11

Consensus, Contract, and Diversity

John Thrasher

1. Introduction

Agreement is at the heart of many theories of political justification, most notably in social contract theories.[1] In its most general form, political contractualism is the view that political life's fundamental principles, norms, or rules are justified because they can garner public agreement in some form. The problem I will identify and attempt to remedy in this chapter is one common to theories of political contractualism. It is found in contractual theories, not because of any distinctive defect in contractual theories; rather, the issue is a general one in political philosophy that is easier to diagnose in these theories, because they wear their justificatory structure on their sleeve. They raise the justificatory bar explicitly in ways that most other theories do not.

The problem I identify here concerns diversity—in thought, beliefs, values, and interests. Diversity of this sort is pervasive in modern, open societies. This pervasive diversity poses a serious problem for any political theory that aims to build the foundations of political legitimacy on a stable conception of justice. Most contemporary contractual theories do this, but so do most intuitionist theories (e.g. most versions of egalitarianism, utilitarianism, or libertarianism). The problem is especially stark in contractual theories, though, and because of this, I look at the problem here in the context of political contractualist theories, first in the work of early modern contractualists, then in the work of Rawls, before turning to two contemporary attempts to solve this problem in the work of Michael Moehler (2018) and Ryan Muldoon (2016). Moehler and Muldoon construct their contractual theories specifically to address the problem of diversity, and hence, these are the most likely ways of avoiding the problem. While both approaches are significant advances on what came before, neither is entirely

[1] Typically, this agreement is construed counterfactually, as in the contractual theories of John Rawls (1971, 1996), David Gauthier (1986), and Gerald Gaus (1996, 2011). Although many have objected to the use of counterfactual models of contractual theories by arguing that hypothetical consent or the idealization of the representative agents undermines the justificatory use of the model, my goal here is not to address those concerns, which have been the subject of considerable debate already (see, e.g., Dworkin 1976; Enoch 2013; Suikkanen 2014; Huemer 2013). Instead, I will address what I take to be a more significant problem with political contractualism, one that persists even if we accept counterfactual models of agreement.

satisfactory. This suggests, I argue, that the problem resists a satisfactory solution. The general failure of contractual theories to solve the problem of diversity poses a serious problem for the viability of those theories.

All is not lost, though. There is a cure for the problem of diversity, one that contractualists are well placed to adopt. The heart of the problem is in the target of justification, not in the contractual method of justification. Seeking consensus on principles of justice as the fundamental source of political legitimacy is the problem. In response, I propose a version of contractualism that rejects seeking consensus on high-level principles of justice as the basis of political justification. In so doing, I develop the framework for political contractualism that avoids the problem of diversity by being more conducive to a dynamic, open society.

2. Contractualism and the Target of Justification

The early modern social contract theorists such as Hobbes (1651), Spinoza (1677), Locke (1681), Rousseau (1762), and Kant (1797) were primarily interested in justifying the authority of a sovereign. This is especially clear in Hobbes's (1651) *Leviathan*. The state of nature, for Hobbes, is a state of moral anarchy where no stable social conception of morality and justice can exist. To rectify this, Hobbes argues that a sovereign must be established whose reason will act as the reason of each member of the public, deciding disputes and fixing the terms of justice and morality. According to Hobbes, the sovereign embodies public reason and, in doing so, defines justice. Sovereignty, for Hobbes, is thus required for justice. The other early modern social contract theorists, though less radical in identifying justice and society directly, also seek to establish justice by establishing a sovereign of some form.

The postwar revival of political contractualism inspired by John Rawls inverts the Hobbesian formula. Instead of justifying a sovereign that defines and provides the normativity for justice, Rawls and later contract theorists use the contractual method to justify the principles of justice directly. Justice, identified through agreement, becomes the direct source of sovereignty, as well as the source of political normativity (i.e. legitimacy).

Social contract theories of whatever sort take the general form of a model with four basic parameters (N, M, R, N^*).[2] The social contract models the reasons that real people (N^*) have for endorsing and complying with a particular set of principles, rules, or norms (R) by showing that model reasoners (N) have reason to choose those principles, rules, or norms (R) in the model of agreement (M). To generate the normativity of this model, there must be a direct link between the

---

[2] This account follows Thrasher (2020) and D'Agostino, Gaus, and Thrasher (2021).

reasons and reasoning of the model reasoners (N) and actual reasoners (N^*). What it means to "share" reasons in this context is a matter of dispute among contract theorists that should not be settled here (see Vallier 2018). But we can stipulate that, at a minimum, we should be able to model the reasoning of the contractors in such a way that it can be coherently represented as the reasoning of actual persons and vice versa. In Thrasher (2019), I have described this as a *representation* condition of contractualism. A slightly different characterization of the problem comes in the form of avoiding what Nic Southwood (2019) has described as the "conditional" and "concessional" problem for contractualism. Whatever its form, the issue is the link between the counterfactual and actual reasons of agents for whom the contractual model is meant to be normatively meaningful. Without such a link, the normativity of the contractual model is suspect.

Each of these parameters (N, M, R, N^*) can be specified in various ways. The shape of a particular contractual theory depends on the precise way these parameters are set. We can characterize the general form of any contractual theory as follows:

Contractualism N chooses R in M, which gives N^* reasons to endorse and comply with R in the real world insofar as the reasons N has for choosing R in M can be shared by N^*.

