our contribution is primarily theoretical but in this section, we provide several examples of historical and contemporary problems that our model helps to elucidate.

4.1 The Mafia

Mafia organizations consist of players who, vis-a-vis the external society, have invested in strength, and band together to prey on weak types. Through their collective action, they are able to extract rents from illegal economies and to provide protection services to the weak. Gambetta (1993) suggests that the Sicilian mafia arose in conditions of near-anarchic state weakness after a 19th century transition to a market economy, in which an excess of strong types created a demand for protection. The combination of formal property rights and weak state enforcement, he suggests, will generate demand for mafia.

Mafia organizations fit into our model in several ways. First, the organization itself,

made up of strong types, must resolve internal conflicts that arise. One way in which groups resolve this problem is to have detailed rules about secrecy and honor, which discipline the internal members of each mafia family. The rules provide for discipline and even trials in the event of the violation of the code of honor within each family. There is also strict enforcement against outsiders who impinge on the property rights of a family. In a particular territory, no criminal activity can take place without the agreement of the family.

Second, mafias can provide arbitration services to those within their territory or under their protection. Chinese triads, Japanese yakuza, and Italian mafia are all known to provide dispute resolution services for legitimate businesses, and in some cases are seen as efficient vis-a-vis cumbersome and costly state courts (Milhaupt and West 2000; Gambetta 1993: 80). For example, a famous Italian mafia boss in Milan (known as the Uncle) acquired a reputation for resolving disputes quickly and effectively, for a fee (Cattino 2019: 123). While the threat of violence may have underpinned his services, he accumulated power through his epistemic ability. And he was paid for it.

Third, the gangs themselves can be modeled as strong types that confront each other. The problem of inter-family violence is much more difficult to solve than intra-family violence. Mafias compete with each other over territory and business, and even an agreement allocating territories may not be stable as one group becomes stronger and another weaker as time passes from the original agreement. Fights among strong types create a risk of violence, which in this context could draw the attention of authorities and undermine everyone's ability to extract rents. Mafia gangs therefore have an interest in cooperating to reduce conflict among themselves.

In some times and places, mafia groups have been able to create regional commissions to resolve these collective problems. The 'Ndrangheta of Sicily, for example created a council in response to a series of violent internal conflicts among its sub-clans. In the United States, a gang war in the 1930s led the so-called Five Families of New York to create a structure known as The Commission to deal with their internal problems. This replaced a prior system in which there was a boss of bosses who would resolve disputes among all others. The rents associated with this position led to continuous competition to capture it, a potentially destabilizing set of contests between strong and strong. The Commission, which included the five New York family heads, as well as Al Capone from Chicago and the head of the Buffalo Mafia, was put in place to resolve disputes and allocate territories. It held meetings every five years for a time, and is reputed to exist to this day.

Mafia dispute resolvers such as the Commission and the Council of the 'Ndrangheta do not publish written decisions. But anecdotal evidence suggests that their role is pre-

cisely as we describe: determining parties' relative strengths and pronouncing decisions based on them (Cattino 2019).

While inter-family violence still continues to exist, the establishment of a clear mechanism of arbitration seems to have resolved the collective action problems among the mafia families to some degree. And while we cannot observe the rents that come from being one of the arbitrators in this situation, no doubt they are plentiful.

4.2 The Epistemic Power of Arbitrators: International Relations

A similarly anarchic setting is international relations, on which there is a large literature discussing conflict and dispute resolution. Inter-state arbitration goes back at least to the ancient Greeks, who used the technique extensively (Ager 1997). It has been argued that it reflected Greek democratic norms, religious commitments, and ideology. Early arbitrations in this system were sometimes sent to the Oracle of Delphi, a religious figure supported by an inter-state council. The Oracle's resolutions were vague. In one case, in which two city-states disputed the ownership of a temple in between their territories, the Oracle gave possession to the first city to sacrifice at the temple, each starting out from their own territory at sunrise on the same day, chosen by the disputants. This resolution essentially turned on a contest of speed. Gradually, the practice of arbitration became more common, and a variety of arbitrators emerged, usually other cities acting through kings, assemblies or others. For example, in 315 B.C.E., a Spartan king was on his way to Italy when he was blown off course and made a landing at Apollonia in Northwestern Greece. He found that the city-state was under attack and acted as an arbitrator, resolving the dispute as a disinterested third party. The Greek practice was even used by parties outside the ancient Greek system, such as Persia and Rome.

