American Political Science Review (2021) 115, 3, 967–981
doi:10.1017/S0003055421000101

Cooperative Capacities of the Rational: Revising Rawls's Account of Prudential Reasoning

JACQUELINE BASU *Stanford University*

John Rawls characterizes political rationality as narrowly self-regarding and therefore incapable of motivating political other-regard, self-moderation, or cooperative behavior. He ascribes these cooperative properties solely to reasonable, or principled, reasoning. This article evaluates Rawls's account of rationality by investigating his characterization of the democratic modus vivendi, which builds upon this account: Rawls asserts that the democratic modus vivendi is inherently unstable because it lacks the cooperative properties of the reasonable. These critiques entail positive claims about rational democratic equilibrium that are contradicted by formal accounts of self-enforcing democracy. The article demonstrates that the democratic modus vivendi can achieve robust stability because the rational can express the cooperative properties that Rawls reserves to the reasonable. By working within Rawls's seminal account of political reasoning to revise the properties he ascribes to rationality, this article offers a novel motivation for theoretical engagement with the rational and its role in political cooperation.

John Rawls parses political reasoning into two distinct modes: the *reasonable* and the *rational*.[1] These modes are fundamentally differentiated by their grounds: the reasonable derives from moral principles of political justice; the rational, calculations of private advantage.[2]

Given this defining difference, Rawls assigns distinct properties to each mode of political reasoning. Rawls presents the following three properties as unique to the reasonable: (1) **political other-regard**, or the long-term prioritization of polity-oriented considerations (Rawls, 48–49 n. 1); (2) **self-moderation**, or the ongoing constraint of narrowly self-regarding ends (Rawls 1999, 25); and (3) the practice of "**virtues of political cooperation**," a suite of other-regarding, self-moderating behaviors that includes compromise, toleration, and civility (Rawls, 157, 194). In direct contrast, Rawls characterizes the rational as incapable of reliably prioritizing polity-regarding considerations, moderating the pursuit of self-regarding ends, or motivating behavior that is functionally equivalent to the cooperative virtues (see, e.g., Rawls, 48–54; 147–150; Rawls 1999, 149–152). As such, Rawls excludes the rational from explanations of political other-regard, self-moderation, and cooperative behavior.

Rawls claims broad validity for this schema, asserting that "no sensible view can possibly get by without the reasonable and rational as I use them" (Rawls, 380). Within his own framework, this account of political reasoning plays a foundational role: his accounts of democratic consensus and stability draw heavily on this characterization of the reasonable and the rational (Rawls, 47–48).

This paper accepts Rawls's basic distinction between the reasonable and the rational, grounding the former in moral principles of political justice, the latter in calculations of private advantage. However, it contests Rawls's account of the properties assigned to each mode. Specifically, it demonstrates that the rational can express the three properties that Rawls reserves to the reasonable: rational agents can reliably prioritize polity-regarding considerations, moderate their pursuit of narrowly self-regarding ends, and engage in cooperative political behavior that is functionally equivalent to the cooperative virtues of the reasonable.

Thus, this paper revises Rawls's typology of rational and reasonable, allowing for the former to express desirable properties that Rawls presents as unique to the latter. It reflects on the implications of this revision for Rawls's accounts of democratic consensus and stability, which build on this framework of political reasoning. More broadly, the argument encourages robust theoretical engagement with the rational and its capacity to motivate political other-regard, self-moderation, and cooperative behavior. The paper concludes by outlining an agenda for further work in this vein.

Properties of the Democratic Modus Vivendi

To assess Rawls's typology of the rational and the reasonable, I evaluate his application of this framework within his account of democratic consensus. Rawls identifies two primary forms of democratic consensus, differentiated by the reasoning that grounds citizens'

Jacqueline Basu, Postdoctoral Research Fellow, Department of Political Science, Stanford University, jbasu1@stanford.edu.

Received: January 06, 2020; revised: January 22, 2021; accepted: February 15, 2021. First published online: March 15, 2021.

[1] This paper primarily addresses the account of the reasonable and rational that Rawls presents in *Political Liberalism* (Rawls 2005, 48–54). Subsequent references to *Political Liberalism* (Rawls 2005) cite only author and page number, omitting publication year.

[2] Following Rawls, I refer to rational motivations using a cluster of related terms, including "ends[,]… interests[,]" and "rational advantage" (respectively, Rawls, 50, 16).

consent: the modus vivendi, grounded solely in the rational, and the overlapping consensus, founded in the reasonable (Rawls, 147–150). Rawls's characterization of each form of democratic consensus reflects the properties he ascribes to its grounding reasoning. Thus, I assess Rawls's characterization of the democratic modus vivendi, and in doing so, I evaluate the account of rationality underlying this characterization.

Drawing on his account of the rational, Rawls attributes three primary failings to the democratic modus vivendi (Rawls, 147–148): Rawls claims that (1) rational democratic consensus is **grounded** in a mode of reasoning that prioritizes self-regarding ends and cannot reliably promote other-regarding, self-moderating considerations; (2) the **object** of rational citizens' consent is limited to the institutional features of democracy and does not entail a reliable commitment to uphold these institutions through practices of cooperative citizenship; therefore, (3) the rational democratic consensus is inherently **unstable** under a changing distribution of political power within the polity.

I demonstrate that Rawls's concerns rest on a series of positive claims about the incentives that motivate democratic compliance and stability, as well as the priorities and behavior promoted by self-interested reasoning. Accordingly, I assess these positive claims by drawing on formal accounts of self-enforcing democracy, which articulate the conditions necessary to maintain stable democratic consensus founded in citizens' self-interest (e.g., Fearon 2011; Przeworski 1991; Weingast 1997a).

These formal accounts counter Rawls's criticisms of the democratic modus vivendi, revealing that (1) rational citizens can reliably prioritize polity-regarding concerns and moderate self-regarding ones; (2) these rational citizens help sustain democratic equilibrium by choosing—as reasonable citizens do—to participate in costly practices of cooperative citizenship; and (3) a rational democratic consensus can achieve robust stability under a changing distribution of power. Thus, the democratic modus vivendi can escape Rawls's critiques, and it has this capacity because the rational can express other-regarding, self-moderating, and cooperative properties that Rawls reserves to the reasonable. As such, this argument challenges Rawls's typology of the rational and reasonable by contesting the properties he ascribes to each mode.

Rawls's characterization of the rational and its role in democratic participation has been widely echoed: even accounts critical of Rawls's adopt his dismissive stance toward political agreements founded in self-interest (e.g., Horton 2011, 122–123; Wendt 2016, 360).[3] This paper is the first to directly and comprehensively assess the positive claims on which Rawls builds his critique of rational democratic consensus. As such, it offers a novel basis for evaluating his underlying account of political rationality. In sum, this paper works within Rawls's seminal framework to promote further theoretical engagement with the rational, focusing attention on its capacity to motivate cooperative political attitudes and behaviors that Rawls reserves to the reasonable.

Rational Democratic Consensus within a Rawlsian Framework

This argument works within the conceptual framework established in *Political Liberalism.* It adopts (a) Rawls's *definitions*, including his basic distinction between the grounds of rational and reasonable evaluation and his minimal account of the modus vivendi as a self-interested democratic consensus, and (b) his *values*: just, stable political institutions and habits of cooperative citizenship (Rawls, 47–48). It demonstrates the validity of the democratic modus vivendi according to this rubric of values and, in turn, demonstrates that the rational can express properties that Rawls reserves to the reasonable.

Theoretical work has been done to investigate cooperation motivated by rational calculation (e.g., Elster 2006; Gauthier 1986), interrogate the distinction between rational and reasonable (Elster 2009), and develop a democratic framework in which citizens' participation is motivated by rational calculation rather than adherence to liberal principles (Ober 2017). However, these accounts do not work within the Rawlsian framework to assess Rawls's own schema of rational and reasonable or evaluate the characterization of rational democratic consensus he builds on this basis.

Rawls's account of democratic consensus *has* drawn much direct criticism, and theories of modus vivendi have a number of proponents. Modus vivendi accounts can be categorized into (1) critiques of the overlapping consensus, (2) accounts that ascribe normative content to the modus vivendi but expand its grounds of consensus beyond self-interest, and (3) practical arguments that self-interest can secure stable democratic institutions. Of these, none comprehensively addresses Rawls's positive claims about the features of a rational democratic consensus or, in turn, reflects more broadly on Rawls's account of political reasoning.

Rawls's ideal of the overlapping consensus has inspired many critiques.[4] Accordingly, many modus vivendi accounts are framed in this light, critiquing Rawls's ideal by presenting the modus vivendi as a preferable object of study. Dauenhauer (2000) presents "A Good Word for a Modus Vivendi" to diagnose the

[3] I focus on Rawls because his framework on political reasoning and democratic consensus has informed much subsequent work, in keeping with his immense influence on contemporary political thought (Forrester 2019). However, others also dismiss self-interested democratic consensus as untenable (e.g., Larmore 1990, 346).