Political contractualism, in this sense, tends to focus on justice as the *target* of justification (R) or some similar analog such as social morality. Sovereignty, authority, and other central political concepts are understood in relation to justice. This need not be the case; other targets of justification are possible. Hobbes makes sovereignty the target of justification, for instance. James Buchanan and Gordon Tullock (1962) develop a contractual theory in *The Calculus of Consent* that makes a supermajority voting rule the target of justification. Nevertheless, most contemporary theorists use justice as the target of justification. Virtually all non-contractual theories also use justice as the target of justification (e.g. Nozick 1974; Schmidtz 2006; Cohen 2008; Estlund 2019).[3]

3. Uniqueness, Stability, and the Paradox of Diversity

The contractual model aims to do two things. First, to show that some set of principles, rules, or norms can be justified. In Thrasher (2020), I have called this the *existence* claim in contractual theories. To meet this condition, the set of principles, rules, or norms that the contract justifies—the target of justification—

---

[3] This is true both for what Chad Van Schoelandt (2020) describes as "functionalist" as well as "non-functionalist" theories of justice, and for ideal and non-ideal theories of justice.

cannot be empty. For most theories, this target will be a conception of justice. Typically, theorists aim to have the contractual model justify one and only one conception of justice. Rawls's model generates justice as fairness, Gauthier's model generates minimax relative concession, Harsanyi's models generate a principle of average utility, etc. However, the need for uniqueness in one's theory of justice can undermine the normativity of many contractual models (Thrasher 2014). This might lead one to reject uniqueness as a desideratum of one's solution to the existence problem. Rejecting uniqueness is not without its dangers, however.

There are two possibilities: Either the result of the contractual model is indeterminate or it may be pluralistic. In the first case, indeterminacy can be caused by a lack of specification or an indecisive selection procedure, for instance, as in Gaus's (2011) social contract model, which uses a social choice procedure that can only specify a maximal set of options, each of which is undominated by the others. In a sense, all the options in this set result from the contractual model. He uses a coordination mechanism to specify the result fully, but because the coordination mechanism is path-dependent, the result cannot be determinately predicted *ex ante*. We see a less stark version of this in *Political Liberalism* (1996), where Rawls argues that his preferred conception of justice—justice as fairness—is merely one of a "family" of reasonable liberal conceptions of justice. Indeterminacy is hard to distinguish from pluralism—the claim that multiple conceptions of justice exist.[4]

In any case, the problem with a non-unique solution is that whatever justification there is for the conception of justice must be normatively compelling to all citizens. If the process for selecting it is random or indeterminate, this will naturally create gaps in the justification for the conception of justice. If the existence problem is solved, then the justification will either be unique or indeterminate (we can fold pluralism and indeterminacy together for our purposes here).

The second aim of the social contract model is to show that whatever conception of justice is justified by agreement will be stable. Stability can mean many things in this context. However, the minimal conception of stability is that the rationale for selecting a particular conception of justice will persist once the model contractors are replaced with real people. So, at least minimally, stability requires what we can call a *representation* condition holds. We can think of this as a stable consensus on the conception of justice. In cases where the conception of justice is unique, consensus will be directly on the conception of justice itself. In contrast, when the model's output is indeterminate, consensus will be indirect and focus on a process or mechanism of generating determinacy. A consensus of this sort is a stable equilibrium where each citizen supports the conception of justice and it is common knowledge that all do so.

---

[4] On some of the issues and challenges with pluralism about justice, see Gjesdal (forthcoming).

There are several ways that consensus of this sort can be undermined. For instance, if the choice in the contractual model is partitioned in a way that changes the basis of choice for the representative choosers, representation is directly imperiled by making choice path-dependent and incoherent (Thrasher 2019). Some forms of consensus are also fragile in the face of weakening the common knowledge condition (Thrasher and Vallier 2015). The first problem can be solved by not partitioning choice in the contractual model. Some have argued that the second problem can be solved by creating credible commitments among political agents that endorse the standards of public reason, the indirect mechanism of consensus (Kogelmann and Stich 2016).

Regardless, there is a related but more significant structural problem that does not suggest an obvious solution and that I have called the *paradox of diversity* (Thrasher 2020). This problem arises out of a fundamental conflict between the two aims of the social contract: existence and stability. The paradox shows that publicly justifying any conception of justice in a diverse society is difficult or perhaps impossible. The reason is that solving the existence condition requires constructing the parameters of the contractual model such that they generate a particular result. Doing so, however, makes the basis of consensus highly specific and narrowly focused. In other words, to generate a specific result, the theorist will need to reduce considerably the diversity of the reasoners in the model. Consensus on this conception of justice will only hold for that narrowly construed, normalized set of reasoners. However, success in the model will not lead to success in the real world because the essential diversity of the actual society will not be modeled in the contractual consensus.

This general problem is amplified in an open, dynamic society, where a consensus on a conception of justice must remain stable as diversity increases in scope and character, that is, as a society gets more diverse in different ways. Call this the problem of *dynamic diversity*. A theory of political contractualism should be able to explain how a society could persist while becoming more diverse and tolerant. Alternatively, if not, it should explain how consensus is possible given substantial diversity, both now and over time. I look at two possibilities in Section 4 and Section 5.

4. The Insulation Strategy

The first approach argues that moral and political diversity can be neutralized for consensus on justice by finding a common source of agreement. One promising approach is to restrict the kinds of reasons that one can bring to bear in the contractual model. Rawls (1971) and many of his followers restrict the reasoning of the contractual agents to some "reasonable" set of reasons. The success of this restriction is the source of considerable controversy. In fact, Rawls uses several

devices in his contractual model to normalize the interests of the parties to the contract, including the veil of ignorance and maximin reasoning. As noted earlier, by the time we get to *Political Liberalism* (1996), Rawls himself seems to question the success of this approach. As Gaus (2017: 30) notes, Rawls seems to conclude that different constructions of the original position will generate different results:

> A theory of justice is characterized by the choice from a certain normalized perspective, but there are multiple partial normalized perspectives that yield different principles. Now if one acknowledges that there are other reasonable normalizations that yield opposing answers, in what sense can one plausibly claim that one has identified the principles of justice for the definitive ordering of social claims based on your preferred normalization?

Moehler (2018) argues that rather than looking for a shared basis of agreement in some substantive view, we should look to a shared conception of rationality. Gauthier (1986) argued that rationality could serve as such a basis. He argued that, regardless of the other ways humans differ, they are alike in their rationality. Following Hobbes, Gauthier argued that rational individuals want to find a shared structure based on a rational agreement for mutual benefit.