One recent study has argued that this practice was not anarchic, but embedded in the hierarchical structure of interstate relations (Grynaviski and Hsieh 2015). By this, the authors mean that the disputants were usually part of a shared federation or subject to a hegemonic influence. Hierarchy, in their view, is a necessary condition for arbitration. We see the causality differently, namely that successful arbitration can facilitate hierarchy. But in any case, the findings of this study are consistent with our theory. The key point is that power on the part of the dispute resolver was often not necessary to resolve the disputes. Consider who was the party conducting the arbitration. If power was necessary, we would expect to find military leaders and kings being the dominant players. Instead, our own calculations show that in 50 of the 114 completed arbitrations identified by Ager, the arbitration was conducted not by kings but by experts who may have

brought epistemic power to bear.

In more modern settings of international relations, conflict is typically modeled as a bargaining game with private information on the level of resolve of the parties, which in turn is determined by nature. There has been a debate in recent years over the benefits of biased vs. unbiased mediators in resolving such conflicts. Kydd (2003) has argued that an impartial mediator may have less credibility than a mediator that is biased, because the impartial mediator has an interest in successfully resolving the dispute above all, and may communicate false information to do so. The biased mediator, by contrast, has the incentive to communicate accurate information to its favored party, which will find the communication credible. This literature only tangentially considers the incentives of unbiased mediators to engage in the process, and tends to do so informally. There are reputational and material benefits from being a successful mediator, but these tend to be assumed rather than modeled.

One solution to the puzzle of the unbiased but non-credible arbitrator is, of course, reputation. If an arbitrator obtains gains from arbitration, then in repeat play the arbitrator may develop a reputation for success and hence be motivated to provide credible information. A next step is to consider the transfer of epistemic power into material power. As the gains from such successes arise, it is possible that the arbitrator will be able themselves to become a strong type, and accumulate power. The combination of power and knowledge dominates either one operating alone.

We often observe powerful states acting as mediators in the international sphere. But there are also instances in which relatively powerless and disinterested actors are called on to mediate. The role of Norway in negotiating the Oslo Accords between Israel and Palestine, or mediating between the Tamil Tigers and the government of Sri Lanka are examples, as is an 1870s arbitration by Pope Leo XIII over ownership of the Caroline Islands between Germany and Spain. These disinterested observers did not get the role through bias, but instead through their epistemic abilities. And they were rewarded in reputation.

The 20th century was one in which a good deal of energy was devoted to institutionalizing arbitration. Beginning with the Hague Peace Conference of 1899, states imagined that a permanent system of arbitration (eventually embodied in the Permanent Court of International Arbitration and the International Court of Justice) would reduce global conflict. It was thought that peaceful dispute resolution would make war less attractive. These and many other international tribunals have had some success, typically in the context of low stakes problems that involve a coordination element. One of us has characterized this as Adjudication in Anarchy (Ginsburg and McAdams 2004): international

dispute resolution works best when it creates self-enforcing solutions in contexts in which the cost of fighting is high relative to settlement.

Yet the presence of arbitrators in the anarchic setting of international relations has not led to any propensity of states to stop investing in strength. In situations in which fighting has already erupted or the stakes of the dispute is high, states are prone to ignore arbitrators. Consider the recent dispute between China and the Philippines involving the South China Sea, a contested maritime territory that China believes is essential to its national interest. China has been building installations and expanding its presence, unilaterally claiming the "prize" in our parlance and signaling strength. The Philippines sought arbitration before a tribunal established under the UN Convention of the Law of the Sea (UNCLOS). A tribunal was constituted, but China refused to participate, and a decision found in favor of the Philippines. In this game, China invested in strength and claimed the prize; the Philippines sought arbitration, but China's continued willingness to fight suggests the arbitration will not be effective. And the presence of a dispute resolution option did not dissuade China from investing in strength. Our model helps to understand these facts.