[4] Chung (2019; 2020), Kogelmann and Stich (2016), and Thrasher and Vallier (2013) demonstrate the instability of the overlapping consensus by developing formal models of deliberation. Dryzek and Niemeyer (2006) argue that the overlapping consensus does not adequately accommodate pluralism (see also Thrasher and Vallier 2018). Talisse (2003) asserts that the overlapping consensus either depends unduly on comprehensive moral doctrines or reduces to a modus vivendi.

instability of the overlapping consensus. Gray (2000) and McCabe (2010) argue that the overlapping consensus cannot sufficiently accommodate value-pluralism and conclude that the modus vivendi offers a preferable focus for liberal theory. Realist scholarship promotes the modus vivendi as a rejection of the "liberal moralist" tradition associated with Rawls (Galston 2010; Horton 2010; Williams 2005). They present the modus vivendi as an alternative to the Rawlsian framework rather than engaging with his particular concerns about rational democratic consensus.

Numerous accounts highlight the modus vivendi's normative content, articulating the values that theories of modus vivendi might pursue. Modus vivendi liberalism (Gray 2000; McCabe 2010) presents the modus vivendi as an avenue toward the liberal value of pluralism. Other modus vivendi theorists promote values like peace (Wendt 2019) or political acceptability (Wall 2019), presenting these as distinct from Rawls's core values but complementary to liberal ends (see also, Sala 2019). Realist accounts of the modus vivendi pursue values external to the liberal project, such as managing political conflict (Horton 2011) or securing political survival (Fossen 2019). However, these proponents of the modus vivendi relax Rawls's definition of its grounds: rather than assuming a narrow foundation in self-interest, they broaden its basis to encompass a wide array of motivations, including culture, norms, and personal values, in addition to interest.[5] Because they modify this defining feature of Rawls's modus vivendi, these accounts cannot address his concerns about democratic consensus founded solely in self-interest or his underlying account of rationality.

Finally, few accounts address Rawls's claim that self-interested democratic consensus is unstable under a changing distribution of power. Hershovitz (2000) most directly evaluates this assertion: using historical example and formal intuitions, he argues that well-designed political institutions establish sufficient incentives to sustain rational democratic equilibrium. Arnsperger and Picavet (2004) argue in similar vein. However, these accounts do not address Rawls's concern that the rational citizens participating in this consensus fail to make choices and take actions that he considers "reasonable." Arnsperger and Picavet (2004) locate the self-interested democratic equilibrium outside the Rawlsian framework, whereas Hershovitz (2000) ultimately dismisses Rawls's concern about reasonableness, as he concludes that it is unnecessary for democratic stability.

This paper uses formal accounts of self-enforcing democracy to show that the democratic modus vivendi can achieve robust stability under a changing distribution of power and that rational citizens uphold this equilibrium by expressing cooperative qualities that Rawls reserves to the reasonable. In doing so, it evaluates Rawls's typology of political reasoning, revising his account of rationality to reflect its capacity to share these properties with the reasonable.

The argument proceeds as follows: the next section rehearses Rawls's account of political reasoning and democratic consensus. The following outlines the formal literature on self-enforcing democracy. The subsequent three sections argue that (a) the stability of the democratic modus vivendi is robust to a changing distribution of political power, (b) rational democratic equilibrium entails consensus on demanding habits of cooperative citizenship, and (c) rational evaluation can reliably prioritize polity-regarding considerations and moderate the pursuit of narrowly self-regarding ends. To conclude, I reflect on this argument's implications for Rawls's account of political reasoning and democratic consensus and outline avenues for further theoretical investigation of rationally motivated democratic cooperation.

RAWLS ON POLITICAL REASONING AND DEMOCRATIC CONSENSUS

This section (1) rehearses Rawls's account of the rational and the reasonable, then (2) outlines the distinct forms of democratic consensus founded on these distinct modes of reasoning, comparing Rawls's reasonable ideal, the overlapping consensus, with its rational foil, the modus vivendi. Finally, (3) it concludes with Rawls's description of how the overlapping consensus develops in practice, an account intended to demonstrate the feasibility of his ideal. The democratic modus vivendi plays a foundational role in this account, providing a basis from which the overlapping consensus evolves.

Political Reasoning: The Reasonable and the Rational

Rawls asserts that organized political cooperation pursues two distinct values: (1) *fairness*, or the coherence of the political framework with a political conception of justice, and (2) the promotion of citizens' *rational advantage* (Rawls, 16). Given these founding values, Rawls identifies two structural components of political cooperation: (1) terms of cooperation that establish "publicly recognized rules and procedures" and (2) citizens' widespread voluntary acquiescence to those institutions (Rawls, 16; see also, 38). Rawls assesses political institutions by their ability to promote both rational advantage and fairness and considers the fulfillment of this condition "implicit in the public culture of a democratic society" (Rawls, 15). Therefore, a democratic agreement satisfies his *institutional* ideal for organized political cooperation.

Using the same values of fairness and rational advantage, Rawls prescribes a basis for citizens' *voluntary acquiescence*. He specifies two modes of political reasoning that citizens might employ: the "reasonable" by which citizens evaluate political fairness and the "rational" by which citizens evaluate the private advantage they gain from cooperation (Rawls, 48–54). These

[5] See, e.g., McCabe (2010, 159), Sala (2019), Fossen (2019), Wall (2019), Wendt (2019; 2016), and Horton (2010), who expand the grounds of the modus vivendi beyond self-interest.

two modes are primarily distinguished by their grounds: the reasonable derives from moral principles of political justice (Rawls, 52, 147), the rational from private calculations of interest (Rawls, 50). Given this grounding distinction, Rawls assigns distinctive properties to each mode of reasoning and draws implications for how they function in a political context. I outline three central distinctions below.

First, Rawls ascribes distinct **priorities** to each mode of reasoning. The reasonable prioritizes the other-regarding concern that "each [citizen] benefits along with others," whereas the rational promotes self-regarding ends—the rational agent "seek[s] ends and interests peculiarly its own" (Rawls, 50).[6] In a political context, then, the reasonable places a high priority on other-regarding ends, extending this regard to all polity members. By contrast, the rational places overriding priority on narrowly self-regarding ends: rational agents form factions, "each of which has its own fundamental interest distinct from and opposed to the interests of the other groups" (Rawls 1999, 150).

Second, Rawls asserts that reasonable other-regard tempers rational self-regard, resulting in **self-moderation**. That is, citizens balance self- and other-regarding priorities by exercising these "distinct and independent" modes of reasoning in conjunction (Rawls, 51–52; see also, Rawls, 54). By employing both modes of political reasoning, citizens rationally pursue private ends, but "their rational conduct" is "constrained by their sense of what is reasonable," ensuring they attend to the effects of their actions on others (Rawls 1999, 25; see also, Rawls 2001, 82). Absent the reasonable, however, the rational lacks this capacity for reliable self-moderation (Rawls, 147–148; Rawls 1999, 150).

Third, Rawls distinguishes the rational and the reasonable by the quality of the **cooperative behavior** each motivates. Rawls asserts that the reasonable promotes longstanding engagement in a collection of cooperative political behaviors that he labels "political virtues" or "virtues of political cooperation" (respectively, Rawls, 194, 157). These encompass an array of cooperative practices and attitudes including "tolerance[,] … being ready to meet others halfway[,] … reasonableness[,] and the sense of fairness" (Rawls, 157). These cooperative virtues express the reasonable orientation toward other-regard and self-moderation: reasonable agents exercise other-regard by attending to the interests of their peers, "tak[ing] into account the consequences of their actions on others' well-being" (Rawls, 49 n. 1); they exercise self-moderation by "act[ing] on [fair] terms, even at the cost of their own interests in particular situations," tempering their pursuit of private ends to ensure that political cooperation adequately benefits their co-citizens (Rawls, xlii).

Rawls describes the reasonable commitment to the cooperative virtues as indicative of an enduring and general regard for all polity members. He attributes these qualities of reasonable cooperation to its principled foundation: the reasonable motivates an *enduring* commitment to the cooperative virtues because it derives from stable principles (Rawls, 148); this cooperative orientation *generalizes* to all polity members because it derives from principles that acknowledge others' moral personhood and therefore "recognize the independent validity of the claims of others" (Rawls, 52).

By contrast, Rawls does not believe the rational capable of motivating cooperative behaviors functionally equivalent to these cooperative virtues. Rawls acknowledges that the rational agent's private sense of the good is not confined to her own well-being (Rawls, 50–51). Given her "special loyalties or attachments," she may act to further the interests of her affiliates, perhaps even by prioritizing their well-being over her own (Rawls, 52 n. 6). However, this cooperative behavior is limited: whereas the cooperative virtues of the reasonable are enduring and reflect a general regard for others, rational cooperation is mutable and particularistic. *Mutable*, as the rational agent might cease cooperation if she reprioritizes her interests and other ends provide greater benefit (Rawls, 50); *particularistic*, as rational other-regard derives from special private attachments rather than a general attentiveness to the well-being of her co-citizens (Rawls, 52 n. 6).

Characterizing Consensus: Modus Vivendi as Foil

Rawls identifies two primary forms of democratic consensus, differentiated by the reasoning that motivates citizens' voluntary consent: (1) the *democratic modus vivendi* in which citizens exercise only the rational, evaluating the regime by its capacity to promote their interests, and (2) the *overlapping consensus* in which citizens consent on the principled basis of the reasonable.[7] Rawls claims that this difference in foundational reasoning creates three central distinctions between the resulting agreements (Rawls, 147–148). Each reflects a criticism of the democratic modus vivendi while casting the overlapping consensus in an ideal light.