As commentators since Hobbes have noted, the main problem with this approach is that although economic or instrumental rationality may be the correct normative conception of rationality, actual people in the real world seem far from economically rational. The "heuristics and biases" literature in economics and psychology catalogs how real people deviate from the core requirements of rationality. To many behavioral economists, this shows that people do not typically abide by the norms of rationality (Tversky and Kahneman 1986; Kahneman and Tversky 1979; Kahneman 2003).[5] Another school of thought, championed by Gerd Gigerenzer (2010), argues that these results indict the theory of rationality, not the reasoning of individuals, which may be "ecologically rational" (see Smith 2008). The experimental economists Vernon Smith and Bart Wilson (2019) have a third view: Economic rationality is a "special" conception of rationality that individuals use in specific contexts but not others. Regardless of who is correct, there seems to be a significant gap between the reasoning of instrumental agents in a contractual model and those outside it, which calls such a model's normativity and external validity into question.

Gaus (2011) raises a different objection in his work. He argues that instrumental agents will not be able to agree to anything like the norms we commonly think of as justice or social morality. These norms require compliance based on them

---

[5] Robert Sugden (2018) argues that this way of thinking about behavioral economics and rationality is mistaken.

being norms rather than on the benefits one gets from following them. Purely instrumental agents, however, cannot credibly commit to following rules where they would benefit from breaking them. Similarly, enforcing those rules and norms requires agents to engage in costly punishment, which instrumental agents will often see as too costly to be worth it. Gauthier (1986) makes a similar argument with regard to the logic of what he calls "constrained" vs. "unconstrained" compliance. For Gauthier, it is rational to give up on instrumental rationality and become someone who can be governed by norms, at least when others are similarly governed by them.

Moehler (2014, 2018) argues that we can avoid some of these problems by conceiving the social contract as having multiple levels. The first level is basic agreement on social norms and rules of justice and morality. This first level is composed of rules that are the "primary source for regulating social behavior because these social, moral rules are tailored specifically to a group's moral history and its conditions of social cooperation" (Moehler 2014: 443). The norms at this first level satisfy the features we all see social and moral rules as having. The problem, according to Moehler, is that deeply pluralistic societies will disagree about these first-level rules and on whether and how much to punish violations of the rules. Any consensus on first-level rules is likely fragile and subject to dissolution as pluralism increases. As consensus on first-level norms strains under increased diversity, traditional contractual theories will, as I argued above, face something like the paradox of diversity. Moehler's (2014: 443) multilevel contract theory has a solution. He writes:

> that the traditional approach to morality reaches its limitations does not mean that moral theory per se reaches its limitations because if the members of deeply pluralistic societies have at least one end in common that they aim to reach despite their differing starting points and resultant conflicts, then the purely instrumental approach to morality applies.

Moehler argues that when consensus on first-level rules is undermined, citizens of a pluralistic society can ascend to a higher-level conception of morality and justice based on their shared commitment to instrumental rationality.

This is a thoroughly Hobbesian solution to the paradox of diversity that places the search for solutions to cases of conflict at the heart of the second-level contract. The principle of justice that this higher-level contract generates is what Moehler calls the "weak principle of universalization." This contractual model justifies this principle by using a bargaining model of agreement and arguing that rational agents in cases of conflict would apply the "stabilized Nash bargaining solution." The consensus-destroying features of dynamism and diversity are dampened here by finding a higher-level consensus immune from disruption. Call this an *insulation strategy* for dealing with the paradox of diversity.

This is an extremely elegant and straightforward way for a justice-centered, consensus-based contractual theory to avoid the implications of the paradox of diversity. The insulation strategy's primary danger is maintaining the firewall between the two levels of morality and justice. This danger goes in both directions, as the instrumental can become corrupted by other substantive moral views, and the substantive moral views may become more instrumental over time. Insofar as the former happens, second-level instrumental morality will not solve cases of conflict. Insofar as the latter happens, the two levels will collapse, and the concerns raised by Gaus and "biases and heuristic" theorists will become salient again.

None of these concerns is a decisive argument against the insulating solution, but they do help us see where the dangers of this strategy lie. In Section 5, I will look at a different approach—the *harnessing strategy*—before arguing why two major institutions of modern pluralistic societies will tend to undermine both attempts to save justice-first, consensus-based contractual theories from the paradox of diversity.

5. The Harnessing Strategy

Moehler's (2018) multilevel contract attempts to insulate the higher-level shared instrumental conception of morality and justice from everyday social life's diverse moral and political views. To mix metaphors, we can think of the higher-level, instrumental morality as a ship that sails on a turbulent sea. The motley crew on board share little in common except their interest in staying afloat, and this shared interest keeps them working together.

Rather than insulating the consensus on justice from diversity, disagreement, and change, it might also be possible to use these destabilizing elements product-ively. This is what Muldoon (2016) attempts to do. Rather than find an insulated point of view above the fray from which to construct a consensus on justice, Muldoon understands his consensus on justice as a "view from everywhere" that incorporates every point of view in society. The key innovation here is the recognition that there are many different perspectives from which to view moral and political questions and each perspective will come with its own doxastic and evaluative commitments and comparative weightings.

For instance, a music lover's perspective will typically have beliefs about the value of music, the relative merits of various kinds of music, and perhaps the importance of supporting civic institutions such as the symphony or local music clubs. We can imagine an equally civic-minded and aesthetically involved perspective that loves film but is indifferent to music. From this perspective, support of symphonies and music venues is not as important as preserving and evangelizing in favor of film. Individuals may have multiple perspectives, and not all the values or beliefs of each perspective will disagree. Muldoon argues that a

diverse society should tolerate and incorporate each of the distinct perspectives in that society. The "view from everywhere" is constructed by giving each perspective a say.

