**Acknowledgements**: This paper is indebted to Jerry Gaus for numerous helpful conversations and comments on earlier drafts. My thanks also to participants at the panel on “The Effects of Disagreement About Justice” at the 2018 PPE Society Annual Meeting, and to two anonymous reviewers at this journal.

Introduction

A condensed history of the contemporary debate about distributive justice starts with Rawls’s difference principle (1999, p. 53), then Nozick’s replies and especially the Wilt Chamberlin argument (1974, pp. 160-3), followed by Cohen’s rejoinders (1995). The debate became increasingly fine grained as it proceeded, with disputes ranging over the correct principle(s) of distribution (Anderson 1999; Arneson 2000; Parfit 1997; Frankfurt 1987; Schmidtz 2002), what is to be distributed (Rawls 2001, pp. 168-75; Sen 1982), whether these principles apply only nationally or globally (Nagel 2005; Pogge 1992), and so on.

The competing proposals seem to have reached a stalemate, with no signs of progress. In order to sidestep this impasse, I address a different question of distribution here, asking “how should people who care, and disagree, about distributive justice go about resolving that disagreement, given the effect of the distribution of goods on their lives?” I suppose that participants in these debates all have settled convictions regarding the various abstract distributive questions; my concern is how they might adjudicate their concrete conflicting justice-influenced claims and commitments. Focusing on adjudicative mechanisms allows us to consider whether any progress can be made with respect to deep normative disagreements about distributive justice. Individuals making competing claims of justice will hardly be content waiting until philosophers find some definitive resolution to the disagreement. Some decision must be made in the absence of consensus, and adjudicative mechanisms are means for making such decisions.[[1]](#footnote-1)

Unlike abstract philosophical analyses, an adjudicative mechanism must be sensitive to the results it produces when implemented in actual social circumstances. An adjudicative mechanism that all parties agreed was inadequate, or failed to produce generally acceptable resolutions, would itself be inadequate, even if in principle it appeared to be optimal. Thus, the question of distributive disagreement must be sensitive to realization conditions of proposed methods of adjudication.[[2]](#footnote-2)

To investigate the results of some adjudicative mechanisms, I develop an agent-based model in which members of a political community must select some distribution of goods. An agent-based model allows one to represent with greater detail individual behavior and resultant group level dynamics, in this case providing greater insight into the process of, outcomes of, and responses to adjudication. I begin by discussing adjudicative mechanisms, establishing two evaluative criteria: a criterion of fairness, and a criterion of pragmatic outcomes (§ 1). I then turn to the set-up of the model, where I develop an account of the behavior of agents, the process of adjudication, and the basic parameters of the model (§ 2). With the model set up, I discuss its results, addressing theoretically significant results as they arise (§§ 3-6).

1. Adjudication

The impasse surrounding the distributive justice debate gives us reason to use adjudicative mechanisms to resolve social conflict when it is based on competing claims of distributive justice. When attempts to convince others fail and some shared resolution is necessary in order to continue social cooperation or coordination, continued argumentation will be of no use. Insistence that one’s claim of distributive justice is the true claim will do little to resolve the conflict, it will only perpetuate it. A promising means of resolving such disagreement is by appeal to an adjudicative mechanism. An adjudicative mechanism, generally speaking, is some process by which an authoritative resolution is provided. Adjudication bypasses entrenched disagreement, since it makes no claims as to the truth of the resolution, only claiming the authority to enforce it.

Adjudication is thus in part a functional concept: an adequate mechanism of adjudication is one that provides resolutions the terms of which agents will obey. And agents will not obey just any result. The general construal of adjudication, which places no constraints on the process of adjudication, allows for just about any means of choosing an outcome, many of which are likely to strain the compliance of agents. For example, agents with competing claims could resolve their conflict by drawing out the complete logical space of outcomes, and then throw a dart to choose the resolution. Clearly, better methods are available.

Agents raise and support their competing claims because they have some reason to do so, and these reasons do not evaporate at the threshold of adjudication. Whatever reasons agents have for supporting their claims will be relevant in their decisions to proceed with and obey the results of adjudication. This leads to a general condition of responsiveness that adjudicative mechanisms must meet—agents will seek methods of adjudication that give their particular claims some minimum weight in selecting a resolution.

Of course, the most satisfactory adjudicative mechanism for some agent is simply the mechanism that is responsive solely to that agent’s particular claim, selecting exactly what she wants. But such a mechanism will hardly resolve the dispute, since other parties will have their own reasons to reject such a mechanism, since it did not give theirreasons sufficient weight. Each agent is thus led to acknowledge the reasons of the other parties, since an effective resolution of the dispute will require that most, if not all, parties obey the terms of the resolution, otherwise the dispute continues.

An appealing criterion for adjudicative mechanisms is thus a *condition of fairness*: the weight of each agent’s claim is to be treated equally in the process of adjudication. If the claims of agents have equal weight, then agents have no basis for rejecting adjudication at the outset—such rejection would require that some agent claim that she deserves greater weight in the process of adjudication, a claim which, if satisfied, would make the process of adjudication unacceptable to the other parties, since they will have reason to reject participating in the process on account being treated unfairly.

Fairness alone, however, will not ensure that adjudication is functional. A fair system may still effectively guarantee that some agents always lose. Take, for example, a simple plurality vote, where the most popular option is chosen, and the vote of each individual is counted equally. Agents who are in a persisting minority have no chance of their claim ever winning. If the agents in a minority position knew this, committing themselves to obeying the outcomes of a plurality vote would require them to contradict their own reasons, since antecedent commitment to the authority of the adjudicative mechanism is in this case effectively a concession.

To be functional, in addition to being fair, a process of adjudication must be able to produce outcomes that agents know can be acceptable to them. The process of adjudication must not be self-effacing, it cannot require that the agent set aside their claims and reasons and accept the result of adjudication, come what may (Gaus 2015; Lloyd 2003). This leads us to the *pragmatic outcome criterion* of adjudication: the process of adjudication must be such that the parties generally find outcomes to be acceptable. Below, I discuss in greater detail what it means for an outcome to be acceptable to an agent, and a measure of the pragmatic outcome criterion is provided. For now, suffice it to say that agents will find an outcome acceptable if it is sufficiently similar to their initial claim, and that a given adjudicative process better satisfies the pragmatic outcome criterion than some alternative insofar as the outcomes it produces tend to be acceptable to a greater number of parties.

There are thus two criteria for evaluating adjudicative mechanisms: the criterion of fairness that applies to the procedure, and the pragmatic outcome criterion of satisfying agent claims which applies to the results of the adjudicative procedure. An adjudicative mechanism that meets both criteria will likely be regarded as binding by parties to the dispute, making it an effective means of resolving the dispute. If agents believe that a process is unfair to them, they will have grounds for antecedently rejecting the outcome as binding, since if their claim is given less weight they are less likely to get an outcome they find acceptable. If agents believe that an adjudicative process will generally produce outcomes that they find unacceptable, then even if the process is fair they will have reason to reject participation in such a process at the outset. The fairness and pragmatic outcome criteria make no appeal to substantive normative claims, as they are criteria for evaluating a political process rather than distributions, avoiding the impasse on normative questions that prompts the need for adjudication in the first place (Van Schoelandt and Gaus 2018).

The focus of this paper will be the pragmatic outcome criterion. I will consider how well four adjudicative mechanisms, all of which meet the fairness criterion, satisfy the pragmatic outcome criterion. My focus is the pragmatic outcome criterion on account of the relative complexity of measuring it. Determining whether a process is fair is straightforward; if the claims of agents are weighed equally, then it is fair. To measure the degree to which an adjudicative mechanism satisfies the pragmatic outcome criterion, by contrast, we must know what proportion of agents in a diverse population find the outcomes acceptable. And doing this requires keeping track of a large array of unique agent parameters, such as the evaluative standards of agents, and the range of outcomes that they find acceptable. To accomplish this task, I develop an agent-based model where diverse agents enter an iterated adjudicative process. By using a computational model, we can keep track of all the relevant parameters for a relatively large population, allowing us to examine the extent to which different adjudicative mechanisms satisfy the pragmatic outcome criterion.

2. Dynamics of Adjudication

The model of adjudication that I develop here proceeds in two steps: some population of agents use an adjudicative mechanism to choose a new distribution, and then each agent responds to the result, based on whether she finds it acceptable or not.