The defining distinction between the overlapping consensus and the democratic modus vivendi lies in the *grounds of citizens' consensus.* Whereas reasonable

[6] Rawls acknowledges that rational priorities are not necessarily egoistic: the rational agent's private sense of the good might include a preference for the good of other individuals or groups (Rawls, 50–51). However, because rational agents develop other-regarding priorities based on "special loyalties or attachments," they do not reliably extend this concern, as reasonable citizens do, to all polity members (Rawls, 52 n. 6).

[7] Rawls acknowledges Baier's (1989) constitutional consensus as a third form of democratic consensus, intermediate between the modus vivendi and the overlapping consensus (Rawls 2005, 149). Like the overlapping consensus, the constitutional consensus derives from principle-based reasoning. This paper develops an account of democratic consensus grounded in prudential, rather than principled, reasoning. It engages with Rawls's overlapping consensus to characterize the values captured in Rawls's ideal. Beyond this, principle-based accounts like the constitutional consensus lie outside its scope.

citizens ground their consent in shared principles of political justice, rational citizens ground their consent in interest-maximizing calculation. By Rawls's definition of rationality, the democratic modus vivendi derives from a mode of reasoning that does not reliably prioritize polity-regarding considerations or moderate the pursuit of self-regarding ends. Accordingly, Rawls describes this rational basis for democratic consensus as a "fortunate convergence of interests," a product of "happenstance" rather than meaningful human decision making (respectively, Rawls, 147, 148).

Rawls's second critique of the democratic modus vivendi lies in the *object of democratic consensus.* Reasonable citizens consent not only to the terms of a democratic agreement but also to the shared conception of justice underlying that agreement. By contrast, Rawls asserts, rational citizens consent only to the *institutional* features of the democratic order: theirs is "merely a consensus on accepting certain authorities, or on complying with certain institutional arrangements" (Rawls, 147).

Finally, Rawls's claim about the *robustness* of democratic stability follows from the two preceding distinctions. Rawls believes that gains in political power corrode rational agents' incentives to uphold the democratic agreement. Therefore, Rawls concludes, "stability with respect to the distribution of power is lacking" in the democratic modus vivendi (Rawls, 148). By contrast, within the overlapping consensus, citizens are morally committed to fair cooperation. As such, shifting interests do not affect their commitment to democratic cooperation.

Evolution of Consensus: Modus Vivendi as Foundation

Finally, Rawls presents his overlapping consensus as a "realistic" ideal rather than a "utopian" vision (respectively, Rawls, 66, 158). He demonstrates its feasibility by describing the mechanism by which it might occur in practice (Rawls, 158–168). The democratic modus vivendi plays a foundational role in this process: the democratic framework is initially established as a modus vivendi, but it evolves into an overlapping consensus over time.

This transformation can occur because both forms of democratic consensus fulfill Rawls's *institutional* ideal of fair terms of cooperation—they simply differ in the reasoning that motivates citizens' consent. Thus, the democratic modus vivendi evolves toward an overlapping consensus as this grounding reasoning evolves from a rational mode to a reasonable one. Rawls credits democratic institutions with catalyzing this transformation of citizens' reasoning (Rawls, 163, 142–143). He concludes that an overlapping consensus can occur if democratic institutions established as a modus vivendi function "effectively and successfully for a sustained period of time," persisting long enough to introduce elements of the reasonable into citizens' political thought and behavior (Rawls, 163; see also, xli). This section rehearses Rawls's account of this process.

Citizens initially establish democratic institutions as a modus vivendi, calculating that democracy presents an interest-maximizing alternative to civil conflict (Rawls, 158–159). By establishing fair terms of cooperation, rational citizens implicitly "endorse an institutional structure satisfying a liberal political conception of justice" (Rawls, xxxviii). However, unlike reasonable actors, they do not accept fair terms *because of* the conception of justice they entail. Instead, each rational citizen is "ready to pursue their goals at the expense of the other, and should conditions change they may do so" (Rawls, 147).

To shift the modus vivendi towards an overlapping consensus, citizens must choose to express reasonable attitudes and behaviors. However, citizens only have reason to express the reasonable if they expect their peers follow suit:

> It is reasonable to expect everyone to endorse and act on [fair terms of cooperation], provided others can be relied on to do the same. If we cannot rely on [our co-citizens], then it may be irrational or self-sacrificial to act from those principles [of political justice]. (Rawls, 54)

Each citizen's decision rests on her expectations about the corresponding choices and behavior of her peers. As such, Rawls concludes, the prevailing rationality of the modus vivendi hinders expression of the reasonable: citizens know their peers to be motivated by interest rather than principle and therefore cannot rely on them to uphold the democratic agreement. Without these assurances, citizens lack motivation to express the reasonable.

Thus, the transformation toward an overlapping consensus hinges on the question of how these interdependent expectations and choices of reasonable behavior might be encouraged. Rawls claims that democratic institutions—if they are well-established, functional, and reliable—provide this catalyzing force:

> The basic political institutions incorporating [liberal] principles … *when working effectively and successfully for a sustained period of time* … tend to encourage the cooperative virtues of political life: the virtue of reasonableness and a sense of fairness, a spirit of compromise and a readiness to meet others halfway. (Rawls, 163, emphasis added)

By encouraging citizens to practice the cooperative virtues, effective democratic institutions introduce elements of the reasonable within the rational equilibrium of the modus vivendi. Further, well-established democratic institutions spark citizens' expectation of *mutual* reasonableness because "if other persons with evident intention do their part, people tend to develop trust in them" (Rawls, 163). By practicing cooperative behaviors, citizens assure others of their willingness to "do their part."

Therefore, Rawls claims, functional democratic institutions catalyze the widespread expression of the reasonable: by eliciting expression of the

cooperative virtues, they induce a virtuous cycle in which citizens reciprocate and reinforce each other's reasonable orientation. Over time, citizens' embrace of the reasonable expands to include a commitment to shared principles of justice (Rawls, 158–168). Thus, effective and successful democratic institutions—established as a modus vivendi—precede and produce the reasonable political culture that brings about the overlapping consensus.

MODUS VIVENDI AS SELF-ENFORCING DEMOCRACY

I challenge Rawls's characterization of a rational democratic consensus and, in doing so, revise his characterization of political rationality. I make this argument by evaluating Rawls's three central criticisms of the democratic modus vivendi:

1. *Robustness*: The stability of a rational democratic consensus is vulnerable to change in the distribution of political power: as citizens gain power, their interest in democratic participation diminishes.
2. *Object of Consensus*: Rational consent to democratic participation encompasses only the institutional and procedural features of democracy.
3. *Grounds of Consensus*: Rational democratic consensus is grounded in a mode of reasoning that lacks the other-regarding priorities and self-moderating capacity of the reasonable.

Each assertion (1) assumes democratic consent founded in self-interested reasoning and (2) makes a positive claim on this basis, whether about the priorities, choices, and behavior associated with self-interested reasoning or about the incentives that motivate democratic participation. Therefore, I evaluate each assertion by comparison with formal accounts of democratic equilibrium that (1) make the same assumption of democratic consent founded in self-interested reasoning and (2) describe the robustness, object, and grounds of democratic equilibrium under these conditions.

I draw upon the formal literature of self-enforcing democracy to make these comparisons. Congruent with Rawls's definition of the democratic modus vivendi, accounts of self-enforcing democracy assume democratic acquiescence motivated solely by self-interested reasoning. Przeworski asserts this standard assumption of self-enforcing democracy: "the only claim I am trying to substantiate is that a theory of democracy based on the assumption of self-interested strategic compliance is plausible and sufficient" (Przeworski 1991, 24; see also Mittal and Weingast 2013, 282–283; Weingast 2004, 162). Given this assumption, accounts of self-enforcing democracy describe the conditions necessary for rational democratic equilibrium. Therefore, they provide a means of evaluating Rawls's positive claims about the democratic modus vivendi and, in turn, his underlying assertions about the properties of rational evaluation.[8]

This section (a) reiterates key assumptions of Rawls's framework observed throughout the paper; (b) presents an overview of self-enforcing democracy; and (c) outlines the remainder of the argument, which evaluates Rawls's account of rational democratic equilibrium by comparison with these formal accounts of self-enforcing democracy.

Self-Enforcing Democracy within a Rawlsian Framework

This argument works within Rawls's framework of assumptions and values. Here, I reiterate four central features of this framework. First, citizens consent to the democratic modus vivendi on a solely prudential basis. Other motivations for democratic participation—for instance, shared principles, norms, or traditions—do not feature in this form of democratic consensus. Formal accounts of self-enforcing democracy share this minimal assumption of self-interested democratic consensus.

Second, the democratic modus vivendi and the overlapping consensus are differentiated by the reasoning that motivates citizens' participation: the former derives from self-interest, the latter, shared principles. By challenging Rawls's critiques of the modus vivendi, I do not suggest that it is *equivalent* to the overlapping consensus: this fundamental distinction in citizens' motivation remains.