Leaving aside the thorny problem of individuating each perspective, we can see this as a model of a diverse society that exists as a large-scale marketplace of ideas built on the Millian ideal that tolerating experiments in living comes with substantial benefits. Out of this diversity, there still need to come some standards that govern society; it will not make sense to let each perspective go its own way. To protect the toleration of diverse perspectives, a conception of justice that distributes stable rights is needed. To generate such a conception of justice, social contract theories, as we have seen, construct a privileged point of view in the contractual model to select some arrangement of principles, norms, and rights. Moehler's multilevel approach introduces a meta, shared point of view in cases of conflict that rises above the diversity of substantive views; how is a view from everywhere able to generate a similarly stable and shared point of view from which to select principles of justice?

Muldoon's solution is to use a bargaining model to generate a distribution of rights without privileging any point of view to act as a representative in the bargaining model. Recall that Moehler's theory constructs the instrumental point of view and then applies a bargaining model to select the specific principle of justice in the form of the weak principle of universalization, thereby pulling a Kantian rabbit out of a Hobbesian hat. Muldoon dispenses with the first step and includes each perspective as a party to the contractual bargain. This creates a problem, though, since very different perspectives will represent the bargaining situation and the rights in question differently. To solve this problem, we need a shared public metric that all the perspectives will reckon roughly the same way. Muldoon argues that the "price" of rights can act as a shared currency. With a common currency for their bargaining, diverse perspectives can bargain over rights. The result of this bargaining procedure will constitute the conception of justice in that society.

As with Moehler's response to the paradox of diversity, we have an ingenious response to the challenges that diversity poses to consensus-based theories of justice in Muldoon's contract theory. Rather than insulating the consensus against diversity, this approach harnesses diversity directly by rejecting the need for a shared point of view to generate a consensus on justice. Call this the *harnessing* strategy for dealing with the paradox of diversity.

One of the puzzles for the insulation strategy is how to keep the levels separated to preserve the conditions for consensus. The main problem for the harnessing strategy is how to preserve the consensus on justice in the face of changing perspectives. Accepting the spirit of the harnessing strategy, Muldoon embraces this dynamism and argues that the contractual procedure is one that can and will be entered into again and again. Consequently, the distribution of rights will

always be up for renegotiation and change as perspectives change. Although this might be troubling in one sense, a society dedicated to toleration and diversity would likely not be troubled by an ongoing negotiation about rights and justice.

The more significant challenge, however, is identifying where the consensus lies in this theory. Recall that the paradox of diversity is about the tension between identifying a discrete conception of justice while at the same time preserving the conditions for consensus. The harnessing strategy attempts to avoid the problem by having the conception of justice move with the consensus. As diversity changes the conditions for consensus, the conception of justice changes too. Diversity bends but does not break the link between justice and consensus.

However, the plausibility of this strategy depends on what we think of as the substance of the consensus. If the consensus is on the outcome of the bargaining procedure, then the harnessing strategy should work, but it is hard to see how this could be the basis of consensus. The key to the system is preserving toleration and the commitment to the "view from everywhere," which is the mechanism that generates consensus and underlies the justification of the outcome of the bargaining procedure. However, this tolerant, open consensus is liable to be undermined by increasing diversity and the possibility of a changing conception of justice.

There is an unstated assumption in the background that no perspective or coalition of perspectives would have enough bargaining power to impose their will sustainably on others. This project's overriding goal is to keep the experiments in living going. However, preserving openness is not explicitly part of the social contract's structural or substantive basis. It is an essential part of the system, but we cannot merely assume it without begging a very important question.

Muldoon seems to be relying, implicitly, on something like Moehler's model of a higher-level contract that preserves the possibility of the first-order dynamic bargaining model. Instead of committing to the instrumental value of peace and cooperation, Muldoon's meta-contractors are committed to the larger Millian project of experiments in living and the marketplace of ideas. However, many perspectives that we are likely to find in a diverse society do not embrace these ideals. If so, how can we assume that the system can remain dynamically stable?

The insulating and harnessing strategies for dealing with the paradox of diversity have their merits and challenges. To simplify, the main challenges are (i) preserving the priority of instrumental morality in cases of conflict and (ii) preventing any perspective (in Muldoon's theory) or faction in the first-level contract (in Moehler's theory) from establishing a firm consensus for their favored conception of justice.

In the rest of this chapter, I will propose a way around these problems in relation to the paradox of diversity. However, in so doing, I will reject the basic framework of justice-centered, consensus-based contractual theorizing. My aim is not to argue that Moehler or Muldoon's (or anyone's) justice-centered, consensus-based theory is *wrong*, but instead that there are specific problems that arise from a

justice-first, consensus-based approach. In a sense, the alternative that I will offer takes its starting cues from these other theories and advances essential elements of them in a new direction.

6. Rejecting the Justice-First, Consensus-Based Approach

Contractual theories necessarily involve abstraction and idealization. Indeed, every theory does, whatever its form, but abstraction is central to contract. Contract theories are models of political justification. All models are abstract representations, idealizations of some sort (Weisberg 2007a, 2007b). The distinctiveness of this or that contractual theory depends on the theorist's choices about what and how to represent, abstract, and idealize.

Virtually all contract theories consider the question of justice and political justification more generally in a vacuum divorced from institutional features of the world. This makes perfect sense if the idea is to develop a theory of justice that is maximally general and portable across different societies with distinct histories and cultures. Although Rawls (1971: 119) allows that his representative choosers "know the general facts about human society," "economic theory," and the "laws of human psychology," he does not elaborate on what these general facts and laws are. Given that Rawls wants to leave open whether the institutional implementation of justice as fairness results in property-owning democracy or democratic socialism (all other institutional structures are ruled out), he must think that what the representatives know in this respect is not very constraining.

Gauthier (1986: 84) is a little clearer on this point. The starting point of his theory is that "morality arises from market failure." By "market failure," he means the Pareto-suboptimal, though rational equilibria that arise from social dilemmas. Although Gauthier never makes this point clear, he must assume that the institutional conditions for markets exist in a state of "moral anarchy." This is a strong assumption. Gauthier presents a theory of morality and justice as purely artificial in the Humean sense; there is no natural undergirding for justice and morality aside from our instrumental rationality. Gauthier (1979, 1991) is not arguing that we are literally *Homo economicus* by nature; instead, his argumentative strategy is to show that morality is compatible with and justifiable by economic rationality.