4.3 An Extension: Building States

War, says Charles Tilly, makes states, in that the need for collective defense provides an incentive for a ruler to provide public goods in exchange for taxation. In our model, this idea would be translate into a strong type acquiring sufficient power that it can then dominate all others.

But might law be helpful in building states as well? We know from prior literature that power is useful for resolving disputes, while successful dispute resolution can confer legitimacy and produce rents, and so dispute resolution is complementary to state-building. It may be that judges can gain a reputation for accurate decisions, and so accumulate their own power. More likely, the judge can deal with enough cases that the King need only intervene in major ones. Judging is thus complementary to state-building, but in exercising full authority the judge is typically going to be outcompeted by specialists in violence.

As our question at the outset made clear, one reason to study arbitration in anarchy is that it offers a way to understand the development of a state and its legal system. Before a centralized authority exists, there would likely be competition among various potential arbitrators. Even assuming, as we do, that arbitrators have an advantage over parties in assessing the strengths of disputants, some arbitrators might be better than others at this

and at determining mutually acceptable settlements. We expect that successful arbitrators will tend to accrue power which can enable them to settle a wider range of disputes. This account is consistent with some models of state formation in which medieval princes gradually extended their writs, getting powerful lords to attend their courts and resolving their disputes without too much fighting. Not every prince was able to do this, however. This heterogeneity may have to do with facts about how power was distributed at some early period, but we think that it may also depend on the relative cognitive capacities of arbitrators. Our theory suggests that there is an epistemic as well as a military path to state-building, acknowledging that the two are complementary.

A paradigmatic example comes from Herodotus, who reports the famous story of Deioces who became king of the Medes. He leveraged his epistemic capacities as an able judge into the kingship (which he held for 53 years and spawned a dynasty):

"As the Medes at that time dwelt in scattered villages without any central authority, and lawlessness in consequence prevailed throughout the land, Deioces, who was already a man of mark in his own village, applied himself with greater zeal and earnestness than ever before to the practice of justice among his fellows. It was his conviction that justice and injustice are engaged in perpetual war with one another. He therefore began his course of conduct, and presently the men of his village, observing his integrity, chose him to be the arbiter of all their disputes.... he showed himself an honest and an upright judge, and by these means gained such credit with his fellow-citizens as to attract the attention of those who lived in the surrounding villages. They had long been suffering from unjust and oppressive judgments; so that, when they heard of the singular uprightness of Deioces, and of the equity of his decisions, they joyfully had recourse to him in the various quarrels and suits that arose, until at last they came to put confidence in no one else.... The number of complaints brought before him continually increasing, as people learnt more and more the fairness of his judgments, Deioces, feeling himself now all important, announced that he did not intend any longer to hear causes, and appeared no more in the seat in which he had been accustomed to sit and administer justice. ... Hereupon robbery and lawlessness broke out afresh, and prevailed through the country even more than heretofore; wherefore the Medes assembled from all quarters, and held a consultation on the state of affairs. ... 'we cannot possibly,' they said, 'go on living in this country if things continue as they now are; let us therefore set a king over us'.... It followed to determine who should be chosen to the office. When this debate began the claims of Deioces and his

praises were at once in every mouth; so that presently all agreed that he should be king.... Deioces continued to administer justice with the same strictness as before. Causes were stated in writing, and sent in to the king, who passed his judgment upon the contents, and transmitted his decisions to the parties concerned: besides which he had spies and eavesdroppers in all parts of his dominions, and if he heard of any act of oppression, he sent for the guilty party, and awarded him the punishment meet for his offence. Thus Deioces collected the Medes into a nation, and ruled over them alone.”