Third, the overlapping consensus represents Rawls's ideal democratic equilibrium. I do not contest Rawls's characterization of this ideal or claim that the modus vivendi is preferable to it. I simply evaluate Rawls's critiques of the democratic modus vivendi, as enumerated above. My primary intent is to show that the rational democratic equilibrium entails self-moderating, other-regarding, and cooperative features that Rawls reserves to the reasonable and to argue that Rawls's account of the rational should be revised to reflect this.

Finally, the democratic modus vivendi provides an institutional foundation for the overlapping consensus: the evolution from modus vivendi to overlapping consensus occurs within a fixed institutional framework and simply reflects a transformation of citizens' motivations for consent. Further, this transformation can only occur if the democratic institutional framework

[8] Methodologically, accounts of self-enforcing democracy adopt a primarily formal approach, often supported by case studies. Thus, Weingast (2004) discusses democratization in Spain; Mittal and Weingast (2013) investigate institutional change in the early United States; and Przeworski (1991) addresses political and economic transitions in Eastern Europe and Latin America. This literature also engages with empirical accounts of democratic transition, consolidation, and stability, e.g., Linz and Stepan (1996), Diamond (1999), and Acemoglu and Robinson (2006). Reciprocally, the concept of self-enforcing democracy appears in empirical work as a goal for processes of democratization (Carugati 2019; Hyde and Marinov 2014).

"work[s] effectively and successfully for a sustained period of time" as a modus vivendi (Rawls, 163). Given this assertion of functionally equivalent and effective democratic institutions, I assume that the democratic modus vivendi, like the overlapping consensus, entails *consolidated* democratic institutions—institutions in equilibrium (Diamond 1999; Linz and Stepan 1996; Schedler 1998). The next subsection outlines the defining features of a rational democratic equilibrium.[9]

Defining Self-Enforcing Democracy

This section defines self-enforcing democracy. Subsequent sections evaluate Rawls's account of self-interested democratic consensus by comparison with the formal accounts rehearsed here.

A self-enforcing agreement has three basic components: (a) the agreement is in equilibrium, or stable; (b) the agreement's stability derives from the voluntary compliance of its participants—the *self*-enforcing nature of the agreement implies that no power external to the agreement is authorized to enforce it; and finally, (c) participants' voluntary compliance is motivated by calculations of private advantage (Telser 1980, 27). A pact is self-enforcing only if "participants all perceive that they are better off under the pact than under the status quo" (Weingast 1997a, 258; see also Przeworski 2006, 312). Thus, a central concern in the study of self-enforcing agreements is to identify the structure of incentives that motivates participants' voluntary cooperation, rendering the agreement stable (Mittal and Weingast 2013, 279).

A stable democratic pact is one form of self-enforcing agreement (Fearon 2011, 1661–1662; Przeworski 2006, 312; Weingast 2004, 173). In addition to the general mechanism of self-enforcement described above, the democratic agreement has particular institutional features. First, the pact establishes a **governing apparatus** authorized to enforce the terms of cooperation and constrain citizens' behavior. By accepting the pact, democratic citizens accept the coercive power it institutionalizes. Thus, the pact formalizes a **power hierarchy** within the polity: once the pact is established, citizens differ in their proximity to governing power, exerting varying degrees of control over the implementation and enforcement of the terms of cooperation. As such, participants in the democratic agreement can be differentiated, at any given time, into groups that are defined by their relative proximity to governing power. Accounts of self-enforcing democracy schematically model this power differential by dividing the democratic polity into **two discrete groups** on this basis: (a) a group "in power"—the parties and organizations that occupy or influence the government apparatus at a given time t—and (b) a group "out of power"—those with diminished access to governing power at time t (Przeworski 1991, 18–19; 2006, 312).[10] Finally, the **sharing and transfer of power** among citizens is a defining feature of democratic governance: citizens' relative proximity to power changes over time, given standard practices of democratic politics (Przeworski 1991, 10–14).

Given these features, a stable democratic pact must satisfy the conditions of a self-enforcing agreement: knowing that the pact institutes these structures of shared power and governance, citizens must have sufficient incentives to endorse and uphold it over time. Of particular concern is the ongoing voluntary compliance of the groups defined by their relative proximity to power: democratic stability requires that, at any given time, "those *in* power adhere to the constitutional rules … they obey election results and eschew transgressing the rights of their opponents," while also "no significant group of citizens or parties *out* of power is willing to attempt to subvert power or secede" (Mittal and Weingast 2013, 282, emphasis added). Self-enforcing democracy implies that citizens throughout the power hierarchy have sufficient incentives to voluntarily comply with the terms of the agreement.

Mittal and Weingast (2013, 285–286) outline four conditions necessary to establish these incentives, securing citizens' ongoing voluntary compliance:

1. **Rules and Limits:** The democratic agreement must institute "structure and process—citizens' rights and a set of rules governing public decision making," establishing "a series of limits on the state" (Mittal and Weingast 2013, 285).

[9] At times, Rawls ascribes characteristics to the democratic modus vivendi that violate the definition of democracy in equilibrium, e.g., when he claims that its rational citizens are "prepared to resist or to violate legitimate democratic law" to pursue factional ends (Rawls 1999, 150). However, such descriptors are incompatible with his account of a democratic modus vivendi that is sufficiently "effective, successful, and sustained" to engender citizens' trust in the efficacy and reliability of the democratic framework, initiating the evolution towards an overlapping consensus (Rawls, 163). This latter modus vivendi must be in equilibrium for the overlapping consensus to be achieved by the mechanism Rawls describes.

Thus, Rawls appears to conflate distinct scenarios under the same name, using the term "modus vivendi" to describe democratic institutions both in and out of equilibrium. It lies outside the scope of the paper to fully disambiguate Rawls's use of the term into distinct senses. Instead, this paper investigates one important sense in which Rawls uses the term: the rational democratic consensus *in* equilibrium, capable of evolving into an overlapping consensus. I show that Rawls makes claims critical of the democratic modus vivendi that are not true of this democratic equilibrium, and he makes claims critical of rationality that do not describe rationality as it is expressed in this context. This paper evaluates these critiques and revises Rawls's characterizations accordingly.

By articulating the properties of a rational democratic agreement in equilibrium, as well as the rational evaluation on which it is founded, this argument lays a foundation for the further work of (1) parsing Rawls's account of the democratic modus vivendi into distinct senses: democracy in equilibrium and out of it, (2) investigating why Rawls's limited conception of rationality leads him to conflate the two, and (3) developing prescriptions for effecting the former rather than the latter.

[10] Two is the *minimum* number of groups defined by their differential proximity to power; under real-world conditions, the number of such groups defies discrete divisions.

2. **Private Advantage:** Citizens must consider participation advantageous to their long-term interests; they believe themselves "better off under the pact than without it" (Mittal and Weingast 2013, 285).
3. **Pact-Defending Behavior:** Citizens must act to uphold the pact and dissuade their co-citizens from transgressing it: "the parties to the pact must be willing to *defend* the pact against transgressions by political leaders … they defend not only the parts of the pact benefitting themselves but also the parts benefitting others" (Mittal and Weingast 2013, 286, emphasis added; see also, Przeworski 1991, 25–26).
4. **Coordination:** Citizens coordinate in choosing to uphold the agreement (Mittal and Weingast 2013, 285). Each has incentives to uphold the pact only if she expects her co-citizens to follow suit. When these coordinated expectations are in place, citizens have reason to practice pact-upholding behavior—including "defend[ing] … the parts benefitting others"—because, reciprocally, "each party anticipates that its rights will be defended by others" (Mittal and Weingast 2013, 286).

This framework of incentives is necessary to secure the long-term advantage of democratic participants and render the agreement self-enforcing.

Finally, I reiterate that a self-enforcing democracy is an agreement in equilibrium. Its citizens consider democratic participation more advantageous than reneging, and they calculate that this is true *in expectation*: "in equilibrium, each party has definite expectations as to what it will receive now and in the future; it attaches a fixed value to future life under democracy" (Przeworski 2005, 269). When citizens evaluate their expected long-term advantage under the pact, this calculation considers the features of democratic participation outlined above: the structures of governance and hierarchy, the sharing of power, the need for coordinated cooperation, and so on. If citizens conclude that the pact serves their long-term advantage, they have incentives to accept short-term losses incurred through normal democratic competition because, on balance, they expect to benefit from ongoing democratic participation (Przeworski 1991, 19).

Destabilization of Rational Democratic Consensus

Finally, I outline the mechanism by which a rational democratic equilibrium might be destabilized, and I highlight a key difference between this outline and Rawls's account of the process. Within a self-enforcing democracy, rational citizens expect their lifetime advantage to be better served under the democratic pact than otherwise. Unforeseen circumstances can, however, trigger changes to this calculation of expected long-term value. These changes to the value of democratic participation can be catalyzed by "endogenous processes, exogenous shocks, and combinations of both" (Greif and Laitin 2004, 639).[11] If such circumstances "alter payoffs so that citizens no longer have incentives to cooperate," this decrease in the expected value of democratic participation threatens pact stability: under these conditions, the expected long-term value of undermining the pact rivals that of upholding it (Mittal and Weingast 2013, 286).