In order to keep matters simple, I will assume that agents are only interested in outcomes of adjudication in two ways: i) how the outcome measures up to their conception of justice, and ii) how the outcome benefits them directly by increasing the amount of goods the agent has, which I will refer to as “self-interest” for shorthand.[[3]](#footnote-3)

Agents begin by strictly preferring justice (as they conceive of it) to self-interest. This means that an agent will aim to promote the distribution she believes to be most just regardless of the consequences it has for the number of goods she receives. Such an assumption aims to capture that situation in which there is a genuine disagreement about justice, rather than a mere conflict of self-interest. As the process of adjudication continues, if an agent’s preferred conception of justice fails to be realized, she will begin to trade-off between justice and self-interest, accepting marginal losses in justice for gains in the amount of goods she receives. If one’s attempts to realize justice are systemically frustrated by others, they will no longer be inclined to act justly themselves. Such a response would occur, if, for example, acting in accordance with justice was a moral norm that individuals would only follow on the condition that others follow it as well (Bicchieri 2005, pp. 117-30; Bicchieri and Chavez 2010; Bicchieri and Mercier 2013). Or, agents may simply be responding in accordance with a norm of reciprocity: “if others are not acting justly, then why should I?” (Rawls 1999, pp. 238-9; Rawls 2005, p. 17; Weithman 2011, p. 49). Supposing that compliance with justice is a moral norm, or that individuals abide by justice on reciprocal terms, then when an agent sees others acting on terms determinately opposed to (what she believes is) justice, she herself will cease to act in accordance with her conception of justice. And, given that self-interest is the only alternative available, this means that she will begin pursuing personal gains in goods at the expense of justice.

This account of preferences departs from typical accounts offered in rational choice theory.[[4]](#footnote-4) While some descriptive accounts of rational choice have discussed how preferences can shift by virtue of natural selection over time (Heifetz, Shannon, and Spiegel 2007; Robson and Samuelson 2011), they have not addressed preferences shifting by virtue of their disappointment within single individuals. That said, preference changes in light of disappointment are not an unfamiliar phenomenon in behavioral treatments of preferences. Generally speaking, the phenomenon of preference change and preference reversal has been observed under a variety of conditions (Tversky and Thaler 1990). Change in preferences in response to disappointment, a case of “sour grapes,” has been discussed at length by Jon Elster, who argues that this phenomenon is a response to cognitive dissonance arising from dissatisfaction (Elster 2016, ch. 3). Whatever the mechanism for such preference shifts, be it a response to cognitive dissonance, a response to perceived norm violation, or following a norm of reciprocity, inclusion of such behavior in the model is meant to capture the intuitive idea that failure to realize what one believes is justice will lead that individual to change her behavior. The context of interest here is a case in which individuals sincerely motivated by claims of justice desire conflicting social states, and for individuals so motivated, I take it, failure to realize aims that they hold to be so important will have some distinctive effect on their behavior.

*2.1 A Simple Case of Adjudication*

The following case serves as a simple illustration of the process of adjudication. Suppose Catherine, Dominic, and Eve are deciding on how to distribute goods in their society. Catherine and Dominic are liberals, and so favor a distribution that is a result of market interaction with some ex post redistribution for welfare purposes. Eve, on the other hand, is a staunch utilitarian, and so prefers that distribution which maximizes aggregate welfare in society. Assume that the parties have identified the two distributions and have named the liberal distribution preferred by Catherine and Dominic “D1” and the utilitarian distribution preferred by Eve “D2.” Suppose further that their preferred means of adjudication is a simple plurality vote. Following the vote, Catherine and Dominic win, implementing distribution D1.[[5]](#footnote-5) Eve sees the new distribution and seeing that a determinately unjust distribution has been implemented (after all, the state could be used to increase aggregate welfare to a significantly higher degree), becomes willing to concede some gains in justice for some gains in self-interest, realizing perhaps that she will not be able to get optimal justice, opting for a less demanding option that benefits her self-interest to a greater degree. She now advocates for distribution D3, which is still quite close to her utilitarian ideal, but now provides her with some marginal advantage in terms of goods compared to her ideal distribution D2, at the cost of producing marginally less aggregate utility.

The process repeats, and again Eve loses the vote. Again, Eve changes her preferences, seeing that even her weakened notion of justice could not be realized, she becomes marginally more self-interested. The process continues, until one of two conclusions are reached. Eve might trade-off between justice and self-interest a sufficient number of times that she happens to begin advocating for the liberal distribution as well, upholding the distribution partially on grounds of justice and partially on self-interest. Though the liberal distribution D1 is some distance away from the utilitarian distribution she believes most just, in this case it might be close enough that given the increase in goods she receives, she is willing to support it, but only after seeing her attempts at realizing what she regards as comparatively more just distributions fail. Or, Eve never finds the liberal distribution acceptable, which leads to her eventually becoming purely egoistic, interested only in maximizing her self-interest. This case obtains when the liberal distribution provides a lower benefit to Eve’s self-interest than some other distribution. The point is best illustrated by considering Eve’s justice perspective, or how she evaluates and orders the similarity of distributions according to how just they are.

**Fig. 1** Eve’s justice perspective when she reaches a consensus with the liberals after the third round

Figure 1 shows Eve’s justice perspective. Since I have assumed that agents begin only concerned with justice, the x-axis is ordered such that the justice score, the left-hand y-axis, is single peaked.[[6]](#footnote-6) By ensuring that the x-axis is ordered this way, we capture the fact that agents consider outcomes primarily in terms of their justice, and so order them according to similarity on the basis of justice. The right-hand y-axis shows the self-interest score of distributions, the second metric along which agents evaluate distributions.[[7]](#footnote-7) According to the process described above, Eve began by advocating for her most preferred distribution, D2. After disappointment in the first round, Eve considers the nearby distributions, D4 and D3, and chooses to advocate for D3 in the following round, since it benefits her self-interest the most within the range she is considering. The second round then takes place and having lost again Eve experiences a shift in preferences once more. If Eve has the perspective shown in Figure 1, then she will come to agree with Catherine and Dominic, since the liberal distribution D1 benefits her most among the distributions under consideration. However, if Eve had the perspective shown in Figure 2, then she would have come to advocate for D5. Suppose also that in this latter case distributions D1-5 are exhaustive of the possibilities. Eve will have become completely egoistic in this second case.

**Fig. 2** Eve’s justice perspective when she becomes an egoist in the final round

The discussion so far has supposed that Eve has no tolerance, the only distribution that she will find acceptable is her most preferred distribution at the time. Such an optimizing stance on justice is unlikely to be found in all members of a given community, as there should be some variance in the range of distributions that agents find permissible (Gaus 2016, 215-20). Some members of the population are likely to conditionally endorse certain distributions as sufficiently just if they believe others support the distribution as well (Gaus 2018). To represent this broader scope of acceptable distributions, agents will have a tolerance radius, setting a range along the x-axis centered on an agent’s most preferred distribution, with distributions that fall within that range being considered acceptable by the agent.

If a distribution is selected that falls within the agent’s tolerance range, then they would have no reason to trade-off on their self-interest, since the distribution is acceptable (albeit suboptimal) on their currently held evaluative standard. If Eve had a tolerance range of two at the outset, she would have experienced no shift in her preferences, since the liberal distribution D1 falls within that tolerance range.

The simple case of Catherine, Dominic, and Eve provides an illustration of the dynamics of the model. Agents will enter with a unique justice perspective, advocating for different distributions to be implemented. They will attempt to resolve the disagreement by using some adjudicative mechanism. Given the results of adjudication, agents update their preferences by extending their search radius of distributions they will consider and shift their support to that distribution which maximizes the goods they receive within that radius. The process is then repeated with the agents’ new preferences.

Two more core characteristics of the model warrant explanation: the dependent variable by which we will measure how well different adjudicative processes meet the pragmatic outcome criterion, and the independent variable, which consists in the different adjudicative mechanisms tested.[[8]](#footnote-8)

*2.2 Outcomes of Adjudication*

Mechanisms meet the pragmatic outcome criterion when they produce results that are generally acceptable to the disputing parties, and they meet the criterion to a greater degree when more agents are able to accept the outcome. Acceptance will be represented by the number of agents who are not maximally self-interested at the conclusion of adjudication, since these are the agents that have found some compromise between pursuing reasonably attainable (from their perspective) justice and their self-interest, whereas maximally self-interested agents have grown completely disaffected from the pursuit of justice.