If knowledge is not inevitably power, it may nevertheless be so eventually, and it can always supplement raw power as a means to develop authority. The history of the common law provides further evidence for this dynamic. In feudal pre-Norman England, private violence was endemic, and took the form of feuds in which violent self-help was the primary mechanism of dispute resolution among notables. With the Norman Conquest, continental ideas about law were introduced, but their extension over the English countryside was gradual and quite incomplete. Well into the thirteenth century, the Kings Peace had not displaced local violence as a mechanism of righting wrongs; indeed, the survival of dueling into the early history of the United States suggests that this pattern was more general. The extension of the Kings justice into the countryside was gradual, and tended to build on local norms and patterns of what counted as a wrong. In other words, the incentives of the powerful to submit to arbitration were not sufficient to allow unification of the legal system. Gradually, in an oft-told tale, the Kings judges extended their writs throughout the land, but the law remained the province of these specialized interpreters, the judges. They would collect the ad-hoc, locally informed decisions made around the country, and eventually recorded them. Once the doctrine of *stare decisis* emerged, local law could be unified in favor of the common law. But only with the gradual increase in legislation in the 18th and 19th centuries was the Kings direct command effectively enshrined in law. (And of course there was resistance from judges who thought their own law more permanent and stable than seemingly arbitrary legislation.) In short, the English pattern reflects the variable quality of arbitration and its gradual centralization in service of state-building. It is a story of the complementarity of law and power in the process.

For contemporary illustrations of the importance of dispute resolution for accumulating power, we need look no further than rebel groups, like the Tamil Tigers or ISIS, which set up courts as one of their early moves. These groups seem to rely on power, and of course want to extract resources from subject populations. As documented by Mampilly (2011) the Tamil Tigers established land courts for the purposes of assessing property val-

ues, facilitating tax authority. Initially they used the government and colonial-era codes for the substantive rules, and so were primarily focused on the way law could facilitate finance, but eventually they took steps to incorporate Tamil cultural norms for inheritance and other aspects of family law, and to excise norms that would not fit their ideology, such as caste. In 1992 they established a broader set of courts, and in 1994, the Tigers adopted their own penal code. Tamil Tiger use of law was so extensive that they established a law college to train 300 lawyers a year (Sivakumaran 2009). All this was remarkably successful in producing the public good of social order: crime was low, and tens of thousands of civil disputes were resolved in Tiger-held territory. This example shows the complementarity of dispute resolution and power in controlling territory.

Perhaps the best contemporary example of government emerging through arbitration is the Islamic Courts Union in Somalia. After the collapse of the Somali government in 1991, Sharia courts stepped in to fill a void in the absence of state power. Only later did they develop enforcement capacity and organized into a broader political movement (Ab-bink 2009; Barnes and Hassan 2007). Initially, funded by court fees, they simply resolved disputes, but eventually set up police forces paid by local businesses. Eventually, they banded together into a Union in the mid-1990s, culminating in 1999 with the creation of a militia called the Islamic Courts Union. Eventually they took over Mogadishu and were able to obtain outside funds from neighboring Eritrea. This is an example of arbitrator power motivating state-formation.

However, rule by judges seems to be a feasible but unstable form. Consider the biblical narrative. In the book of Samuel, the aging prophet turns over judicial power to his two sons, but they turn out to be ineffective and corrupt. The leaders of the people then approach Samuel, and demand a king to replace the judges that had ruled the country for centuries. Samuel warns the leaders that a monarchy will be tyrannical, and that once they have one they will not be able to gain relief:

This is what the king who will reign over you will claim as his rights: He will take your sons and make them serve with his chariots and horses, and they will run in front of his chariots. Some he will assign to be commanders of thousands and commanders of fifties, and others to plow his ground and reap his harvest, and still others to make weapons of war and equipment for his chariots. He will take your daughters to be perfumers and cooks and bakers. He will take the best of your fields and vineyards and olive groves and give them to his attendants. He will take a tenth of your grain and of your vintage and give it to his officials and attendants. Your male and female servants and the best of your cattle and donkeys he will take for his own use. He will take

a tenth of your flocks, and you yourselves will become his slaves. When that day comes, you will cry out for relief from the king you have chosen, but the LORD will not answer you in that day. (1 Samuel 8: 11-18)