This creates a puzzle. Rawls models his representative contractual agents as reasonable democratic citizens. Gauthier models his representative agents as economic agents. Neither, however, assumes that the background conditions of democratic governance or a market society rely on a thick set of conventions and institutions that are themselves potentially subject to strictures of contractual justification. Moehler partly avoids this problem by taking his contractors' first-order moralities and institutions as given. Muldoon cannot avoid this issue, since the source of diversity will be what the contractors bring to the table in the first

instance and their interactions over time, with at least some of that time being spent in markets, commerce, and politics.

Contemporary contractual theory starts from the recognition that modern societies are diverse in terms of their citizens' beliefs, lifestyles, and pursuits. Contractual theories aim to generate a conception of justice to regulate such a society, given its diversity. However, this ignores the sources of diversity in modern societies. There are many, but most are related in some way to the market and commercial institutions that organize daily life in modern societies as well as the contestatory political institutions of modern democratic governance.

In the market, individuals and firms aim to outdo one another in satisfying the wants of consumers. Innovation is the motor of competition in the market, and the general churn of creative destruction ensures that the market economy is constantly in flux. However, we miss something important if we think of the economy as only about trade and profit. Commerce, employment, and trade affect virtually all aspects of modern societies: our lifestyles, food, entertainment, education, and travel. All the things that shape our conceptions of ourselves are profoundly influenced and shaped by markets. Critics of capitalism lament this fact, envisioning solidaristic communities driven by shared values or beliefs rather than relentless market change. We need not reject capitalism to admit that its critics inadvertently compliment it by recognizing its inherent dynamism.

Political institutions are not directly analogous to markets, since collective choice differs in many essential ways from individual exchange. Nevertheless, there is a sense in which a free and democratic political system harnesses and directs individual choices in largely mutually advantageous ways. The market's invisible hand may not have any direct analog in democratic politics. However, the discipline of the electoral system does share some similarities with the discipline of the market. Whether this creates results that generally conduce toward the common good is an open question, but it is responsive to the preferences and interests of the voters.[6]

Another feature that both well-functioning and mature market and democratic institutions share is that they do not prevent the entry of competitors. In the language of Douglass North, John Wallis, and Barry Weingast (2009), they are "open-access orders." Stability in these systems is assured by preserving openness and maintaining dynamic disequilibrium, not by fixing the conditions of the system in a constant state. In a democracy, the greatest threat is that one political party or leader would prevent contestation through elections or disqualify opposition political parties or candidates. In markets, one of the greatest threats is

---

[6] There is considerable debate on whether democratic institutions are "efficient" in a market-like way. Donald Wittman (1989, 1995) argues that they are, while Daron Acemoglu (2003) and others (for different reasons) argue they are not. See also Acemoglu and Robinson (2006), Caplan (2011), Somin (2016), and Achen and Bartels (2016).

monopolies and anti-competitive behavior, where the current winners in the market prevent competitors from entering. In both cases, the institutions work best when there is open access to new entrants who can compete with those currently in the system.

The key feature of a liberal and open society—one characterized by markets and democracy—is that it is dynamic. Like a functioning market process, the political and social institutions are always in disequilibrium at any given time. This is partly because markets are so central to the political and social institutions of a dynamic liberal society. If the goal of the open society as such is to preserve openness that will ensure change and disruption, thinking about the justification of any such system in terms of a fundamental justification of basic principles of justice that are meant to be insulated from the diversity and change that the system will create is likely to be misleading at best and counterproductive at worst.

Traditional contractualist theories attempt to justify high-level principles of justice that order society and authorize coercion. If we think of society as an interactive system, we can think of these principles as fixed points or stable equilibria. If, however, liberal society as a whole is, like the market, always in the process of disequilibrium, then thinking of set principles of justice as either a normative or justificatory ideal is likely to be both misleading and counterproductive. It needs to be more accurate, because the system is dynamic and not characterized by equilibrium stability, and counterproductive, because aiming at such equilibrium states could undermine the real benefits of the system, namely its dynamism and responsiveness.

We are faced again with the problem of the paradox of diversity and dynamic diversity. However, now the problem is amplified by assuming the existence of market and democratic institutions. In this context, insulation will be hard to maintain, but in any case, it is only needed if we accept the need for a consensus on principles of justice. The thought here is that justice is an outgrowth of various institutional structures rather than a higher-level principle that structures those institutions. We can avoid diversity's problems for a consensus-based theory of justice by rejecting the "justice-first" approach to contract theory. As I will argue in Section 7, doing so allows us to embrace and harness fully the diversity of modern open societies while also embracing some of the justificatory aims of traditional contract theory.

7. Dynamic Contractualism

An attempt to develop a dynamic contractual theory that rejects the justice-centered approach will differ considerably from what we find in traditional

contractual theories, since it will not aim at principles of justice in the Rawlsian sense or at the idea of a good society as one well-ordered by those principles. Because of this, it does not require a consensus on those principles for its justification, which makes public justification in the traditional sense unnecessary. This does not mean, however, that no basis of agreement is necessary for the dynamic approach. The contractualist project of justifying the basic rules of society through agreement is still compelling. What changes when we reject consensus-based, justice-centered theorizing is the *target* of justification, not its general form. However, changing the target does change many of the features that we associate with the contract theories of Rawls (1971) or Gauthier (1986), as we will see.

Capitalism and democracy are central to modern societies, as I have argued above. Does either system require a consensus on principles of justice to work? The answer must be no for a straightforward reason: Both institutional forms evolved without a clear consensus on principles of justice. This point is evident for democracy, especially American democracy. Whatever principles American democracy was founded on were not recognizably principles of justice in the Rawlsian sense. Most were excluded from democratic participation, and many were seen as the property of others. In no sense could the American political constitution, at least as it was first implemented and interpreted, be seen as just in something like a Rawlsian sense.