When agents become completely self-interested, they are no longer able to conceive of their political community as one interested in justice, becoming disaffected from the public sphere. Rawls, discussing peoples’ inability to affirm the public conception of justice, writes:

[t]he meaning of "affirm" here can be given by noting two ways in which we react when the strains of commitment seem to us excessive. In the first way we become sullen and resentful, and we are ready as the occasion arises to take violent action in protest against our condition. In this case the least advantaged are bitter; they reject society's conception of justice and see themselves as oppressed. The second way is milder: we grow distant from political society and retreat into our social world. We feel left out; and, withdrawn and cynical, we cannot affirm the principles of justice in our thought and conduct over a complete life.[[9]](#footnote-9)

While Rawls was more narrowly focused on the reaction of the least advantaged, there is no reason that his comments do not apply to any person supporting a conception of justice which is completely disregarded by the public social and political culture. Agents who are assumed to have interest in justice will likely experience disaffection when attempts to realize what they believe to be justice are systemically frustrated or undermined, even more so when positions where they seek compromise are all rejected. Rawls identifies two responses to such disaffection; at worst, agents can become antisocial and actively harm others, and at their most benign agents withdraw from public life into their own private sphere. Either result is costly: excessive violence and antisocial behavior is evidently harmful for others in the community, and withdrawal from social life provides a serious cost to both the person withdrawing and the general social community, in that the possibility of forming valuable relationships, like those present in economic, social, cultural, or religious communities, with one another is foreclosed.

Adjudicative mechanisms will thus be appraised in term of the number of disaffected agents that they generally produce. The fewer maximally self-interested agents an adjudicative process yields, the better it meets the pragmatic outcome criterion.

*2.3 Disagreement Over Outcome Measures*

The pragmatic outcome criterion used here is hardly the only measure for adjudicative outcomes. A question thus naturally arises. Who is to say that the kind of disagreement which obtains with respect to justice does not also obtain with respect to measures of adjudicative outcomes?[[10]](#footnote-10) The pragmatic outcome criterion aims to minimize the number of individuals disaffected from social life. But one may also adopt criteria which minimizes variance in satisfaction of justice claims, which maximizes the preference satisfaction of some subset of the population, which maximizes the average of preferences satisfied, and so on. What justifies our appeal to the pragmatic outcome criterion over alternatives?

Recall that the motivation for the pragmatic outcome criterion is functional. Agents in the model are driven to realize their conception of justice. The realization of justice will require that others comply with the rules and requirements that the conception sets. Every agent, by virtue of their interest in realizing justice, also has an interest in maintaining obedience by other agents to the result of adjudication (and, of course, they also wish to maximize the chance of their preferred conception being chosen). Agents are thus concerned primarily with satisfying the preferences of others to the extent that they will follow the result of adjudication, and no more. Maximizing the rate of acceptance of outcomes of adjudication requires minimizing the number of disaffected individuals, and so we see that each agent recognizes the appeal of the pragmatic outcome criterion. The pragmatic outcome criterion thus serves as a functional criterion shared by every agent regardless of their conception of justice, and so appeal to it will be justified despite the background disagreement about justice.

This argument for the pragmatic outcome criterion involves a significant simplifying assumption, namely that compliance is a binary condition which depends only on a threshold of preferences being satisfied (with that threshold being determined by an agent’s tolerance). On a more complex account, individuals may have some probability of complying with the result which increases with increased preference satisfaction. Surely in actual conditions the degree to which preferences are satisfied will matter as much as certain thresholds being met. Consider, for example, a case where some subset of the population has their preferences systemically satisfied to a lower degree than other members of the population. Intuitively, in light of such treatment members of that subset may begin to be noncompliant with the results of adjudication in an attempt to improve their social position.

Dependence between compliance and degree of preference satisfaction is an eminently plausible feature of social dynamics. Introducing such features into the model, however, would significantly complicate our present focus, brining on significant costs by reducing tractability. One would need to introduce variant rates of compliance among agents, relations between preference satisfaction and compliance, diversity among these values, and the possibility of variance along how preference satisfaction influences compliance rate. Such factors are sure to feature into real world rates of acceptance of outcomes of social adjudication. The present aim, however, is to identify what kinds of social choice mechanisms best satisfy the pragmatic outcome criterion, and in pursuit of this aim we isolate a number of relevant features, i.e. the adjudicative mechanisms and agents’ responses to them, and hold fixed or set aside others. To maintain tractability of the model, then, I set aside these considerations of variant acceptance rates dependent on the degree of preference satisfaction and assume the binary threshold account of compliance. That said, while the present account will surely be incomplete (as any model will be) with respect to understanding social mechanisms of adjudication, it will nevertheless advance that understanding by identifying what mechanisms are most reliable and effective at choosing outcomes deemed best according to the pragmatic outcome criterion.

*2.4 Mechanisms of Adjudication*

I will consider four mechanisms of adjudication, which I draw from conventional views in political philosophy, all of which plausibly meet the fairness criterion of adjudication.[[11]](#footnote-11) The fairness criterion, recall, required that each agent’s claim was to be given weighed equally in the process of adjudication. The four mechanisms that I will consider are: a social contract-inspired mechanism of selecting the most acceptable distribution, a Borda count vote, a minimax relative concession bargaining solution, and a simple plurality vote.

2.4.1 Contract Selection

The social contract-inspired mechanism, which I will simply call the “contract mechanism,” is based on selecting that distribution which the greatest number of agents finds acceptable. As the name indicates, this mechanism is inspired by the social contract tradition, which aimed at understanding just social relations on the basis of terms that all the relevant parties (at the appropriate level of idealization) could endorse (Hobbes 1991; Locke 1980; Rawls 1999; Rousseau 1997). The mechanism considered does not require a full-blown social contract: it will not require that *all* agents endorse the distribution. In the context of diversity assumed here such agreement is impossible, and inevitably the adjudicative mechanism demanding unanimous consent would reach an impasse.

The more relaxed model of the social contract used here consists in asking each agent what distributions they would find acceptable, and then implementing the modal response.[[12]](#footnote-12) This adjudicative mechanism aims to establish the largest social contract, with respect to the distribution of goods, possible in the given population.

This social contract mechanism meets the fairness criterion, since it gives each agent equal weight in their acceptance of a distribution and chooses only that distribution that receives the greatest support among agents in equal relative positions.

2.4.2 Borda Count

The second adjudicative mechanism that I consider is a Borda count vote. On a Borda count ballot, agents rank the N available options from first to last. When ballots are tallied, a top ranked option receives N points, the next choice receives N-1, and so on to the last option, which receives one point. The Borda count has the advantage that all “run-offs” occur at once. This makes the model more tractable by setting aside concerns of agenda setting in the voting process.

As a democratic mechanism, Borda count meets the fairness criterion. As is standard of democratic institutions, every agent counts equally in the voting process, with all ballots being offered the same weight, and any permutation of options is an admissible submission, placing no restrictions on what preference orderings are admissible.

2.4.3 Minimax Concession

Minimax relative concessions is a bargaining solution that has received considerable philosophical attention, most notably as David Gauthier’s preferred solution in his contractarian theory (Gauthier 1987, pp. 136-9). As a bargaining solution, minimax relative concessions (“minimax concession” hereafter) holds that agents will accept the outcome that minimizes the maximum concession. A concession is the amount an agent gives up relative to their ideal outcome, in our case this is the distance on an agent’s justice perspective between their current preferred distribution and a proposed distribution.

Adjudication, on the minimax concession mechanism, considers each outcome and registers the concession that each agent would make on their justice perspective if that outcome were to be chosen. The outcome that produces the lowest concession is then selected, or, if multiple distributions produce maximal concessions, as they are likely to, that distribution which produces the lesser number of maximal concessions is chosen. In the case of ties, the outcome is decided at random, with each tied distribution having an equal chance of being chosen.

There are two appealing characteristics of minimax concession that make it worth testing. First, it tracks fair outcomes: if an outcome exists where agents make equal relative concessions, then it is chosen by the minimax concession mechanism (Gauthier 1987, p. 140). This captures the intuitive approach of resolving disagreement by “splitting the difference” between disagreeing parties, with each party moving an equal relative distance from their ideal, where this is possible. This decision heuristic also meets the fairness criterion, since the concessions of agents are weighed equally—in splitting the difference no side is favored.

Second, minimax concession can be used as a representative of bargaining as an adjudicative mechanism. The circumstances being considered within the model are somewhat distinct from traditional bargaining situations: agents care not only for the amount of goods they receive but also the justice of the outcome, and the process is iterated with agents experiencing a preference shift after each round of adjudication. The performance of bargaining may be affected under such non-standard conditions. And, if the results of minimax concession can be shown to generalize to bargaining solutions in general, then minimax concession will serve as a representative of bargaining generally. I address this point in more detail in the results of the default case (§3.2).

2.4.4 Simple Plurality

The fourth and final mechanism I consider is a simple plurality vote. By using another democratic mechanism, we can test whether the results of Borda are simply a product of general democratic characteristics, such as aggregating equally weighed responses, or if the additional informational content contained in a Borda ballot, the relative position of options on individual preference orderings, is an important element of the mechanism. Hence, we have a means of testing whether the information provided by ranked choice voting contributes to the quality of adjudicative outcomes. Since the simple plurality rule is a democratic mechanism, like the Borda count it will meet the weak fairness condition.