Faced with this grim warning, the people choose to ignore the Prophet, and they persist in demanding a strong leader. Frustrated, Samuel asks God for advice and the almighty advises him to give the people what they want. Shortly thereafter, a man named Saul appears, and Samuel anoints him as the first King. The rest, as they say, is history. Things go fairly well for a while, and Sauls successors David and Solomon enjoy a period of military expansion and economic stability. Thereafter, however, there is a political split into two kingdoms, and the remaining three dozen or so Kings of the Jewish people are mostly mediocrities who are not household names today. The monarchy ends with the Babylonian invasion and sacking of Jerusalem in the 6th century B.C.E.

Why is rule by judges an unstable form? In the progression of governance forms from priest to judge to king, the endpoints are fairly understandable. Priests rule because they rely on internalized fear of divine punishment, and are rewarded with tithes. Kings rule because they rely on fear of temporal punishment by the sword, and are able to use this power to obtain taxes. How can judges govern? They have neither the power of the divine nor the power of the sword. But they can, and apparently in ancient Israel did for several centuries, leverage dispute resolution into temporal power, producing a system that was stable so long as external threats are not too high. That condition is rarely satisfied under modern circumstances. Perhaps the problem is that, as in our model, rule by judges motivates investment in strength, leading to more conflict than can be handled by arbitration alone.

5 Conclusion

A complete theory of arbitration in anarchy must identify why it is that people would invest in strength in the first place, why they would bring cases to judges (ex ante compatibility), as well comply with them (ex ante compatibility), and why judges are motivated to give accurate solutions. Only with an account that integrates these problems can we show how judges can be effective without external power. In this article, we have integrated these various issues into a single framework and demonstrate the possibility of arbitration in anarchy. The framework helps illuminate certain aspects of historical examples, extending the prior literature. We further explain at least one channel for leveraging judicial decision-making into state power. This route appears to be relatively rare, but not

unknown in human history.

6 References

- Abbink, J. 2009. The Islamic Courts Union: The Ebb And Flow Of A Somali Islamist Movement. In *Movers and Shakers*, pp. 87-113. Leiden: Brill.
- Ager, Sheila. 1997. *Interstate Arbitrations in the Greek World 337-90 BC*. Berkeley: University of California Press.
- Ayboga, E. 2016., et al. *Revolution in Rojava*. Trans. Janet Biehl. London: Pluto Press.
- Baczko A, Dorronsoro G. and Quesnay, A. 2017. *Civil War in Syria: Mobilization and Competing Social Orders*, New York
- Barnes, C, Hassan, H. 2007. The rise and fall of Mogadishu's Islamic Courts. *Journal of Eastern African Studies* 1: 151-160.
- Benson, Bruce L. 1989a. Enforcement of Private Property Rights in Primitive Societies: Law without Government. *Journal of Libertarian Studies* 9:126.
- Catino, Maurizio. 2014. How Do Mafias Organize? Conflict and Violence in Three Mafia Organizations. *European Journal of Sociology* 55(2): 177-220
- Catino, Maurizio. 2019. *Mafia Organizations: The Visible Hand of Criminal Enterprise*. New York: Cambridge University Press.
- Fuller, Lon. 1964. *The Morality of Law*. Cambridge: Harvard University Press.
- Gambetta, Diego. 1993. *The Sicilian Mafia: The Business of Private Protection*. Cambridge: Harvard University Press.
- Ginsburg, Tom Richard McAdams. 2004. *Adjudication in Anarchy*. William and Mary Law Review.
- Grynaviski, Erik and Amy Hsieh. 2015. Hierarchy and Judicial Institutions: Arbitration and Ideology in the Ancient World. *International Organization* 69: 697-729.
- Hadfield, Gillian and Barry Weingast. 2012. What is law? A coordination model of the characteristics of legal order. *J. Leg. Anal.* 4(2):471-514
- Hadfield, Gillian and Barry Weingast. 2013. Law without the state: legal attributes and the coordination of decentralized collective punishment. *J. Law Courts* 1(1):334
- Kydd, Andrew. 2003. Which Side are you On? Bias, Credibility and Mediation. *Am. J. Pol. Science.* 47: 597-611
- Mampilly, Zachary C. 2011. *Rebel Rulers: Insurgent Governance and Civilian Life During War*. Ithaca, NY: Cornell Univ. Press
- McAdams, Richard. 2015. *The Expressive Powers of Law*. Cambridge: Harvard University Press.
- Milhaupt, Curtis and Mark West. 2000. *The Dark Side of Private Ordering: An Institutional and Empirical Analysis of Organized Crime*. University of Chicago Law Review