However, unexpected changes of circumstance do not *directly* dictate the value that rational individuals place on democratic participation, nor, therefore, do they directly dictate the stability or instability of democratic institutions. Rather, citizens calculate how unexpected change shapes their interests and behavior based, in large part, on how they expect their peers to respond. This is because citizens cannot unilaterally transgress the pact. Przeworski (1991, 28) observes that "isolated individuals do not shake social orders … only organized political forces have the capacity to undermine the democratic system." Like stability, instability results from coordinated action. If stability reflects a coordinated effort to uphold democracy, instability reflects a coordinated effort to undermine it.

Thus, if enough citizens lack incentives to continue cooperating under unforeseen circumstances, they might constitute an "organized political force" capable of threatening democratic stability. If, however, enough citizens prefer to continue coordinating around pact-upholding cooperation, they can discourage such attempts to renege.[12] Democratic equilibrium can weather unforeseen circumstances if citizens coordinate around pact-upholding behavior. However, as Rawls observes of the coordination required to secure an overlapping consensus, "[while] that is the hope; there can be no guarantee" (Rawls, 65).[13]

Within Rawls's account, the stability of the modus vivendi depends *entirely* on chance conditions: he describes its stability as simply "contingent on circumstances remaining such as not to upset the fortunate convergence of interests" (Rawls, 147; see also, xli). By contrast, accounts of self-enforcing democracy show that unexpected circumstances influence rational individuals' calculation of their long-term interest in democratic participation but do not represent the whole of this calculation. Rather, because citizens must coordinate their political action with their peers, much of this

[11] Exogenous shocks occur external to the agreement, e.g., through economic crises (Przeworski 2005, 265). Endogenous change occurs when institutions are structured such that their operation induces gradual alterations to citizens' assessment of the long-term value of cooperation (Greif and Laitin 2004, 639).

[12] Mittal and Weingast (2013, 280, 286–287) note that institutions that can be adapted to meet changing conditions can generate incentives that promote this cooperative coordination.

[13] An overlapping consensus can be destabilized if unreasonable citizens come to occupy significant political power (Rawls, 65). Nevertheless, within Rawls's account of the overlapping consensus, this potential for failure does not diminish the value of investigating the conditions for success. The same, I argue, is true of the democratic modus vivendi—particularly because a successful modus vivendi lays the foundation for an overlapping consensus.

calculation rests on their interdependent choices, expectations, and behavior.

Argument Overview

By describing the democratic modus vivendi as a rational democratic consensus that "work[s] effectively and successfully for a sustained period of time," Rawls describes a democratic agreement in equilibrium (Rawls, 163). The remainder of the paper evaluates Rawls's characterization of rational democratic equilibrium by comparison with that developed in formal accounts of self-enforcing democracy. Through this comparison, I assess Rawls's positive assertions about the robustness, object, and grounds of rational democratic equilibrium.

Drawing on the definition of self-enforcing democracy, I demonstrate that (a) the *stability* of a rational democratic consensus can withstand variation in the distribution of political power; (b) the *object of consensus* is not limited to democratic institutions; it also comprises cooperative behavior functionally equivalent to Rawls's reasonable "cooperative virtues"; and (c) although self-interest serves as the *grounds of consensus*, there are various modes by which interest can be pursued; rational democratic equilibrium is grounded in a mode of rationality that is, like the reasonable, reliably other-regarding and self-moderating.

ROBUSTNESS UNDER A CHANGING DISTRIBUTION OF POWER

This section addresses Rawls's claims about the *robustness* of a rational democratic equilibrium: that (1) individuals' interest in upholding this equilibrium erodes as they gain access to power; consequently, (2) the democratic agreement's "stability with respect to the distribution of power is lacking" (Rawls, 148). Rawls argues these claims by analogy, equating the stability of the democratic modus vivendi with that of "a treaty between two states whose national aims and interests put them at odds" (Rawls, 147). I reject this analogy, demonstrating that the institutions and incentives entailed in a democratic pact differ from those of a treaty. I show that within a self-enforcing democracy, (1) citizens' incentives to uphold the pact do not vary with their proximity to power; therefore, (2) the democratic equilibrium can withstand a changing distribution of political power.

Self-Enforcing Agreements and Their Participants

I begin by contrasting the conditions required for democratic stability with those entailed in a treaty. Like the self-enforcing democracy, a treaty is a form of self-enforcing agreement: stability depends on the voluntary compliance of the agreement's participants, as no power external to the agreement is authorized to enforce it (Telser 1980, 27). Therefore, both forms of agreement are stable so long as "all politically significant groups" participating in the agreement choose to comply with the pact rather than undermine it (Burton, Gunther, and Higley 1992, 3). Voluntary compliance, in turn, is motivated by calculations of private advantage; pacts are self-enforcing only if "participants all perceive that they are better off under the pact than under the status quo" (Weingast 1997a, 258).

However, although treaty and democracy are both self-enforcing agreements, they establish distinct institutions. The relationship between sovereign states remains *anarchic* under a treaty: states do not create governing institutions authorized to enforce compliance with the agreement. By contrast, the relationship among citizens becomes *hierarchical* under the democratic agreement: democratic institutions establish a government of citizens with enforcing power (Waltz 1979, 79–128). Thus, democratic citizens differ in their proximity to governing power at any given time, exerting varying degrees of control over its use. While citizens are differentiated by their proximity to governing power, states remain undifferentiated in this respect.[14]

Because treaty and democracy establish distinct institutional structures and create distinct relational structures among their participants, they also produce distinct sets of "politically significant groups" whose compliance must be secured. The democratic agreement contains a set of groups defined by their proximity to power: democratic citizens can be divided, at any given time, into a group "in power" and one "out of power." By contrast, an anarchic agreement lacks a hierarchical governing structure, and it therefore lacks this set of groups defined by their relative proximity to governing power.

By definition, a self-enforcing pact provides incentives for all participants—all "politically significant groups"—to comply with its terms. Democratic citizens recognize that the pact differentiates them into such groups based on their relative proximity to power, and they accept that this hierarchy is dynamic, subject to change through processes of democratic competition. Thus, a democratic agreement is only self-enforcing if, at any given time, citizens throughout this power hierarchy have adequate incentives to accept and uphold the democratic agreement (Przeworski 1991, 30–31).

Stability and the Distribution of Power

Rawls conflates two forms of "power" when he likens the democratic modus vivendi to a treaty: the anarchic "power" of sovereign states ebbs and flows outside the terms of the pact, whereas the hierarchical "power" of citizens is established by the pact itself. Because the latter is granted and regulated by the agreement,

[14] Waltz (1979, 81) observes that, unlike international anarchy, "domestic politics is hierarchically ordered. The units … stand vis-à-vis each other in relations of super- and subordination… . Political actors are formally differentiated according to the degrees of their authority." See also Przeworski (2010, 127–128).

citizens assess its structure and dynamics when offering their consent. This observation allows us to address Rawls's claims about the robustness of the democratic modus vivendi under a changing distribution of power.

First, I counter Rawls's assertion that citizens' interest in democratic participation erodes as they gain access to power. By definition, a self-enforcing democratic pact entails the existence of group-level incentives that, in expectation, motivate compliance with its terms: at any given time, citizens *occupying* power have incentives to obey the boundaries of the pact rather than abuse their power by changing the rules, transgressing the rights of their opponents, or refusing to cede power. Simultaneously, citizens *out* of power have incentives to accept the leadership of the governing coalition, rather than rejecting its authority or taking extraconstitutional action to override it (see, e.g., Mittal and Weingast 2013, 279; Przeworski 2005, 270; Weingast 2004, 162). If a democratic pact lacks these incentives for citizens throughout the power hierarchy to comply, it fails to meet the definition of a self-enforcing agreement.

Second, I address Rawls's assertion that the democratic modus vivendi is unstable under a changing distribution of power. The sharing and transfer of power is a defining feature of democratic governance. When citizens consent to the pact, they accept that their proximity to power is subject to change, given standard practices of democratic politics.[15] Citizens only have reason to accept these power-sharing implications of democratic governance if, as described, groups across the power hierarchy have ongoing incentives to sustain the agreement. These conditions promote the peaceful sharing and transfer of power: political winners have incentives not to entrench their hold on power by amending the rules of the game or disenfranchising their opponents, and political losers have incentives to accept the temporary loss of power because they do not fear these consequences.[16] In short, the definition of self-enforcing democracy implies the existence of incentives that ensure that the transfer of power does not disrupt democratic governance. This counters Rawls's claim that "stability with respect to the distribution of power is lacking."

THE OBJECT OF CONSENSUS AND DEMOCRATIC INCENTIVES

This section identifies the incentives that motivate politically significant groups to uphold the democratic pact and shows how these incentives persist despite changes in the distribution of power. It demonstrates that these incentives necessitate a widespread consensus on practices of citizenship functionally equivalent to the "cooperative virtues" that Rawls attributes to the reasonable. Thus, I counter Rawls's assertion that the *object of consensus* under the democratic modus vivendi is "merely a consensus on accepting certain authorities, or on complying with certain institutional arrangements."

Incentives for Politically Significant Groups

Within a self-enforcing pact, rational individuals calculate that, in expectation, democratic participation promotes their long-run advantage (Przeworski 2005, 269).[17] This calculation assumes the features that define a democratic agreement: the institutionalization of governing power and the power-sharing inherent in democratic governance.[18] Given these considerations, acquiescent citizens calculate that they benefit from the ongoing stability of the democratic pact.