3. The Default Case of Adjudication

Having conclude the set-up of the basic elements of the model of adjudication, I now turn to the results of the model in the default case, where the claims of agents have equal weights and agents always respond honestly (assumptions which are later relaxed).

To recapitulate, adjudication consists in each agent making an input, according to the adjudicative mechanism being used, e.g. in the Borda mechanism agents submit their choices ranked from first to last. The outcome for that round of adjudication is then chosen in accordance with the rule of adjudication being used. After an outcome is chosen, agents evaluate the result, judging whether it is acceptable to them. If the outcome falls outside the tolerance range of the agent’s current preferred distribution (which need not be the distribution they believe is most just after the first round), then the outcome is deemed unacceptable, and the agent experiences a shift in preferences towards greater self-interest. This shift was illustrated above in Eve’s justice perspective shown in Figures 1 and 2. When an agent experiences a preference shift, they expand the radius of options that they will consider supporting in the next round. This radius is always centered on their ideally just distribution, the peak on their justice perspective, and their new preferred outcome is that distribution within the radius that maximizes the number of goods the agent has. After the shift, the process is repeated, with a new round of adjudication taking place, and the process continues to repeat until the number of agents experiencing preference shifts each round reaches zero—everyone who would become completely self-interested will at this point have become so.

To replicate the conditions of diversity discussed at the outset of the paper, the justice-perspective, self-interest maximizing distribution, and tolerance range of agents are all determined randomly, with integers being chosen with equal probability within some given range.

*3.1 Tolerance and Stability*

**Fig. 3** The resulting number of egoists under various tolerance ranges

The first notable result of the default case is shown in figure 1.3, which shows the number of “egoists,” or agents who become completely self-interested in the course of adjudication, that each adjudicative mechanism produced under different maximum tolerance ranges. For this case, the population is fixed at 5000 agents, and the number of different distributions at 6.[[13]](#footnote-13) Note that setting the maximum tolerance radius to *T* means that only 1/(*T*+1) of the population will have any specific tolerance value within the radius, since an agent’s tolerance is set to a random integer within the range.

The effect of increasing tolerance is quite clear. As a greater proportion of the population becomes relatively more tolerant, the resulting number of egoists decreases. Contract, Borda, and plurality all follow the same trend, producing almost the same number of egoists, leaving minimax concession as the sole exception. Minimax concession only experienced a significant reduction in egoists once the maximum tolerance range reached five, a situation where half the population needed to find almost all of the outcomes acceptable before minimax concession could produce a reduced number of egoists.[[14]](#footnote-14)

What is most interesting about these results is that no adjudicative mechanism is able avoid making the complete population egoistic until the maximum tolerance range is increased to at least 2, or two thirds of the population has a tolerance of 1 or greater. Some minimum threshold of tolerance is thus necessary for reaching a stable resolution.[[15]](#footnote-15) Without this minimum threshold being met, the adjudicative process is susceptible to what I will call the “chaining effect.” To illustrate, let us consider a simple case of Catherine, Dominic, Eve, and Frederick, all of which represent some portion of the population, and further suppose that their justice perspectives are identical to the perspectives of those they represent. Suppose their ideal distributions are C, D, E, and F respectively, and that adjudicative mechanism being used is a plurality vote. This process is illustrated in table 1, showing the preferred distribution of each representative at each round, marking the winner of each round with an asterisk. In the first round, suppose that Catherine represents marginally more individuals than of the other representatives, causing her preferred distribution C to be chosen. This result is found unacceptable by Eve and Frederick, and those they represent, who then experience a preference shift to F and D respectively. In the subsequent round, Dominic and Frederick have a majority, leading to the selection of D. This result is unacceptable to both Catherine and Eve’s groups, causing them to experience a preference shift to their self-interest maximizing distribution X, which they happen to share. In the next round, Catherine and Eve get their new preferred result to be chosen, leading Dominic and Frederick to shift to their self-interest maximizing distributions Y and Z respectively.

Table 1. An Illustration of the Chaining Effect

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Catherine | Dominic | Eve | Frederick |
| Round 1 | C\* | D | E | F |
| Round 2 | C | D\* | F | D\* |
| Round 3 | X\* | D | X\* | D |
| Round 4 | X | Y | X | Z |

In each round, a new majority is produced which chooses an outcome that causes other groups to experience a shift in preferences, creating a new majority in the subsequent round, which in turn leads to yet another different distribution being chosen. If in response to the previous outcome a new determinate majority is formed each round, with members moving closer to their self-interest maximizing preference every other round, over time every member will become an egoist. I call this the chaining effect, on account of the new majority bringing the previous majority closer to maximum self-interest every other round. As was seen in Figure 3, each adjudicative mechanism was susceptible to this effect at a maximum tolerance range of 1 or lower.

Tolerance stops the chaining effect. If enough agents shift to include the previous winner within their tolerance range, then that outcome will remain chosen in the subsequent round. Suppose, for instance, that in the previous case Catherine’s tolerance range, and the thus the tolerance range of those she represents, included distribution D. In this case, the majority of the population would have agreed upon distribution D in the second round, and supposing that only Eve found it unacceptable, only she and the agents she represented would become egoists, while the rest of the population would have found the result acceptable.

Some minimal threshold of tolerance in the population is thus essential for ensuring that adjudication does not break down. If agents are unable to accept any outcome but their ideal, adjudication breaks down.

*3.2 Performance in Default Conditions*

I now consider the relative performance of the different adjudicative mechanisms in greater detail. Figure 4 and Table 2 below show the average outcome over twenty trials for each adjudicative mechanism. The parameters for these twenty runs were: a population of 500, a maximum tolerance range of 3, and a range of 10 possible distributions. The large population size and range of distributions was chosen in order to ensure that the population would typically be completely diverse by effectively ensuring that no preference orderings repeat. These parameters form the default case and will be held fixed in modified tests below where power asymmetry and strategic manipulation are introduced, serving as a baseline by which to compare the different cases of adjudication. The selection of values is to some extent arbitrary, but the arbitrariness needn’t worry us—as the parameter sweeps in Figures 2, 10, and 11 show, the relative performance of the adjudicative mechanisms is insensitive to the parameter values used.

**Fig. 4** Performance of adjudicative mechanisms under default conditions

Table 2 Reliability of Adjudicative Mechanisms under Default Conditions

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Contract | Borda | Minimax | Plurality |
| Median Steps to Equilibrium | 1 | 1 | 23.5 | 1 |
| Number of Trials that Reached Equilibrium | 20 | 20 | 2 | 20 |

Figure 4 shows the median number of egoists produced in a typical trial, and Table 2 shows the median number of steps that the adjudicative process needed to reach equilibrium. Equilibrium, as I use it, is achieved when a distribution is selected by adjudication, and then always selected in subsequent rounds until adjudication concludes, which is when no more agents experience a shift in preferences.

As Figure 4 shows, minimax concession is the only outlier. Contract, Borda, and plurality all had relatively comparable performances, with plurality producing marginally fewer egoists than either of the other two, and Borda marginally fewer egoists than contract selection. The outlying performance of minimax concession can be seen in table 2 as well; it took significantly more rounds of adjudication to reach an equilibrium, and equilibrium was reached in only two of twenty trials, whereas all other mechanisms reached equilibrium every trial, and typically in the first round.[[16]](#footnote-16)

This significant underperformance of minimax concession is explained by its susceptibility to the changing effect. The effect obtained on every trial of minimax concession and was only stopped in those two rounds that an equilibrium was reached.

Since it is a bargaining solution, we may ask whether the results of minimax concession are indicative of the performance of other bargaining solutions under similar conditions. If it can be shown that susceptibility to the chaining effect is not a quirk of the minimax relative concession solution, but rather a characteristic that follows from it being a bargaining process, then this result can be generalized to bargaining approaches as such— and this would lead us to conclude that bargaining as a method of adjudication would be ineffective because they are intimately related to generally leads to the breakdown of adjudication.

The chaining effect obtains when the following conditions are met. First, there needs to be sufficient diversity among agents such that in the initial round the majority of agents find the outcome unacceptable. Second, agent tolerance must not be so great as to find most outcomes acceptable. If agents were typically extremely tolerant then there would be insufficient preference shifts each round to produce a shifting majority. Third, the method of adjudication must be *over-responsive*, meaning that there is some outcome O for which there is no proper subset of the population V that is decisive for O. When some set of agents V is decisive for some outcome O, if every member of V finds O acceptable, then O will be chosen in the subsequent round. Notice that decisiveness does not require that O be the top option for each member of V, it need only be the case that each member of V accepts O, and so does not experience a preference shift when O is chosen. We can say that a mechanism is more over-responsive when there are more outcomes for which it does not yield decisive sets.