67(1): 41-103

Sivakumaran S. 2009. Courts of armed opposition groups: Fair trials or summary justice? J. Int. Crim. Justice 7(2):489-513

Appendix - Proofs

Proof of Proposition 1

Supposing everyone else is playing as described, we find conditions under which the given strategy is a best response. If everyone else is playing the given strategy, we first find best responses for a weak player when they are faced with the choice to fight or not. The choice to not fight would deliver a sure payoff of zero. The choice to fight will deliver a payoff of $-qc + (1 - q) \left(r \left(\frac{b}{2} - c \right) + (1 - r)b \right)$, as a function of mixing parameters. Thus, mixing with probability r between fighting and not fighting will be a best response if and only if

$$-qc_h + (1 - q) \left(r \left(\frac{b}{2} - c_h \right) + (1 - r)b \right) = 0. \quad (6.1)$$

Now, given that in the second stage a weak player earns a payoff of zero by fighting, it's clear that a strong player would earn a strictly positive payoff by fighting, and hence fighting is a dominant strategy for a strong player.

It remains to check that becoming strong with some probability is itself optimal. The expected payoff to becoming strong and then fighting, the left side of the below equation, must equal the payoff to staying weak and earning zero in order for mixing to be a best response. That is, mixing between strong and weak with probability q is optimal if and only if

$$q \left(\frac{b}{2} - c_l \right) + (1 - q)(r(b - c_l) + (1 - r)b) - x = 0. \quad (6.2)$$

Thus, in order for the equilibrium as described to exist, we require the mixing probabilities q and r to simultaneously solve (6.1) and (6.2). Simple algebra reveals that the unique solution are the given equilibrium probabilities,

$$q_{NA}^* = \frac{2(bB - Bx - 2c_l x)}{bB}, \quad r_{NA}^* = \frac{2b(b + 2c_h - 2c_l) - 4x(b + c_h)}{b(b + 2c_h - 2c_l) - 2x(b + 2c_h)} \quad (6.3)$$

where $B = b + 2c_h - 2c_l$.

Some restrictions on the parameter values are required for coherency. In particular, the requirement that $q_{NA}^* > 0$ is satisfied with the assumptions $b/2 > c$ which has already been made. Next we need $r_{NA}^* > 0$. This becomes an upper bound on q_{NA}^* . In particular we need $q_{NA}^* < 2x/B$. Naturally, in order for any weak type to fight, there can't be too many strong types. And so we don't have to check $q_{NA}^* < 1$ as $q_{NA}^* < 2x/B$ suffices. The exact inequality here becomes $x > \frac{bB}{2b + 2c_h}$. This, also naturally says, that in order that not too many become strong, the cost needs to be quite high. So making this assumption, we

get both $r_{NA}^* > 0$ and ${}_{NA}^*q < 1$. Lastly we get $r_{NA}^* < 1$ for free from our earlier argument that we need $x < b/2$.