However, these incentives to comply with the agreement depend on the pact-compliant behavior of other citizens. If citizens subvert the negotiated agreement, they affect their peers' incentives to uphold those terms.[19] Citizens throughout the power hierarchy can transgress the pact: power-holders can violate their co-citizens' rights or refuse to cede power, while opposition groups might refuse to abide by government decisions or seek power extraconstitutionally.

However, citizens cannot *unilaterally* transgress the democratic pact without inviting punishment. Because the agreement authorizes the enforcement of its terms through coercive force, transgressions by individuals or small groups are likely to be costly and unsuccessful (see, e.g., Diamond 1999, 67–68; Przeworski 1991, 28). Therefore, efforts to subvert the agreement require coordinated action (Weingast 2004, 170). If a sufficiently large bloc of citizens supports transgressive action, this bloc, or its leaders, can successfully undermine the agreement at low cost. If, however, most citizens reject such attempts, the benefits to be gained from transgressing the pact are likely to be exceeded by the cost and risk of the attempt.[20] In this case, groups and their leaders are unlikely to attempt transgression: "rather than risk failure, leaders are deterred from

[15] Przeworski (1988, 64) notes that "democracy is possible when the relevant political forces can find institutions that would provide a reasonable guarantee that their interests would not be affected in a highly adverse manner in the course of democratic competition."

[16] Mittal and Weingast (2013, 279) observe that "stable democracy requires that incumbent officials who lose elections must have incentives to step down and those out of power must be willing to eschew force as a means of taking control of the government." See also Przeworski (1991, 26).

[17] Telser (1980, 42) asserts, "A self-enforcing agreement is possible if and only if the expected future gains from adherence to it exceeds [sic] the current gain from a violation of the agreement." See also Elster (1993, 175).

[18] Przeworski (1991, 33) observes that self-enforcing democratic institutions "make even losing under democracy more attractive than a future under nondemocratic alternatives."

[19] Mittal and Weingast (2013, 279), Przeworski et al. (2000, 16–18), and Przeworski (1991, 28) describe how citizens' incentives to comply are affected by the actions of their peers.

[20] Mittal and Weingast (2013, 297) conclude that "citizen coordination against leaders who transgress their rights is central to maintaining a constitution. This coordination threatens leaders with the withdrawal of support necessary to retain power, providing the incentives not to contemplate transgressions."

violating the rules by adverse citizen reaction" (Weingast 2004, 170).

By displaying a pact-upholding orientation and a willingness to reject attempts at transgression, citizens establish key incentives for groups to uphold the agreement, creating the conditions necessary for democratic stability.[21] The self-enforcing democratic pact requires citizens to collectively "defend democracy against transgressions" (Weingast 2004, 162; see also Weingast 1997a, 251). Thus, citizens must reject transgressive attempts by both (1) groups *in* power, which might seek to entrench their hold on power or transgress the rights of other citizens, and (2) groups *out* of power, which might seek to seize power extraconstitutionally or secede (see, e.g., Diamond 1999, 70; Weingast 1997a, 260; Weingast 2004, 162–163). So long as rational agents calculate that the democratic agreement furthers their long-term advantage, they have incentives to hold group behavior in check.

Therefore, citizens themselves create the linchpin of the incentive structure that elicits ongoing compliance from politically significant groups and secures these incentives regardless of the distribution of power. Citizens' coordinated behavior ensures that groups throughout the power hierarchy have reason to abide by the pact rather than undermine it. By discouraging group-level transgressions, citizens promote the widespread compliance necessary for democratic stability.

Rational Consensus on Cooperative Action

This discussion counters Rawls's assertion that the object of consensus under the democratic modus vivendi (1) is "merely a consensus on accepting certain authorities, or on complying with certain institutional arrangements" and therefore (2) lacks properties that he ascribes to the reasonable.

First, I have demonstrated that an object of consensus limited to "institutional arrangements" cannot generate the incentives necessary for self-enforcing democracy. Instead, consent to democratic institutions entails a corresponding consensus on the actions required to secure them.[22] Citizens must agree on what constitutes a transgression of the democratic agreement and actively respond to transgressions. When "all citizens hold the same views about transgressions and citizen duty," citizens can coordinate their responses, curtailing destabilization of the pact (Weingast 1997a, 251). Without this consensus on pact-upholding behavior, groups lack incentives to comply and the pact lacks a self-enforcing incentive structure.

Second, I counter Rawls's claim that this object of consensus lacks content that Rawls ascribes to the reasonable. Rawls labels practices of political compromise, moderation, and toleration as reasonable "cooperative virtues." These practices express other-regarding considerations and a self-moderating capacity: reasonable citizens demonstrate *other-regard*, or attentiveness to others' interests, by "tak[ing] into account the consequences of their actions on others' well-being" (Rawls, 49 n. 1); they practice *self-moderation* by constraining their pursuit of private ends to preserve the democratic agreement. This practice of the cooperative virtues is *enduring* and reflects a *general regard* for all polity members.

While Rawls associates the cooperative virtues with principled reasoning, I have shown that the rational can incentivize functionally equivalent behavior: rational citizens act in accordance with the cooperative virtues when they practice the pact-upholding behaviors necessary to maintain rational democratic equilibrium. First, maintaining democratic stability "requires that [citizens] defend not only the parts of the pact benefitting themselves but also the parts benefitting others" (Mittal and Weingast 2013, 286). Thus, rational citizens have incentives to engage in behavior that is functionally equivalent to the *other-regard* motivated by reasonable principles. To uphold self-enforcing democracy, rational citizens attend to their co-citizens' interests by rejecting attempts to transgress their rights; accepting the decisions these co-citizens authorize when in power—even when these conflict with their own agenda and interests—and compromising their ideal terms of cooperation to ensure that others benefit adequately under the pact (Weingast 1997a, 251–252; 2004, 171–172).

In turn, rational citizens within a self-enforcing democracy practice ongoing *self-moderation*, which Rawls defines as citizens' willingness to "act on [fair] terms, even at the cost of their own interests" (Rawls, xlii). They incur short-term costs to reject group-level transgressions, even when they might benefit from the pact-transgressive behavior of their affiliated groups.[23] Rational citizens have incentives to assume these short-term costs if they expect the lifetime benefit of stable democratic institutions to exceed them.

Finally, I counter Rawls's claim that rational cooperative behavior is necessarily *mutable* rather than enduring and *particularistic* rather than general. Within a self-enforcing democracy, citizens have incentives to maintain an ongoing cooperative orientation toward all polity members: given that democratic participation serves their long-term interests, (1) each recognizes that democratic stability depends on the widespread, coordinated cooperation of her peers. In turn, (2) each knows that her peers will only cooperate to uphold the

[21] Przeworski (1991, 29), Diamond (1999, 66), Weingast (1997b, 22), Weingast (1997a, 252), Fearon (2011, 1662), Mittal and Weingast (2013, 279–280), and Hyde and Marinov (2014, 329–330) discuss the role this coordinated citizen defense plays in democratic stability.

[22] Weingast (1997b, 23) notes that "citizens in stable democracies possess a relatively common set of understandings about the appropriate boundaries of government and about their duty in the face of violations of these boundaries."

[23] Weingast (1997a, 257) observes that within a self-enforcing democracy, "citizens agree that the rules must be defended and that appeals to violate them must be opposed, even by the intended beneficiaries of the violation." See also Diamond (1999, 70) and Weingast (1997b, 12).

pact if democratic participation furthers their interests. Finally, (3) each recognizes that her own behavior affects her peers' calculation of their interest in democratic participation, as their interest, like hers, depends on widespread cooperation.

Thus, each rational agent has incentives to maintain cooperative habits, upholding the pact and the interests of her co-citizens: by this means, she motivates her peers to reciprocate her cooperative orientation, securing her own interests under the pact (Elster 2006, 188; Weingast 2004, 172–173). Weingast (1997a, 262) summarizes this mechanism: "citizens aid those who are threatened because the potential victims will later fail to come to their aid if they fail to come to the victims' aid." Thus, within a self-enforcing democracy, rational citizens' cooperative behavior is not limited in the ways Rawls assumes: it is general rather than particularistic, as rational actors have incentives to secure their co-citizens' interest in democratic participation, whether or not they share personal attachments; it is enduring rather than mutable, as these incentives persist as long as citizens expect a long-term benefit from democratic participation.

Rawls believes that a rational democratic equilibrium arises in the absence of these cooperative practices: because the cooperative virtues stem from the reasonable, he claims, they appear only *after*—indeed, due to—the consolidation of democratic institutions (Rawls, 163). However, I have shown that self-interest can motivate these behaviors and, indeed, that maintaining a self-enforcing democracy depends on their widespread expression. I conclude that the object of consensus under a democratic modus vivendi shares a key property with the reasonable: rational democratic consensus entails consent to the cooperative behavior that renders democracy stable.