When an adjudicative mechanism is over-responsive in the context of diversity and bounded tolerance supposed here, if the previous outcome lacked a decisive set, then the subsequent changes in preferences in the population are sufficient to change the outcome of adjudication in the next round. If enough outcomes lack decisive sets, if the adjudicative mechanism is highly over-responsive, then each round a new outcome will be reached, a different set of agents experience a preference shift, and so in the subsequent round a different outcome is reached, and so on. This is the chaining effect. Notably, by design neither contract selection, Borda count, nor simple plurality are over-responsive, since under each mechanism each option can have a decisive set.

By hypothesis, the diversity and bounded tolerance conditions obtain here, so it remains to be shown that bargaining solutions generally are over-responsive, rather than just minimax, in order to prove that they are susceptible to the chaining effect. Bargaining will be over-responsive when the outcome of bargaining is a function of all participant preferences and those preferences are given equal weight. If the outcome depends on the inputs, and variation in any input produces variation in the outcome, as it is assumed to under the conditions of equality under consideration here, then it follows that when agent preferences shift in response to a previous outcome, the subsequent result of bargaining should be different from the previous. And, of course, we have assumed that agent preferences are weighed equally here. Bargaining solutions, insofar as they respond to the preferences of all bargainers, will be over-responsive, and hence subject to the chaining effect.

Since bargaining solutions by definition respond to the preferences of all bargainers, and since we have assumed that conditions are diverse and individuals have bounded tolerance ranges, a reply in defense of bargaining accounts must hold that bargaining should not be iterated, and that after the first round of bargaining the resolution is settled indefinitely. That this stops the chaining effect from taking place is clear, since without a shifting winner each round, some set of the population, those who found the initial outcome acceptable, will not be driven to complete egoism. However, one trades away much functionality in adopting this response. In light of the conditions of diversity supposed here, it seems inevitable that conflicts will return, and resolutions called into question. Functional adjudication, insofar as it can resolve conflict diachronically, must be able to settle these disputes as well, and in foreclosing the possibility of future iterations one cannot do so with a one-shot bargain. Thus, insofar as social conflicts necessitating adjudication are recurrent, a functional adjudication cannot adopt such a condition of non-iteration.

Under the conditions of fairness supposed by the default case bargaining generally is an ineffective adjudicative mechanism. The failure of minimax concession shows that iterated bargaining constrained by fairness and characterized by shifting preferences among participants will render the complete population egoistic. Conversely, we see that the other contract or democratic means of adjudication are effective at identifying outcomes with decisive sets of agents, as is evident by their reaching equilibrium from the very first round. We may thus conclude that under conditions of fairness and diversity we have reason to reject bargaining approaches to adjudication in favor of contractual or democratic means. These results, it should be noted, are a product of adjudication under idealized conditions of fairness, conditions which I now relax.

4. Power Asymmetry and Adjudication

Two highly demanding assumptions were made in the default case. First, it was assumed that all agents have equal power in the process of adjudication, and second, it was assumed that agents would always be honest. In this section I relax the former assumption, and in the following section I relax the latter.

To model asymmetric positions between agents in the adjudicative process, I introduce “power,” represented within the model as additional responses that the agent makes in the process of adjudication. So, for example, under the plurality mechanism an agent with 100 power gets 100 additional votes. This, in effect, weighs the agent response to be more significant, agents have more power in that they have a greater degree of influence over the outcome of adjudication.

My aim here is to assess the influence of power asymmetries on adjudicative processes, which means setting aside the complications of a detailed analysis of the distribution of power in social settings. To accomplish this, I introduce several simplifying assumptions. First, power is a function of the number of goods an agent has. An agent has more power if, and only if, they have more goods. This allows for the model to accommodate power asymmetries without introducing further complications like social relations among agents. Second, I assume that for any given agent, the absolute maximum amount of goods that they can accrue is the same. This sets aside complex concerns regarding the distinction between production and distribution. Third, I use a simple linear formula to calculate the power of an agent:

1. Power = height – decay \* distance

As stated above, power refers to the number of additional responses that an agent will have. “Height” refers to the absolute maximum amount of additional responses that an agent can have. “Decay” refers to the rate at which decreases in amounts of goods produce decreases in power, and “distance” refers to the number of distributions between the current distribution and the agent’s self-interest maximizing distribution on the shared distributional ordering, which arranges distributions sequentially from D1, D2…Dn.[[17]](#footnote-17) Call the distribution under which an agent would have the maximal amount of goods that agent’s maximal distribution. To calculate an agent’s power, then, take the maximal amount of additional responses they can have, and subtract from that value the distance between their maximal distribution and the current distribution multiplied by the decay value. This simple linear model of power allows us to test asymmetric adjudicative positions among agents without need for any more complicating variables. Surely, in the real world, power will be a more complex phenomenon, involving far more variables that would not have a linear relationship, but introducing such complications would take away from our present focus.

*4.1 Power Profiles*

The specification of power parameters will largely depend on the kind of power structure that we envision characterizing the agent population. There are infinitely many different possible power profiles, or distributions of power within a social setting, and so to focus our attention we must restrict our consideration to some subset of these profiles. I will focus on four which are based on fairly common power structures been observed in human relations at one point or another: aristocratic, oligarchic, competitive, and egalitarian power profiles.

To keep matters simple, let us suppose that a profile can have a relatively high or low height, and a relatively high or low decay. The high height is 500, or a maximal additional weight to agent responses equal to 10 percent of the normal population, and the low height is 100, or and additional weight equal to 2 percent of the population. The high decay is 100, and the low decay is 20. The values are in some sense arbitrary, one can scale them up or down will little change insofar as the power profile remains the same. What is relevant is restricting the degrees of difference between the profiles to be relatively small, two binary dimensions in this case, in order to make the analysis tractable.[[18]](#footnote-18)

In an aristocratic power structure, those in power have significantly more influence than those not in power, but internal status is highly relevant, leading to significant asymmetries among those who have power as well. In the real world, a prince would have more influence than a duke, and both would have more influence than a commoner. Thus, the aristocratic profile is characterized by a high height and a high decay. The oligarchic profile similarly envisions agents in power having significantly more weight than agents not in power but sees the status among “oligarchs” as relatively equal—thus being modelled with a high height and a low decay. Both the aristocratic and oligarchic power profile, true to the exclusive membership of such institutions of power in the real world, will have constraints on who may have access to such power. This is modelled by adding the condition that if an agent is at a distance of greater than two from their maximal amount of goods, then they get no additional influence.[[19]](#footnote-19) A commoner, no matter their relative status among other commoners, will not be able to enter the ranks of the nobility or oligarchy until they accumulate the same amount of goods as those in power, meaning that power gains with respect to gains in goods is discontinuous until one reaches some threshold of goods.

The competitive and egalitarian profiles, by contrast, have no constraint on who can have additional influence, as long as they have the corresponding amount of goods. The competitive profile uses the low height of 100, capturing the idea that competition erodes the maximal amount of power any one agent can have, and uses a low decay value, showing that there is a wide range of economic positions that agents may occupy in a competitive economic environment. The egalitarian profile is meant to capture a circumstance where agents attempt to mitigate the influence of relative wealth on adjudication to the greatest degree possible but are unable to remove it completely. The low height and high decay values are thus what characterize the egalitarian power profile. For ease of comparison, the characteristics of the four power profiles are shown in table 3:

Table 3. Power Profiles

|  |  |  |
| --- | --- | --- |
|  | Height | Decay |
| Aristocratic | High (500) | High (100) |
| Oligarchic | High (500) | Low (20) |
| Competitive | Low (100) | Low (20) |
| Egalitarian | Low (100) | High (100) |

*4.2 Modelling Additional Responses*

In the case of minimax concession and plurality it is straightforward to model power: in minimax concession an agent’s response is added additional times (which counts as multiple agents making the concession), and in the case of plurality an agent gets to vote additional times.

In the case of contract and Borda, however, the relative complexity of the response available to agents introduces the possibility of different uses of power. First, the agents could use their additional influence to continue to submit their full set of preferences, thus in case of contract all acceptances are repeated, and in the case of Borda the complete ballot is submitted additional times. Or, agents could use their influence to only support their most preferred option, in the case of contract using their additional responses to send only their most preferred option, and in the case of Borda only the top ranked option of the ballot is weighed additionally. Both these approaches find intuitive expressions in social setting. In former, what I call “wide” case, the complete preferences of the agent are given more weight, placing them in the position of a social decision-maker of high standing. In the latter “narrow” case, the agent’s most valued options are those that receive the support of their influence, like a campaigner supporting their preferred candidate despite the fact that they may find other candidates appealing. To accommodate this distinction between the wide and narrow versions of contract and Borda, I test both models of power.