Thus we make the following assumptions on the parameters:

$$x < \frac{B}{2}, \quad x > \frac{bB}{2c_h + 2b}, \quad x > \frac{B}{2} - c_l. \quad (6.4)$$

The analysis above demonstrates that subject to assumptions (6.4) an equilibrium exists described by (6.3), as desired.

Proof of Proposition 2

We require the sensibility assumptions (6.4) that defined the equilibrium, but given these the comparative static results are trivial.

Proof of Proposition 3

Note first by Bayes rule that if someone chooses A in stage 3 yet ends up in stage 4 they conclude their opponent played NA, and as such their posterior about the probability their opponent is strong is

$$Q = \frac{q(1-w)}{q(1-w) + (1-q)(1-y)}. \text{ And note: } Q > q \iff w < y \text{ and } \frac{1-y}{w-y} < q \leq 1. \quad (6.5)$$

Now, we begin at the end. At stage 4, for a weak type who played NA in stage 3 their belief is still their prior q . For them to mix over F and NF, since the latter delivers a zero payoff, we require that

$$W(NA, q) \equiv -qc_h + (1-q) \left(r(1-y) \left(\frac{b}{2} - c_h \right) + b(1-r)(1-y) + by \right) = 0 \quad (6.6)$$

which says they can either play against a strong type who will always fight, or a weak type who arbitrated before and thus won't fight now, or a weak type who also played NA before and thus mixes now. We denote this value to the game at stage 4 to a weak type who plays NA in stage 3 and thus has belief q as $W(NA, q)$ which is zero in this equilibrium. Similarly we will soon need $W(A, Q)$ defined analogously. But note that $W(A, Q) = 0$ since given the assumptions above if a weak type is indifferent between

fighting and not in stage 4 with belief q they will strictly prefer not fighting with belief Q and hence payoff is zero.

Next in stage 3 we require indifference for the weak types between A and NA:

$$\begin{aligned} W_{arb} &= q(-k(w) + (1-w)w(A, Q)) + (1-q)(y(b/2 - k) + (1-y)W(A, Q)) \\ &= -qk(w) + (1-q)y(b/2 - k) \text{ using } W(A, Q) = 0 \\ &= W_{noarb} = W(NA, q) = 0. \end{aligned}$$

Hence, using the second line, we can write the indifference condition as

$$y = \frac{kqw}{(1-q)\left(\frac{b}{2} - k\right)}. \quad (6.7)$$

Next define the value of the game to the strong type in stage 4 if they arrive their having the belief of either q or Q (ie whether they played NA or A in stage 3 respectively):

$$S(\kappa) = \kappa \left(\frac{b}{2} - c_l \right) + (1-\kappa)(r(b - c_l) + (1-r)b).$$

And so the indifference condition for strong types at stage 3 is that (LHS is playing A, RHS is playing NA)

$$q \left(w \left(\frac{b}{2} - k \right) + (1-w)S(Q) \right) + (1-q)(y(b - k) + (1-y)S(Q)) = S(q). \quad (6.8)$$

Finally, we need indifference at stage 1, and since we have seen that weak types eventually get zero, and strong types eventually $S(q)$ this simply reads

$$S(q) - x = 0. \quad (6.9)$$

So the equilibrium parameters (q, y, w, r) are pinned down by equations 6.6, 6.7, 6.8, 6.9. Solving for these we get the following:

$$w^* = \frac{b(b-2k)(2x(c_l(b+2c_h) - k(b+2c_h-2c_l)) - (b+2c_h-2c_l)(b(c_l-k) + 2c_lk))}{2(b+2c_h)(bc_l - k(b+2c_l))(b(c_l+k) - 2k(c_l+x))},$$

$$q^* = \frac{b(c_l+k) - 2k(c_l+x)}{bc_l}$$

$$r^* = \frac{1}{2}b \left(\frac{1}{c_l} - \frac{1}{k} \right) + 1$$

$$y^* = \frac{b(b + 2c_h - 2c_l)(b(c_l - k) + 2c_l k) - 2bx(c_l(b + 2c_h) - k(b + 2c_h - 2c_l))}{(b + 2c_h)(bc_l - k(b + 2c_l))(b - 2(c_l + x))}$$

It remains to constrain the parameters so that the equilibrium mixing probabilities are indeed probabilities.