Nevertheless, American democracy, over several centuries, has become increasingly inclusive and open. This should be puzzling to the justice-centered theorist. Consensus on principles of justice was not needed for fundamental rights to be respected for all and to expand over time. There are constitutional limits to democratic power, but these limits are not principles of justice in the traditional sense. The same case can be made for capitalism. One need not have a thoroughly progressive notion of history to admit that on most dimensions concerning human welfare, capitalism and democracy have generally been forces for good. They have also gotten more humane and inclusive over time.

In what sense is justice the "first virtue" of social institutions if justice is not needed for these intuitions to evolve in positive directions? The suggestion here is that a consensus on principles of justice is neither necessary nor sufficient to undergird or generate an open, dynamic society. This does not mean that justice has no role in an open and dynamic society. We can agree with Rawls (1971: 3) that "Each person possesses an inviolability founded on justice that even the welfare of society as a whole cannot override," without thinking that the social order should aim at or be founded on a conception of justice. Or so I will argue.

If justice is not the target of justification, what is? Recall that the problem of the paradox of diversity is a tension between generating agreement and maintaining

consensus on a fixed conception of justice. The assumption is that society must endorse and comply with a single conception of justice. In this way, high-level principles of justice constrain the institutional environment and the decisions of public persons within the society. In Rawls's later work, the idea of public reason relates justice directly to public discourse and action. We can think of this as a "macro" idea of justice. Here justice acts as the "public moral constitution," as Rawls (1980: 539) described it.

Another approach, however, what we might call a "micro" idea of justice, could focus on the basic norms that make cooperative social life possible. Some notions of property, contract norms, reciprocity, etc. are necessary for well-functioning market societies to get off the ground. However, theorists going back to at least Hume (1739) have shown that, with increasingly rigorous demonstrations, these norms can and do arise as conventions (e.g. Lewis 1969; Vanderschraaf 2018). The contractual model can be used to test these norms, helping us see whether they are justified and how they might be improved in characteristic circumstances (Curry et al. 2019). In this sense, the social contract becomes an equilibrium selection mechanism that evaluates potential or existing norms and institutions rather than a mechanism for generating high-level principles that organize those norms and institutions. While the system as a whole does not converge on a single equilibrium or fixed point of justice, the dynamic process of a liberal society can be considered a distributed series of mini-games with conventional equilibria in each instance. The process of dynamic stability is the transition from each mini-game to the next.

Consider this simple nested, conflictual coordination game (Table 1). Even in this highly simplified version of a situation with two model contractors, we can already see the complexities of generating a social contract as a dynamic series of

Table 1 Nested Conflictual Coordination Game.

	A	B	C	D	E	F
I	8,8	0,8	0,0	0,0	0,0	0,0
II	8,0	6,6	0,0	0,0	0,0	0,0
III	0,0	0,0	5,3	0,0	0,0	0,0
IV	0,0	0,0	0,0	3,5	0,0	0,0
V	0,0	0,0	0,0	0,0	2,2	0,1
VI	0,0	0,0	0,0	0,0	1,0	1,1

mini-games. This game has six pure-strategy equilibria and fifty-seven mixed-strategy equilibria. To put it differently, there are sixty-three possible conventions or agreements from which to choose. Whichever convention is selected, however, the parties to the convention or contract will need to be able to justify rationally their selection of a particular strategy that leads to a convention. This project of justification is the substance of most social contract theories.

The difference between traditional justice-centered, consensus-based contractual theories and the dynamic, norm-based, micro-theory as it is being developed here is that there needs to be a set of high-level principles meant to order or select the equilibria in each of the mini-games. However, there is still an appropriate target for public justification. Rather than at principles of justice, justification in the dynamic liberal society aims at basic constraints or limiting principles on the system that are applied at the micro-level. These norms and institutional constraints build up across society as a whole and likely exhibit macro-convergence tendencies. The norms and institutional constraints form the basis of social organization and, as such, may look like implementations of high-level principles. However, those principles are just an explication of the lower-level norms and constraints. The targets of justification are more like structural features of the society than substantive or regulative ideals.

In addition to the substantive norms of property, reciprocity, rights, etc., this dynamic contractual theory will need basic limiting principles for those rights and norms. At its foundational level, the limiting principles preserve the openness that allows for the development of additional conventions and norms, especially those of a market economy, and democratic government guarantees openness. This is not a principle of morality or justice but a meta-principle that limits public institutional forms or principles that limit access. In a liberal, open society, this will mean, minimally, that individuals are mainly treated impersonally by basic institutions of society. Functionally, this is equivalent to a procedural conception of equality and a basic conception of personal freedom, both of which follow from a rejection of natural authority. These limiting principles operate similarly to Hart's (1961) rule of recognition with regard to other norms and rights. Although there is no space here to elaborate on this point in more detail, these limiting principles are not justified directly in the contractual model. Instead, they are the constitutive conditions of any contractual theory.

This approach avoids the paradox of diversity that afflicts justice-centered, consensus-based contractual theories by not relying on consensus at all. We should not mistake this as the claim that dynamic liberalism is opposed to justice or liberal norms and principles. The claim here is that these principles of justice or liberal norms are neither necessary nor sufficient for preserving the openness and maintaining a dynamic liberal society. Nevertheless, a dynamic liberal society will be recognizably liberal, but it is liberal because it is open and dynamic, not open and dynamic because it is liberal. Liberal principles of justice—or something like

them—will emerge out of the structural features of the dynamic liberal society; they are not the basis of that society. The upshot is that there is no need to generate a contractual consensus on basic principles of justice to institute a dynamic liberal society.