GROUNDS OF CONSENSUS AND THE RATIONAL POWER

Finally, I assess Rawls's account of the reasoning that *grounds* a rational democratic equilibrium. According to Rawls's account of political reasoning, a rational basis for democratic consent lacks key properties of the reasonable—namely, the capacity to reliably prioritize other-regarding political ends and moderate self-regarding ones. The previous section, however, demonstrated that the rational can motivate cooperative practices that reliably prioritize other-regarding, self-moderating considerations. The following section suggests that Rawls's account of the rational must be expanded to account for this capacity: it identifies two modes of pursuing interest—one self-regarding and one cooperative—where Rawls reduces interest to the former. It concludes that rational citizens establish and sustain self-enforcing democracy by choosing the cooperative mode of interest rather than its self-regarding cousin.

Modes of Self-Interest: Rivalrous and Equitable

Rawls asserts that self-interest cannot motivate reliable habits of cooperation. Thus, his account of self-interest should be expanded to reflect this capacity. Whereas Rawls posits a single form of self-interest, I differentiate it into two modes. Adopting Danielle Allen's terminology, these are (1) a "rivalrous" mode that matches Rawls's definition and (2) an "equitable" alternative capable of motivating ongoing cooperative habits (Allen 2004, 134, 137–138). Rational democratic equilibrium demands that citizens choose the equitable orientation toward interest rather than the rivalrous alternative Rawls assumes.

As Rawls himself observes, rational agents select the ends they pursue and decide their relative priority (Rawls, 50–51). Rivalrous and equitable modes of self-interest, then, are differentiated by the relative priority given to the preservation of relationships—for instance, the political relationships constituted by a democratic agreement. *Rivalrous* agents choose not to prioritize the preservation of these relationships relative to their other interests. Because they do not value these bonds, rivalrous agents act from "a commitment to [their] own interests without regard to how they affect others" (Allen 2004, 134). If this attitude is widespread in a polity, its politics will reflect the myopic pursuit of factional interests. Allen observes, "no consensually based form of social organization can, over the long term, sustain relationships of cooperation in the face of unrestrained self-interest" (Allen 2004, 137). Because the democratic political relationship demands ongoing cooperation, self-interest pursued rivalrously cannot bring about this end.

By contrast, a rational agent chooses the *equitable* approach to self-interest if she places high priority on preserving relationships. Equitable agents recognize that their partnerships continue only so long as their peers have reason to participate (Allen 2004, 134–139). Therefore, equitable agents practice self-moderation and other-regard as they pursue private ends, recognizing that this cooperative orientation is necessary to maintain stable relationships (Allen 2004, 126, 135–136). In a political context, citizens with an interest in maintaining self-enforcing democracy have incentives to pursue equitability and avoid rivalrousness, as the former bolsters their peers' interests in upholding the pact, while the latter diminishes them (Allen 2004, 126).

Thus, I contest Rawls's reductive account of interest, which constrains it to its rivalrous form. I have presented an equitable alternative, which rational agents choose when they value the continuation of their relationships. Interest pursued in this equitable mode shares characteristics with the reasonable: it awards high priority to other-regarding considerations and motivates self-moderating constraints. To uphold self-enforcing democracy, rational citizens must choose to forego rivalrousness in their pursuit of self-interest and adopt this equitable orientation instead.

Finally, I reflect on Rawls's claim that rational democratic equilibrium derives from a "fortunate convergence of interests" (Rawls, 147). Insofar as citizens' choices and actions help constitute their co-citizens' interests in democratic participation, I argue that the resulting convergence of interests ought not merely be attributed to luck. Rather, each citizen contributes to this convergence by choosing to pursue her interest equitably, providing incentives for her peers to reciprocate this equitability. The convergence of interests reflects citizens' coordinated choice to pursue interest in its equitable mode, giving each other key incentives to support the pact. Indeed, this coordinated choice of equitability follows the same logic of interdependence by which Rawls describes citizens' coordinated choice of the reasonable: citizens have reason to adopt a cooperative orientation "provided others can be relied on to do the same" (Rawls, 54).[24]

CONCLUSION

Redrawing the Democratic Modus Vivendi

I conclude that democratic consensus achieved as a modus vivendi can escape Rawls's critiques of its (a) *grounds*, (b) *object*, and (c) *robustness* to variation in the distribution of political power. Drawing upon formal accounts of self-enforcing democracy, I showed that (a) citizens help establish the convergence of interests by choosing an *equitable* orientation toward self-interest; (b) equitable citizens consent to not only democratic institutions but also practices of cooperative citizenship; and finally, (c) supported by citizens' cooperative choices and actions, the self-enforcing democratic pact can withstand variation in the distribution of political power.

This argument, therefore, counters Rawls's claim that "no sensible view can possibly get by without the reasonable and rational as I use them" (Rawls, 380). It demonstrates that a rational democratic equilibrium overcomes Rawls's critiques precisely because the rational can display characteristics that Rawls reserves to the reasonable: other-regarding political priorities, a self-moderating capacity, and motivation for ongoing cooperative behavior. I conclude that Rawls's account of the rational should be revised accordingly to allow for this equitable alternative to rivalrous rationality.

[24] Rawls refers to these mutual expectations of cooperation as "trust" (Rawls 2005, 163). Vallier (2019) also grounds a concept of trust in moral motivations and associates the concept with reliable cooperation: "trust depends upon empirical expectations that others will comply with recognized rules of peaceful conduct" (Vallier 2019, 44). Though I do not claim that cooperation motivated by pragmatism engenders the rich notion of trust that Rawls and Vallier derive from moral motivations, I note that it *does* entail "empirical expectations that others will comply."

Pursuing an Account of Equitable Self-Interest

This argument motivates further theoretical engagement with the equitable mode of rationality and its role in motivating political cooperation. In particular, by parsing rationality into equitable and rivalrous modes, the paper highlights the choice that rational agents make between them. It promotes an agenda that investigates the factors that shape this decision making and generates prescriptions for promoting the equitable outcome. I outline five strands for further work in this vein:

1. **Balancing Priorities:** This paper demonstrates that the rational citizen's calculation of her political interest entails the ongoing interaction of self- and other-regarding priorities. Thus, it invites analysis of the mechanism by which rational individuals weigh and balance these priorities. Rawls's account of political reasoning does not pursue this line of inquiry, as he isolates self-regarding political priorities from other-regarding ones by assigning them to distinct modes of reasoning.
2. **Coordinated Choice:** The paper motivates investigation into the origins and maintenance of an "equitable society"—the widespread, coordinated choice to pursue interest equitably rather than rivalrously—which is necessary to sustain self-enforcing democracy.
3. **Distribution of Equitability:** Rawls attributes democratic stability to the distribution of *political power* in the case of the modus vivendi and to the distribution of *moral belief* in his account of the overlapping consensus. This paper focuses attention on the distribution of *equitability* as a consideration salient to democratic stability: the choice of equitable self-interest promotes self-enforcing democracy, whereas that of rivalrous self-interest undermines it, and citizens make this choice interdependently with their peers.
4. **Cross-Methodological Collaboration:** The paper's argument draws on both normative and formal political theory. Its findings encourage further collaboration among empirical accounts of consolidated democracy and democratic transition, formal accounts of self-enforcing democracy, and normative accounts of democratic participation. Interaction between these approaches can further illuminate the role self-interest plays in the establishment and maintenance of democratic cooperation, as well as the factors that encourage citizens to pursue interest in its equitable, rather than rivalrous, mode.
5. **Range of Motivations:** While the paper presents interest as a meaningful motivation for democratic cooperation, it does not discourage engagement with other motivations for cooperative behavior—Rawls's moral principles, for example, or the shared norms and traditions invoked by modus vivendi theorists. Rather, it encourages further examination of the interaction between interest and other modes of reasoning as they factor into citizens' political decision making.

Equitable Self-Interest: A Shared Object of Inquiry

Rationality plays a foundational role in Rawls's ideal accounts of democratic consensus and stability, as he builds his overlapping consensus on a modus vivendi basis. Developing a realistic account of the overlapping consensus, as he intends, demands an accurate representation of the self-enforcing democratic modus vivendi on which it is founded. To that end, this paper corrects Rawls's account of the modus vivendi by introducing the equitable alternative to rivalrous rationality and outlining its role in sustaining rational democratic equilibrium. The agenda presented above can further inform the account of the democratic modus vivendi that lays the groundwork for Rawls's ideal.

More broadly, this agenda should have wide appeal for scholars of democratic cooperation and stability—whether Rawlsian accounts that build ideals of democratic consensus on a modus vivendi foundation, realists and modus vivendi theorists who are developing frameworks that do not rely on demanding moral conceptions, or formal and empirical accounts that seek a conceptual vocabulary to describe citizens' political reasoning within a rational democratic equilibrium. Insofar as these accounts engage with rationality as a motivation for democratic participation, this agenda on equitable self-interest offers a common object of inquiry and invites collaboration among them.

ACKNOWLEDGMENTS

My warmest thanks to Lanier Anderson, Emilee Chapman, Jennifer Cryer, Sung Mi Kim, Minh Ly, Alison McQueen, Josh Ober, Rob Reich, Quentin Skinner, Chloe Stowell, and Barry Weingast for their guidance and feedback as this paper developed. I also thank the organizers and participants of the Stanford Political Theory Workshop for a productive discussion of this paper in its early stages. Finally, I am grateful to the *APSR* editorial team and three anonymous reviewers for their helpful comments and direction.