*4.3 Performance Under Power Asymmetries*

With the details of the power case settled, we can now assess the relative performance of the adjudicative mechanisms when power is in play. The same parameters used in the default case are used here, meaning that the population is fixed at 5000, the range of distributions at 10, and the maximum tolerance range at 3.

**Fig. 5** Performance of adjudicative mechanisms under conditions of power asymmetry

**Fig. 6** Performance of adjudicative mechanisms in the power case, excluding the minimax mechanism

Figures 5 and 6 show the performance of adjudicative mechanisms under the different power profiles, with Figure 6 excluding the minimax concession result for greater ease of comparison between the other mechanisms. Comparing the results with those observed in the default case (Figure 4), most striking is the dramatic improvement of minimax concession, since under no power profile did it make the complete population egoistic. This should be expected, since asymmetry of influence in adjudication makes bargaining less over-responsive; the increased relative influence of agents in power is sufficient to produce decisive sets of agents for some options.

Perhaps, then, the rejection of bargaining as a means of adjudication was too quick; real world contexts are likely to involve power asymmetries, and so under real world conditions bargaining will be more effective than it was in the idealized default case. This claim might be supported by the fact that all other adjudicative mechanisms (with the exception of the wide contract) experienced a decline in performance under the non-ideal conditions of power asymmetry. Even so, the relatively worse performance of the other adjudicative mechanisms is still better compared to the improved performance of minimax concession, meaning that under conditions, ideal or non-ideal, the contract, Borda, or plurality mechanisms will perform better than minimax concession (and bargaining mechanisms generally). In addition to providing further support to the contractual and democratic mechanisms, this also provides novel support for the fairness criterion—agents not only have an interest in the weak fairness condition based on their interest in their own claim, they also, insofar as minimizing the number of egoists is society is valuable, have reason to support the fairness condition on the grounds that it produces more functional social orders.

While most adjudicative mechanisms (barring minimax concession) experienced a decline in performance, the wide contract mechanism experienced only a marginal decline in performance, and in the aristocratic and competitive power profiles experienced no change at all relative to the default case. The wide version of power in the contract mechanism was subject to a sort of cancellation effect; since agents in power submitted the complete set of acceptances, other more widely accepted distributions are still relatively competitive with the most preferred options of those in power. The diversity of agents in power produces a diverse range of acceptances, often approximating the preferences of the population with a higher magnitude. The narrow version of the contract, by contrast, would only yield additional support for powerful agents’ most preferred options, foreclosing the possibility of overlap over options which are regarded as suboptimal for agents in power. Thus, in the narrow contract those who preferred options that were not among the most preferred options of the powerful were effectively excluded from the social contract.

The power case thus offers us three general lessons. First, minimax concession, it seems, was indeed subject to the chaining effect. Having weakened one condition necessary for the chaining effect to obtain, the equal bargaining position of parties, minimax concession was no longer subject to it with as great a frequency. Despite the improvement in performance, however, all other mechanisms nonetheless outperformed minimax concession. This leads to the second general lesson. Since the other three mechanism outperform minimax concession in both the default case and in cases of power asymmetries, and they perform better in the default case, agents have a novel reason to endorse the fairness criterion. When the adjudicative process is fair, a better outcome, in terms of minimizing the number of egoists, is chosen. Third, where power leads to diverse agents having similar degrees of power, as in the competitive power profile for the wide version of the contract, power asymmetries have minimal effect. The diversity of the population, in such a case, leads to an approximation of default conditions at simply higher magnitudes.

5. Strategic Response and Adjudication

I now relax the second assumption of the default case, that agents vote honestly. In this section, I will once again assume that agents have equal relative influence in the process of adjudication, setting aside power asymmetries until the following section. It is likely that some agents will lie in the process of adjudication in order to increase the likelihood that outcomes that are higher ranked on their perspective are chosen. Introducing strategic responses allows us to model the influence of such strategies on the process of adjudication.

*5.1 Strategic Responses*

Strategic behavior will vary with the mechanism under consideration. In contract and minimax concession the strategic response will be straightforward. In the contract mechanism, agents can improve the chance of their most preferred option winning by always claiming that only their most preferred option is acceptable to them, regardless of their tolerance range. In minimax concession, the agent can exaggerate their concession on every option except for their most preferred choice. For distributions that are not an agent’s most preferred, they will claim that their concession is maximal.

Plurality provides a more interesting case, since agents are already only supporting their most preferred option. One way that agents can manipulate their ballot is by voting for the candidate most likely to win within their range of acceptable results. This kind of behavior is in line with the motivations of strategic behavior, since it maximizes the probability that an agent gets a result that they find acceptable. With this kind of a strategy, agents vote on the basis of the expected utility of various outcomes, weighing them by the probability that they obtain, given the current pattern of voting.

Borda, due to the high information content of its ballots, provides agents with a variety of ways to manipulate their ballot. One common strategy for the manipulation of Borda ballots consists in agents “burying” competitors or placing other options that might win lower on the ballot than the agent would have placed them had they been honest. Agents can perform two kinds of burying, forward- and backward-looking. Forward-looking burying consists in agents placing at the bottom of their ballot the current front-runner of voting (assuming it is not their most preferred option). Backward-looking manipulation consists in the agent burying the previous result of adjudication (assuming again that it was not their most preferred option). Using these two strategies as a basis, Borda agents have a third strategy available to them, one that incorporates both backward- and forward-looking manipulation. I call this the compound strategy, which consists in agents assigning the previous winner of adjudication the lowest spot on the ballot, and the current front-runner the second lowest spot.

*5.2 Strategic Manipulation and Adjudicative Performance*

The primary strategic parameter is the frequency of strategic agents in the population. Some portion of the population is randomly assigned the strategic type, which responds strategically every round. Figure 7 shows a parameter sweep of varying percentages of the population being strategic.

**Fig. 7** Resulting number of egoists under varying rates of strategic manipulation

Notice, first, that the performance of minimax concession drastically improves as the rate of strategic agents increases. This occurs because as more agents become strategic minimax selection begins to approximate a plurality vote in default conditions. At the limit case, where all agents are strategic, the only distribution that any agent will not respond as making a maximal concession for is their most preferred option. The result of minimax concession will thus be the distribution that is most preferred among the greatest number of agents, which would also always be the result of an honest plurality vote.

The second notable result is the failure of adjudication when the backward-looking strategy is introduced to the Borda mechanism. The decline in performance of the backward-looking Borda strategy is a result of the backward-looking strategy undercutting the possibility of attaining equilibrium. Once a significant portion of the population uses the backwards-looking strategy, the probability of reaching an equilibrium becomes incredibly slim, since a significant portion of the population will bury the previous result of adjudication, making it unlikely to win. As a result, the backward-looking strategy makes such ranked choice voting susceptible to the chaining effect.

Compound and forward-looking Borda strategies suffer no such set-back, due to the cancellation of different instances of forward-looking manipulation. Agents bury the current front-runner at the time of their vote, and, if each agent buries the current front-runner any gains from previous manipulation are counteracted, since the preferred option of earlier voters will in turn be buried once they become the front-runner. The compound strategy, which gives priority to backward-looking manipulation, still produces this cancellation effect, allowing an equilibrium to be reliably reached.

Having seen the influence varying proportions of strategic voters have on the outcome of adjudication, I set the probability of agents being a strategic type at 30 percent for the representative trials. Recent empirical research indicates that actual voters manipulate their ballot at approximately this rate (Spenkuch 2018). Setting the rate of strategic voters at thirty percent provides the most reliable reflection of the effectiveness of adjudicative mechanism in real contexts.

**Fig. 8** Performance of adjudicative mechanisms with a 30% rate of strategic manipulation

The results of the representative trials approximate the results of the various mechanisms under the 25 percent case in the Figure 7. Compared to the default case, contract was subject to no influence. Just like minimax, strategic manipulation in the contract mechanism simply approximates honest plurality voting, which as we saw in the default case was a relatively effective adjudicative mechanism. Similarly, the compound Borda case was subject to minimal change in performance relative to the default case, suggesting the cancelation effect was indeed effective at stopping the chaining effect. By contrast, the forward-looking Borda case was subject to a modest decline in performance, and the backward-looking case subject to severe declines in performance, for reasons discussed above. Minimax experiences an improvement in performance relative to the default case, since approximating plurality voting was sufficient to undercut the chaining effect. Lastly, plurality was subject to a modest decline in performance. Considering the changes in contract, minimax, and plurality, we arrive at the perhaps surprising result that a simple plurality vote is an effective mechanism of adjudication.