As we have seen, the fundamental question of justification remains, even in this decentralized, norm-based, and realist approach.[7] It is, however, relocated to the basic norms and institutions that undergird an open society. The contractual method can be used here to generate justification. Since consensus on basic principles of justice and morality is not required, there is no reason to normalize or constrain the diversity of the parties' reasoning to the contract (N) in the contractual model (M). Each may endorse the norms and institutions of their society from a different point of view or perspective. So long as they converge on the same equilibria, it is irrelevant whether their reasoning is conventional (Hardin 2003; Sabl 2012), instrumental (Moehler 2018), moral (Harman 1975; Harsanyi 1982; Gauthier 2013a and 2013b), or some combination (Gaus 2011). This is because the micro-norms and institutions, as well as the limiting principles, are not themselves a conception of justice but merely the foundational limits of a constitutional order. Because of this, consensus in the Rawlsian sense is not essential. Unlike traditional social contract theories, the diversity of non-public conceptions of morality or justice does not threaten agreement on the basic limiting principle and micro-norms.

8. Conclusion

I have argued that justice-centered, consensus-based theories of justice face a problem in the form of the *paradox of diversity*, especially regarding *dynamic diversity*. In response to this problem, I have argued that we should reject the goal that political contractualism should be in the business of justifying principles of justice through a consensus. Instead of a static consensus on principles of justice, we should see the contractual model as justifying background micro-norms and institutions that create the conditions for a liberal, open society. A consensus on a liberal conception of justice, I have argued, is neither necessary nor sufficient for generating a liberal social order.

The positive account of an alternative approach to political contractualism presented here is only a starting point. Especially given its novelty and differences from traditional justice-centered, consensus-based contractualism, any such account requires considerable argument before we can see it as a genuine

---

[7] The realism of this approach should be obvious, even though it is not explored in detail here. See Hankins and Thrasher (2022) for an account of political realism in the same spirit as the proposal developed here.

competitor. The goal was merely to show that a rejection of justice-centered, consensus-based theorizing is also a rejection of traditional forms of political contractualism. The project for theorists attracted to the basic justificatory standards of contractualism will be to develop and refine contractual theories compatible with open, dynamic societies. Doing so, I have argued, will require a wholesale rethinking of what contractual theories do and how they should do it.

References

Acemoglu, Daron. 2003. "Why Not a Political Coase Theorem? Social Conflict, Commitment, and Politics," *Journal of Comparative Economics* 31/4: 620–52.

Acemoglu, Daron, and James A. Robinson. 2006. *Economic Origins of Dictatorship and Democracy*. Cambridge: Cambridge University Press.

Achen, Christopher H., and Larry M. Bartels. 2016. *Democracy for Realists: Why Elections Do Not Produce Responsive Government*. Princeton, NJ: Princeton University Press.

Buchanan, James, and Gordon Tullock. 1962. *The Collected Works of James M. Buchanan*, iii: *The Calculus of Consent: Logical Foundations of Constitutional Democracy*. Indianapolis, IN: Liberty Fund Inc.

Caplan, Bryan. 2011. *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton, NJ: Princeton University Press.

Cohen, G. A. 2008. *Rescuing Justice and Equality*. Cambridge, MA: Harvard University Press.

Curry, Oliver Scott, Matthew Jones Chesters, and Caspar J. Van Lissa. 2019. "Mapping Morality with a Compass: Testing the Theory of 'Morality-as-Cooperation' with a New Questionnaire," *Journal of Research in Personality* 78: 106–24.

D'Agostino, Fred, Gerald Gaus, and John Thrasher. 2021. "Contemporary Approaches to the Social Contract," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Winter 2021 edn), https://plato.stanford.edu/entries/contractarianism-contemporary/, accessed 2 September 2023.

Dworkin, Ronald. 1976. "The Original Position," in Norman Daniels, ed., *Reading Rawls: Critical Studies on Rawls' 'A Theory of Justice'*," 16–52. Stanford, CA: Stanford University Press.

Enoch, David. 2013. "The Disorder of Public Reason: A Critical Study of Gerald Gaus's The Order of Public Reason," *Ethics* 124/1: 141–76.

Estlund, David. 2019. *Utopophobia*. Princeton, NJ: Princeton University Press.

Gaus, Gerald. 1996. *Justificatory Liberalism: An Essay on Epistemology and Political Theory*. New York: Oxford University Press.

Gaus, Gerald. 2011. *The Order of Public Reason: A Theory of Freedom and Morality in a Diverse and Bounded World*. Cambridge: Cambridge University Press.

Gaus, Gerald. 2017. "Is Public Reason a Normalization Project? Deep Diversity and the Open Society," *Social Philosophy Today* 33: 27–52.

Gauthier, David. 1979. "Thomas Hobbes: Moral Theorist," *Journal of Philosophy* 76/10: 547–59.

Gauthier, David. 1986. *Morals by Agreement*. Oxford: Clarendon Press.

Gauthier, David. 1991. "Why Contractarianism?" in Peter Vallentyne, ed., *Contractarianism and Rational Choice: Essays on David Gauthier's Morals By Agreement*, 15–30. Cambridge: Cambridge University Press.

Gauthier, David. 2013a. "Achieving Pareto-Optimality: Invisible Hands, Social Contracts, and Rational Deliberation," *Rationality, Markets and Morals* 4/78: 191–204.

Gauthier, David. 2013b. "Twenty-Five On," *Ethics* 123/4: 601–24.

Gigerenzer, Gerd. 2010. *Rationality for Mortals: How People Cope with Uncertainty*. New York: Oxford University Press.

Gjesdal, Adam. Forthcoming. "Agnosticism and Pluralism about Justice," *Journal of Ethics and Social Philosophy*.

Hankins, Keith, and John Thrasher. 2022. "Hume's Politics and Four Dimensions of Realism," *Journal of Politics* 84/2: 1007–20.

Hardin, Russell. 2003. *Liberalism, Constitutionalism, and Democracy*. New York: Oxford University Press.

Harman, Gilbert. 1975. "Moral Relativism Defended," *Philosophical Review* 84/1: 3–22.

Harsanyi, John. 1982. "Morality and the Theory of Rational Behavior," in Amartya Sen and Bernard Williams, eds, *Utilitarianism and Beyond*, 39–62. Cambridge: Cambridge University Press.