REFERENCES

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

Allen, Danielle S. 2004. *Talking to Strangers: Anxieties of Citizenship since Brown v. Board of Education*. Chicago: University of Chicago Press.

Arnsperger, Christian, and Emmanuel B. Picavet. 2004. "More than *Modus Vivendi*, Less than Overlapping Consensus: Towards a Political Theory of Social Compromise." *Social Science Information* 43 (2): 167–204.

Baier, Kurt. 1989. "Justice and the Aims of Political Philosophy." *Ethics* 99 (4): 771–90.

Burton, Michael, Richard Gunther, and John Higley. 1992. "Introduction: Elite Transformations and Democratic Regimes." In *Elites and Democratic Consolidation in Latin America and Southern Europe*, eds. John Higley and Richard Gunther, 1–37. Cambridge: Cambridge University Press.

Carugati, Federica. 2019. *Creating a Constitution: Law, Democracy, and Growth in Ancient Athens*. Princeton, NJ: Princeton University Press.

Chung, Hun. 2019. "The Instability of John Rawls's 'Stability for the Right Reasons.'" *Episteme* 16 (1): 1–17.

Chung, Hun. 2020. "The Well-ordered Society under Crisis: A Formal Analysis of Public Reason vs. Convergence Discourse." *American Journal of Political Science* 64 (1): 82–101.

Dauenhauer, Bernard P. 2000. "A Good Word for a Modus Vivendi." In *The Idea of a Political Liberalism: Essays on Rawls*, eds. Victoria Davion and Clark Wolf, 204–20. Lanham, MD: Rowman & Littlefield.

Diamond, Larry. 1999. *Developing Democracy: Toward Consolidation*. Baltimore, MD: Johns Hopkins University Press.

Dryzek, John S., and Simon Niemeyer. 2006. "Reconciling Pluralism and Consensus as Political Ideals." *American Journal of Political Science* 50 (3): 634–49.

Elster, Jon. 1993. "Constitution-making in Eastern Europe: Rebuilding the Boat in the Open Sea." *Public Administration* 71 (1–2): 169–217.

Elster, Jon. 2006. "Altruistic Behavior and Altruistic Motivations." In *Handbook of the Economics of Giving, Altruism and Reciprocity, Vol.* I, eds. Serge-Christophe Kolm and Jean Mercier Ythier, 183–206. Amsterdam: Elsevier.

Elster, Jon. 2009. *Reason and Rationality*. Princeton, NJ: Princeton University Press.

Fearon, James D. 2011. "Self-enforcing Democracy." *The Quarterly Journal of Economics* 126 (4): 1661–708.

Forrester, Katrina. 2019. *In the Shadow of Justice: Postwar Liberalism and the Remaking of Political Philosophy*. Princeton, NJ: Princeton University Press.

Fossen, Thomas. 2019. "Modus Vivendi Beyond the Social Contract: Peace, Justice, and Survival in Realist Political Theory." In *The Political Theory of Modus Vivendi*, eds. John Horton, Manon Westphal, and Ulrich Willems, 111–27. Cham, Switzerland: Springer.

Galston, William A. 2010. "Realism in Political Theory." *European Journal of Political Theory* 9 (4): 385–411.

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

Gray, John. 2000. *Two Faces of Liberalism*. New York: The New Press.

Greif, Avner, and David D. Laitin. 2004. "A Theory of Endogenous Institutional Change." *American Political Science Review* 98 (4): 633–52.

Hershovitz, Scott. 2000. "A Mere Modus Vivendi?" In *The Idea of a Political Liberalism: Essays on Rawls*, eds. Victoria Davion and Clark Wolf, 221–30. Lanham, MD: Rowman & Littlefield.

Horton, John. 2010. "Realism, Liberal Moralism and a Political Theory of Modus Vivendi." *European Journal of Political Theory* 9 (4): 431–48.

Horton, John. 2011. "Modus Vivendi and Religious Conflict." In *Democracy, Religious Pluralism and the Liberal Dilemma of Accommodation*, ed. Monica Mookherjee, 121–36. Dordrecht, Netherlands: Springer.

Hyde, Susan D., and Nikolay Marinov. 2014. "Information and Self-enforcing Democracy: The Role of International Election Observation." *International Organization* 68 (2): 329–59.

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

Larmore, Charles. 1990. "Political Liberalism." *Political Theory* 18 (3): 339–60.

Linz, Juan J., and Alfred Stepan. 1996. *Problems of Democratic Transition and Consolidation: Southern Europe, South America, and Post-communist Europe*. Baltimore, MD: Johns Hopkins University Press.

McCabe, David. 2010. *Modus Vivendi Liberalism: Theory and Practice*. Cambridge: Cambridge University Press.

Mittal, Sonia, and Barry R. Weingast. 2013. "Self-enforcing Constitutions: With an Application to Democratic Stability in America's First Century." *The Journal of Law, Economics, & Organization* 29 (2): 278–302.

Ober, Josiah. 2017. *Demopolis: Democracy before Liberalism in Theory and Practice*. Cambridge: Cambridge University Press.

Przeworski, Adam. 1988. "Democracy as a Contingent Outcome of Conflicts." In *Constitutionalism and Democracy*, eds. Jon Elster and Rune Slagstad, 59–80. Cambridge: Cambridge University Press.

Przeworski, Adam. 1991. *Democracy and the Market: Political and Economic Reforms in Eastern Europe and Latin America*. Cambridge: Cambridge University Press.
Przeworski, Adam, Michael E. Alvarez, José Antonio Cheibub, and Fernando Limongi. 2000. *Democracy and Development: Political Institutions and Well-being in the World, 1950–1990.* Cambridge: Cambridge University Press.
Przeworski, Adam. 2005. "Democracy as an Equilibrium." *Public Choice* 123 (3–4): 253–73.
Przeworski, Adam. 2006. "Self-enforcing Democracy." In *The Oxford Handbook of Political Economy*, eds. Barry R. Weingast and Donald A. Wittman, 312–28. Oxford: Oxford University Press.
Przeworski, Adam. 2010. *Democracy and the Limits of Self-government*. Cambridge: Cambridge University Press.
Rawls, John. 1999. *The Law of Peoples, with "The Idea of Public Reason Revisited*." Cambridge, MA: Harvard University Press.
Rawls, John. 2001. *Justice as Fairness: A Restatement*, ed. Erin Kelly. Cambridge, MA: Harvard University Press.
Rawls, John. 2005. *Political Liberalism, Expanded Edition.* New York: Columbia University Press.
Sala, Roberta. 2019. "Modus Vivendi and the Motivations for Compliance." In *The Political Theory of Modus Vivendi*, eds. John Horton, Manon Westphal, and Ulrich Willems, 67–82. Cham, Switzerland: Springer.
Schedler, Andreas. 1998. "What is Democratic Consolidation?" *Journal of Democracy* 9 (2): 91–107.
Talisse, Robert B. 2003. "Rawls on Pluralism and Stability." *Critical Review* 15 (1–2): 173–94.
Telser, L. G. 1980. "A Theory of Self-enforcing Agreements." *The Journal of Business* 53 (1): 27–44.
Thrasher, John, and Kevin Vallier. 2013. "The Fragility of Consensus: Public Reason, Diversity and Stability." *European Journal of Philosophy* 23 (4): 933–54.
Thrasher, John, and Kevin Vallier. 2018. "Political Stability in the Open Society." *American Journal of Political Science* 62 (2): 398–409.
Vallier, Kevin. 2019. *Must Politics Be War? Restoring Our Trust in the Open Society*. New York: Oxford University Press.
Wall, Steven. 2019. "Liberal Moralism and Modus Vivendi Politics." In *The Political Theory of Modus Vivendi*, eds. John Horton, Manon Westphal, and Ulrich Willems, 49–66. Cham, Switzerland: Springer.
Waltz, Kenneth. 1979. *Theory of International Politics*. Reading, MA: Addison-Wesley.
Weingast, Barry R. 1997a. "The Political Foundations of Democracy and the Rule of Law." *American Political Science Review* 91 (2): 245–63.
Weingast, Barry R. 1997b. "Democratic Stability as a Self-enforcing Equilibrium." In *Understanding Democracy: Economic and Political Perspectives*, eds. Albert Breton, Gianluigi Galeotti, Pierre Salmon, and Ronald Wintrobe, 11–46. Cambridge: Cambridge University Press.
Weingast, Barry R. 2004. "Constructing Self-enforcing Democracy in Spain." In *Politics from Anarchy to Democracy: Rational Choice in Political Science*, eds. Irwin L. Morris, Joe A. Oppenheimer, and Karol Edward Soltan, 161–95. Stanford, CA: Stanford University Press.
Wendt, Fabian. 2016. "The Moral Standing of Modus Vivendi Arrangements." *Public Affairs Quarterly* 30 (4): 351–70.
Wendt, Fabian. 2019. "Why Theorize Modus Vivendi?" In *The Political Theory of Modus Vivendi*, eds. John Horton, Manon Westphal, and Ulrich Willems, 31–47. Cham, Switzerland: Springer.
Williams, Bernard. 2005. "Realism and Moralism in Political Theory." In *In the Beginning was the Deed: Realism and Moralism in Political Argument*, ed. Geoffrey Hawthorn, 1–17. Princeton, NJ: Princeton University Press.