To conclude, it is worth considering which versions of strategic manipulation in the Borda mechanism are most plausible. The forward-looking and compound of versions of Borda, I believe, will be too informationally demanding to be feasible in real contexts. For individuals to bury the front runner of the ballot, they must have highly accurate information regarding the preference orderings of all other agents. More plausibly, agents will use the backwards-looking strategy in real contexts, given the relative ease, and efficacy, of using the strategy (when used by a sufficient number of the population the previous outcome is *never* chosen). If this less informationally demanding Borda strategy is used, then Borda will experience a significant decline in performance in purely strategic contexts.

6. Adjudication under Mixed Conditions

This final case considers the performance of adjudicative mechanisms under the non-ideal conditions where power asymmetries and strategic manipulation both obtain. This case consists simply in combining the parameters of the power and strategy case. Each adjudicative mechanism is tested under the four different power profiles, with one third of the population being strategic. On account of its greater feasibility compared to other strategies, and to simplify the presentation of results, only the backward-looking strategy of Borda is tested here.

**Fig. 9** Performance of adjudicative mechanisms under mixed conditions including both power asymmetry and strategic manipulation

Figure 9 shows the performance of each adjudicative mechanism under the mixed conditions of power and strategy. Comparing these results to the pure cases of power (Figures 5 and 6) and strategy (Figure 8), several changes can be noted. First, while many changes in performance are small, almost every mechanism was subject to a decrease in performance under mixed conditions. Minimax concession experienced a marginal change in the other direction, having a performance that was generally the same as under pure strategy, which is an improvement over its performance in pure power. Perhaps unintuitively, this speaks to the consistency of plurality voting, since as discussed above (§5.2) strategic voters under minimax concession have behavior that approximates a plurality ballot.

The wide version of Borda experienced a significant decline in performance when compared to its pure strategy results, suggesting that interactive effects between strategic behavior and power produced conditions where establishing an equilibrium was very unlikely. The narrow version of the Borda power case, by contrast, experienced a significantly improved performance with respect to its pure strategy result. This suggests that the introduction of power asymmetries is enough to counteract the destabilizing effects that backwards-looking manipulation had for Borda’s performance; individuals in power had sufficient influence to ensure that in subsequent rounds strategic voting was still insufficient to lead to a different result.

*6.1 Adjudicating Actual Social Conflict*

With the performance of adjudicative mechanisms tested under several different conditions, what can be said with respect to our initial question of how agents should adjudicate otherwise irresolvable disputes about distributive justice? The results here point to an answer that is perhaps disappointing: the precise mechanism of adjudication to be used will likely depend on many local conditions, such that the idealized model developed here cannot recommend any precise mechanism. Despite a number of idealizing assumptions outcomes from the model nonetheless showed a considerable degree of sensitivity to the conditions of adjudication, such as the rate of strategic manipulation and kind of power structure in place.

While the results of the model can offer no determinate recommendation, certain notable results still warrant attention, serving as pro tanto recommendations for and against certain adjudicative mechanisms. Contract and plurality exhibited a high degree of consistency across cases, with the destabilizing conditions of power asymmetries and strategic voting producing only marginal decreases in performance, if any were produced at all. The model thus offers some preliminary reason to prefer contract or plurality mechanisms of adjudication. Minimax concession, by contrast, was consistently the worst performing mechanism. Generalizing the performance of minimax concession to bargaining generally, we have pro tanto reason against using bargaining procedures as a means of adjudicating disputes (§3.2).

These suggestions, are, of course, only pro tanto. A more complete, and empirically sensitive, theory of adjudication will need to be developed before a more comprehensive answer can be offered to the question of how we should adjudicate social disputes.

Conclusion

With the significant results of the model noted above,[[20]](#footnote-20) I will conclude by highlighting what proved to be a central factor in establishing an equilibrium in the process of adjudication—allowing for agents to trade-off between their ideal of justice and their self-interest. Self-interest has often been disregarded by political philosophers discussing distributive justice, often considering it to be irrelevant to questions of distribution. This presupposition does not hold here, if agents could not trade-off some of their interest in justice for gains in self-interest, then the plurality of agents supporting the existing equilibrium will be insufficient to sustain it.[[21]](#footnote-21) If agents were to insist on only ever endorsing their most preferred outcome (based on their conception of justice) then we would find ourselves in conditions similar to those where agents have relatively low levels of tolerance (Figure 3). In such cases, it is almost inevitable that the entire population becomes egoistic, leading to the impossibility of a peaceful and stable social order.[[22]](#footnote-22)

Some are sure to bemoan this result as a testament to the crooked nature of human beings. It seems that the only way to secure a cooperative social order in light of significant disagreement on distributive justice is to “buy off” people, paying them to betray their reasoned convictions.

This pessimistic response is not the only reading of the results. For we see also that by allowing people to pursue their self-interest, we make peaceful cooperation among socially-inclined agents a real possibility, even in light of significant diversity. The pervasive diversity observed at the outset of the essay is thus not so prohibitive as to render peaceful social life on agreeable, though suboptimal for some, terms impossible. Contexts characterized by such deep diversity are precisely those where it seems there is no possible way forward for diverse persons to live in a shared social world, except perhaps through the use of violence or coercion. But the results of the model show that this is not so. Agents, when allowed to pursue their self-interest at some trade-off to realizing their conception of justice, will be able to live together in a social world they find generally acceptable, despite being faced with deep and irresolvable conflict for which there is no forthcoming resolution. Self-interest is thus a crucial element in making a peaceful social order possible, without it agents would be left with only zealous insistence that their view is right, and so must either give up on the possibility of shared social life with other diverse persons or else take their dispute to blows (Hobbes 1991, ch. 5 para. 3).

Appendix

**Fig. 10** Resulting number of egoists under varying population size

**Fig. 11** Resulting number of egoists under varying distribution range

**Fig. 12** Resulting number of egoists under varying power height

**Fig. 13** Resulting number of egoists under varying power decay value

References:

Anderson, E. S. (1999). What Is the Point of Equality? *Ethics*, *109*(2), 287–337.

Bicchieri, C. (2005). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge: Cambridge University Press.

Bicchieri, C., & Chavez, A. (2010). Behaving as Expected: Public Information and Fairness Norms. *Journal of Behavioral Decision Making*, *23*, 161–178.

Bicchieri, C., & Mercier, H. (2013). Self-Serving Biases and Public Justifications in Trust Games. *Synthese*, *190*(5), 909–922.

Brams, S. J., & Fishburn, P. C. (1978). Approval Voting. *American Political Science Review*, *72*(3), 831–847.

Cohen, G. A. (1995). *Self-Ownership, Freedom, and Equality*. Cambridge: Cambridge University Press.

Elster, J. (2003). “The Market and the Forum: Three Varieties of Political Theory.” In *Debates in Contemporary Political Philosophy: An Anthology*, edited by Derek Matravers and Jon Pike, 325–41. New York: Routledge.

Elster, J. (2016). *Sour Grapes: Studies in the Subversion of Rationality*. Cambridge: Cambridge University Press.

Frankfurt, H. (1987). Equality as a Moral Ideal. *Ethics*, *98*(1), 21–43.

Gaus, G. (2015). Public Reason Liberalism. In S. Wall (Ed.), *The Cambridge Companion to Liberalism* (pp. 112–140). Cambridge: Cambridge University Press.

Gaus, G. (2016). *The Tyranny of the Ideal: Justice in a Diverse Society*. Princeton: Princeton University Press.

Gaus, G. (2018). Self-organizing moral systems: Beyond social contract theory. *Politics, Philosophy, and Economics*, *17*(2), 119–147.

Gauthier, D. (1987). *Morals by Agreement*. Oxford: Oxford University Press.

Heifetz, A., C. Shannon, and Y. Spiegel. (2007). “The Dynamic Evolution of Preferences.” *Economic Theory* 32 (2): 251–86.

Hobbes, T. (1991). *Leviathan*. (R. Tuck, Ed.). Cambridge: Cambridge University Press.

Lloyd, S. A. (2003). Hobbes’s Self-effacing Natural Law Theory. *Pacific Philosophical Quarterly*, *82*(3–4), 285–308.

Locke, J. (1980). *Second Treatise of Government*. (C. B. Macpherson, Ed.). Indianopolis: Hackett.

Nagel, T. (2005). The Problem of Global Justice. *Philosophy and Public Affairs*, *33*(2), 113–147.

Nozick, R. (1974). *Anarchy, State, and Utopia*. New York: Basic Books.

Parfit, D. (1997). Equality and Priority. *Ratio*, *10*(3), 202–221.

Pogge, T. W. (1992). Cosmopolitanism and Sovereignty. *Ethics*, *103*(1), 48–75.