Hart, H. L. A. 1961. *The Concept of Law*. Oxford: Clarendon Press.

Hobbes, Thomas. 1651. *Leviathan*, ed. Noel Malcolm (2012 edn). Oxford: Clarendon Press.

Huemer, Michael. 2013. *The Problem of Political Authority: An Examination of the Right to Coerce and the Duty to Obey*. Basingstoke: Palgrave Macmillan.

Hume, David. 1739. *A Treatise of Human Nature*, ed. David Fate Norton and Mary J. Norton (2005). Oxford Philosophical Texts. New York: Oxford University Press.

Kahneman, Daniel. 2003. "Maps of Bounded Rationality: Psychology for Behavioral Economics," *American Economic Review* 93/5: 1449–75.

Kahneman, Daniel, and Amos Tversky. 1979. "Prospect Theory: An Analysis of Decision under Risk," *Econometrica* 47/2: 263–92.

Kant, Immanuel. 1797. *The Metaphysics of Morals*, ed. Mary Gregor (2012). Cambridge: Cambridge University Press.

Kogelmann, Brian, and Stephen Stich. 2016. "When Public Reason Fails Us: Convergence Discourse as Blood Oath," *American Political Science Review* 110/4: 717–30.

Lewis, David. 1969. *Convention: A Philosophical Study*. Cambridge, MA: Harvard University Press.

Locke, John. 1681. *Two Treatises on Government*, ed. Peter Laslett. 1989. Cambridge: Cambridge University Press.

Moehler, Michael. 2014. "The Scope of Instrumental Morality," *Philosophical Studies* 167/2: 435–51.

Moehler, Michael. 2018. *Minimal Morality: A Multilevel Social Contract Theory*. Oxford: Oxford University Press.

Muldoon, Ryan. 2016. *Social Contract Theory for a Diverse World: Beyond Tolerance*. New York: Routledge.

North, Douglass C., John Joseph Wallis, and Barry R. Weingast. 2009. *Violence and Social Orders: A Conceptual Framework for Interpreting Recorded Human History*. Cambridge: Cambridge University Press.

Nozick, Robert. 1974. *Anarchy, State, and Utopia*. New York: Basic Books.

Rawls, John. 1971. *A Theory of Justice* (rev. edn). Cambridge, MA: Belknap Press.

Rawls, John. 1980. "Kantian Constructivism in Moral Theory," *Journal of Philosophy* 77/9: 51–72.

Rawls, John. 1996. *Political Liberalism*. New York: Columbia University Press.

Rousseau, Jean-Jacques. 1762. "On the Social Contract," in *Basic Political Writings*, 153–252, ed. and tr. Donald A. Cress (2011). Indianapolis, IN: Hackett Publishing.

Sabl, Andrew. 2012. *Hume's Politics: Coordination and Crisis in the "History of England."* Princeton, NJ: Princeton University Press.

Schmidtz, David. 2006. *The Elements of Justice*. Cambridge: Cambridge University Press.

Smith, Vernon. 2008. *Rationality in Economics: Constructivist and Ecological Forms*. Cambridge: Cambridge University Press.

Smith, Vernon, and Bart Wilson. 2019. *Humanomics: Moral Sentiments and the Wealth of Nations for the Twenty-First Century*. Cambridge: Cambridge University Press.

Somin, Ilya. 2016. *Democracy and Political Ignorance: Why Smaller Government Is Smarter* (2nd edn). Stanford, CA: Stanford University Press.

Southwood, Nicholas. 2019. "Contractualism for Us as We Are," *Philosophy and Phenomenological Research* 99/3: 529–47.

Spinoza, Benedict. 1677. *Theological-Political Treatise*, ed. Jonathan Israel, tr. Michael Silverthorne and Jonathan Israel (2007 edn). Cambridge: Cambridge University Press.

Sugden, Robert. 2018. *The Community of Advantage: A Behavioural Economist's Defence of the Market*. Oxford: Oxford University Press.

Suikkanen, Jussi. 2014. "Contractualism and the Conditional Fallacy," *Oxford Studies in Normative Ethics* 4: 113–37.

Thrasher, John. 2014. "Uniqueness and Symmetry in Bargaining Theories of Justice," *Philosophical Studies* 167/3: 683–99.

Thrasher, John. 2019. "Constructivism, Representation, and Stability: Path-Dependence in Public Reason Theories of Justice," *Synthese* 196/1: 429–50.

Thrasher, John. 2020. "Agreeing to Disagree: Diversity, Political Contractualism, and the Open Society," *Journal of Politics* 82/3: 1142–55.

Thrasher, John, and Kevin Vallier. 2015. "The Fragility of Consensus," *European Journal of Philosophy* 23/4: 933–54.

Tversky, Amos, and Daniel Kahneman. 1986. "Rational Choice and the Framing of Decisions," *Journal of Business* 59/4: S251–S278.

Vallier, Kevin. 2018. "Public Justification," in Edward N. Zalta, ed., *The Stanford Encyclopedia of Philosophy* (Spring 2018 edn), https://plato.stanford.edu/archives/spr2018/entries/justification-public/, accessed 2 September 2023.

Vanderschraaf, Peter. 2018. *Strategic Justice: Convention and Problems of Balancing Divergent Interests.* New York: Oxford University Press.

Van Schoelandt, Chad. 2020. "Functionalist Justice and Coordination," *Social Theory and Practice* 46/2: 417–40.

Weisberg, Michael. 2007a. "Three Kinds of Idealization," *Journal of Philosophy* 104/12: 639–59.

Weisberg, Michael. 2007b. "Who Is a Modeler?" *British Journal for the Philosophy of Science* 58/2: 207–33.

Wittman, Donald. 1989. "Why Democracies Produce Efficient Results," *Journal of Political Economy* 97/6: 1395–424.

Wittman, Donald A. 1995. *The Myth of Democratic Failure: Why Political Institutions Are Efficient.* Chicago: University of Chicago Press.