First, in order that $0 < w^* < 1$ we require assumptions that effectively say: we need x , the cost of getting strong, large enough so that there are sufficiently few strong types to discourage all arbitration, but x and c_l small enough so that there are enough strong types to discourage some fighting. Specifically, the assumptions become:

$$w^* > 0 \iff x > c_l \left(\frac{2c_h c_l}{c_l(b + 2c_h) - k(b + 2c_h - 2c_l)} - 1 \right) + \frac{b}{2} \quad (6.10)$$

and.

$$w^* < 1 \iff c_l < \frac{bk}{b + 2k} \text{ and } x \leq \frac{b^3 c_l - b^3 k + 2b^2 c_l^2 - 8bc_l^2 k - 4bc_l k^2 + 8c_l^2 k^2}{2b^2 c_l - 2b^2 k - 8c_l k^2}. \quad (6.11)$$

Now, to get $0 < 1q^* < 1$ we require that the prize being fought for is high enough that some people want to become strong, but not high enough that everyone does, Specifically, we require:

$$1 > q^* > 0 \iff 2(c_l + x) > b > \frac{2k(c_l + x)}{c_l + k} \quad (6.12)$$

Next, to get $1 > r^* > 0$ we just need $c_l > k$ (already assumed) and

$$b < \frac{2c_l k}{c_l - k}. \quad (6.13)$$

This second condition is a bit counter-intuitive, because it says the condition for some weak types to fight is that the prize should be low. But really the prize being low is ensuring q isn't too big, which is what allows some weak types to want to fight.

Finally, to ensure that some weak types want to arbitrate we need them to have some chance of winning, which means we must constrain the number of strong types, but to make sure they don't all want to, we need their to be enough strong types. That is we

need the cost of becoming strong to be high but not too high:

$$1 > y^* > 0 \iff \frac{b^2k + bc_hc_l + bc_hk - 2bc_lk - 2c_hc_lk}{2k(b + c_h)} > x > \frac{(b + 2c_h - 2c_l)(bc_l - bk + 2c_lk)}{2(bc_l - bk + 2c_hc_l - 2c_hk + 2c_lk)}. \quad (6.14)$$

Lastly, we need to ensure that $y^* > w^*$ so that the actions taken in this equilibrium are indeed sequentially rational. Recall that weak types that play 'arbitrate' but are rebuffed update their belief about q upward if and only if $y^* > w^*$. Specifically, we must assume that

$$k < \frac{bc_l}{2c_l + 2x - b} \text{ and } c_h > \frac{-b^2c_l + b^2k + 2bc_l^2 - 4bc_lk + 2bc_lx - 2bkx + 4c_l^2k + 4c_lkx}{2bc_l - 2bk + 4c_lk - 4c_lx + 4kx} \quad (6.15)$$

These say that in order for weak types to arbitrate a lot, we need the cost of arbitration to be low, and their cost of fighting to be high (relative to c_l).

Thus, assumptions (6.10) - (6.15) and the analysis above ensure that the equilibrium of the given form exists and is unique.

Proof of Proposition 4

Routine algebra reveals that $q_{NA} < q_A$ is true if and only if

$$x > \frac{(b + 2c_h - 2c_l)(bc_l - bk + 2c_lk)}{2(bc_l - bk + 2c_hc_l - 2c_hk + 2c_lk)}$$

This is precisely assumption 6.14, and so given the set of assumptions underlying proposition 3, this proposition trivially follows.

Proof of Proposition 5

Similarly, it is straightforward to see that $r_{NA} > r_A$ if and only if

$$x > \frac{(b + 2c_h - 2c_l)(bc_l - bk + 2c_lk)}{2(bc_l - bk + 2c_hc_l - 2c_hk + 2c_lk)}$$

and so given the set of assumptions underlying proposition 3, this proposition trivially follows.