Rawls, J. (1999). *A Theory of Justice* (Revised). Cambridge: Harvard University Press.

Rawls, J. (2001). *Justice as Fairness: A Restatement*. Cambridge: Harvard University Press.

Robson, A. J., and L. Samuelson. (2011). “The Evolutionary Foundations of Preferences.” In *Handbook of Social Economics*, edited by Jess Benhabib, Alberto Bisin, and Matthew Jackson, 221–310. Amsterdam: North Holland.

Schmidtz, D. (2002). How to Deserve. *Political Theory*, *30*(6), 774–799.

Sen, A. (1982). Equality of What? In *Choice, Welfare, and Measurement* (pp. 353–371). Cambridge: Harvard University Press.

Sen, A. (2017). *Collective Choice and Social Welfare*. Expanded Edition. Cambridge, MA: Harvard University Press.

Spenkuch, J. L. (2018). Expressive vs. strategic voters: An empirical assessment. *Journal of Public Economics*, *165*, 73–81.

Tversky, A., and R. H. Thaler. (1990). “Anomalies: Preference Reversals.” *The Journal of Economic Perspectives* 4 (2): 201–11.

Vallier, K. (2010). Production, Distribution, and J. S. Mill. *Utilitas*, *22*, 103–125.

Van Schoelandt, C., & Gaus, G. (2018). Political and Distributive Justice. In S. Olsaretti (Ed.), *The Oxford Handbook of Distirbutive Jusitce* (pp. 283–305). Oxford: Oxford University Press.

Weithman, P. (2011). *Why Political Liberalism: On John Rawls’s Political Turn*. Oxford: Oxford University Press.

1. The results discussed here may naturally be applied to other accounts of social decision-making to account for how to proceed under the absence of consensus. Cf. Elster 2003 on the need to proceed without consensus in deliberative democracy. [↑](#footnote-ref-1)
2. On realization conditions see Gaus 2016, pp. 21-6. [↑](#footnote-ref-2)
3. “Goods” is used in a wide sense, describing those things which are desired by agents and which are the objects of distributive concern. The goods being distributed in the model may interpreted to be whatever the reader’s preferred account of distributive justice suggests they should be. [↑](#footnote-ref-3)
4. My thanks to an anonymous reviewer for drawing my attention to this point. [↑](#footnote-ref-4)
5. For present purposes I assume that the distribution of goods can be changed with relative ease. Of course, this is a simplification, which abstracts from factors tying the distribution of goods to their production. See Vallier 2010. [↑](#footnote-ref-5)
6. See Gaus 2016, pp. 43-56 for a formal model of justice perspectives. [↑](#footnote-ref-6)
7. In order to simplify future computation, I will suppose that for all agents when the x-axis is ordered sequentially, e.g. D1, D2…Dn, then the self-interest score is single-peaked, with the peak being in different locations for the agents. [↑](#footnote-ref-7)
8. On may be concerned here with whether individual preferences will be coherent. On the assumptions here, individual preferences must be synchronically consistent, but may be diachronically inconsistent (as has been observed in behavioral treatments, e.g. Tversky and Thaler 1990). The synchronic coherence of individual preferences is shown by indexing individual preferences to times. For any individual i, then, their preference relation must be indexed to some time t, Rt. Let R0 be i’s initial preferences, defined as discussed above. We assign subsequent preferences with the function F: Rt, e 🡪 Rt+1 where e is the set of exogenous conditions sufficient to cause individual preferences to shift, in our case that is a distribution’s falling outside i’s tolerance range. As defined in the model, the top option in Rt is the justice and self-interest optimizing option, and the placement of other options is determined by their distance from the top option on i’s justice perspective, generating an ordering. Hence at any given time i's preferences form a coherent ordering, though over time the ordering changes in response to exogenous factors allowing for the possibility of diachronic inconsistency, modelling the behavioral results which inspire this account. I thank an anonymous reviewer for drawing this concern to my attention. [↑](#footnote-ref-8)
9. Rawls, *Justice as Fairness: A Restatement*, p. 128. [↑](#footnote-ref-9)
10. My thanks to an anonymous reviewer for raising this point. [↑](#footnote-ref-10)
11. This project thus naturally extends social choice investigations into dynamic social contexts. For a philosophically oriented, and accessible, overview of the social choice project see Sen 2017. [↑](#footnote-ref-11)
12. This procedure is identical to approval voting, see Brams and Fishburn 1978. [↑](#footnote-ref-12)
13. Parameter sweeps of both population size and distribution range are shown in the Appendix. The relative performance of each adjudicative mechanism at the different variables is subject to almost no variation, making the choice of distributive range and population size is arbitrary. Hence, in order to keep values fixed, a population size of 5000 and distribution range of 6 are used. [↑](#footnote-ref-13)
14. That the complete population could become completely egoistic in the case of low tolerance values is, of course, a completely counterfactual result. It is hardly the case that in any real social group every single person will become an egoist. This result, I believe, is best understood as a case where adjudication would have broken down in actual conditions. Agents, seeing that no result is forthcoming, will withdraw from adjudication, treating results as unauthoritative. The costs associated with complete egoism, however, will likely still obtain. Agents withdraw from the social process of resolving conflict, at best foregoing the benefits of having a resolution for a shared problem, and at worst take it upon themselves to enforce their preferred outcome. [↑](#footnote-ref-14)
15. Note that even if tolerance improves upon the pragmatic outcome criterion there may still be reason to withhold from making individuals maximally tolerant. Trivially, if every agent is made maximally tolerant the pragmatic outcome criterion will be maximally satisfied. Nevertheless, under such conditions it no longer seems plausible to interpret the agents as concerned about justice, since they will accept the outcome of adjudication regardless of how it performs with respect to justice on their perspective. The social context we are concerned with here is one in which agents care deeply and disagree about justice, such that increasing tolerance too highly will constitute a departure from the social situation of interest. Moreover, surely we have some reasons to encourage individuals to develop a conception of justice which they then attempt to advance. There is, however, a relevant question as to what social reformers should do when given the opportunity to influence individual preferences. From the above, it follows that they should optimize between individual commitments to justice and tolerance. Too strong a commitment to justice makes social cooperation impossible, and too little trivializes justice. The significance of justice, whatever the correct conception, and satisfaction of the pragmatic outcome criterion are thus countervailing forces on the selection of adjudicative mechanisms and shaping of individual preferences in actual social contexts. [↑](#footnote-ref-15)
16. Plurality was the exception here, taking two rounds to reach equilibrium on 9 rounds of 20. [↑](#footnote-ref-16)
17. Recall that it was assumed that all agents’ self-interest preference is single-peaked when the distributions are ordered sequentially. See §2.1. [↑](#footnote-ref-17)
18. See Figures 12 and 13 for parameter sweeps of power height and decay respectively. For the height test decay was held fixed at 100, and for the decay test height was held fixed at 500. As the Figures show, the only mechanism with significant variation is median selection, the relative performance of the remainder of the mechanisms is insensitive to the parameter values. [↑](#footnote-ref-18)
19. Note that with a radius of 2 approximately half the population will have access to some amount of power. The criterion is fairly inclusive in order to ensure that some gradient of rank can obtain in the aristocratic model, since a lower radius would not allow one to model high degrees of variation among those in power. [↑](#footnote-ref-19)
20. The results I have in mind are: the relative failure of bargaining mechanisms (§3.2), the instrumental justification of the condition of fairness (§4.3), and the robustness of both contract and simple plurality mechanisms (§6). [↑](#footnote-ref-20)
21. Three mechanisms were put forward to explain this shift: resolution of cognitive dissonance, response in accordance with a norm of reciprocity, or response to the observation of a norm of justice being eroded. Realistically, things will hardly be so uniform. Actual individuals will likely experience a shift for many diverse reasons, and will have preferences over a number of domains which they may shift toward. To take just one example, a politically disillusioned individual may find themselves more attracted to their religious commitments. The mechanism of shifting preferences, and the preferences among which they agents shift are thus likely to be far more complicated than the discussion here has assumed. Allowing for a greater range of trade-offs, however, does not challenge the result of the model showing that when no trade-off between justice and other interests is permitted adjudication cannot reach a stable outcome. Allowing a greater range of responses does not change the result which obtains when no response is permitted. [↑](#footnote-ref-21)
22. A population becoming completely egoistic is best treated as indicative of the breakdown of adjudication in real contexts (see note 9). In such a case, Rawls’ comments on the anti-social behavior of those disaffected from the public conception of justice still apply (§2.2)—the majority of individuals will still see that their conception of justice is unwelcome among the public, adopting anti-social attitudes of isolation or aggression. [↑](#footnote-ref-22)