The Virtue of Deliberation:
Sophrosyne & Epistemic Democracy

Michael A. Neblo
Assistant Professor of Political Science
Ohio State University
neblo.1@osu.edu


N.B.: This paper is not really the one that I intended to write, nor is it complete. It consists more in an over-long setup for claiming that sophrosyne is a key deliberative democratic virtue, and only briefly sketches the direct argument itself. I plan to develop the notes in the last section into a stand-alone piece working out the detailed argument. Please do not cite or circulate.
Rule #20 Unethical Play: Poker is an individual game. Any action or chat intended to help another player is unethical and is prohibited (Tournament Rules of Poker).

In poker the interest of the player lies in preventing “signaling” – i.e., the spreading of information to the opponent. This is usually achieved by irregular and seemingly illogical behavior...signaling lead[s] to a delicate problem of balancing in actual playing, i.e., in the process of trying to define “good,” “rational” playing (von Neumann and Morgenstern, 1944:54).

President Harry Truman was an avid poker player most of his adult life, though he knew that the game was considered unseemly and antisocial by many, especially for a respectable, public man like himself. Truman was unbowed, however. He wrote to his future wife Bess, explaining that though he considered himself religious, “I like to play cards…but I don’t feel badly about it. I go when I feel like it and the good church members are glad to hear what it’s like” (Truman, 1983: 23). Though part of the attraction was fraternal, Truman took the game very seriously on its own terms, and was an able player by all accounts. In March of 1946, he rode the train with Winston Churchill from Washington to Missouri, where Churchill delivered his legendary “Iron Curtain” speech. The night before the speech, the two played poker in Truman’s train car, along with some of their aides and a few journalists. David Brinkley, then a young reporter sent to cover the speech, recounts that “[A]fter we’d been playing for a while, it was evident Churchill didn’t know the game very well ... When Churchill excused himself to go to the bathroom, Truman said to the rest of us: ‘This man saved the free world. Lose.’ So the rest of the night, we were folding with flushes and three of a kind” (Brinkley, 1996).

---

1 The original quotation contained a fair bit of mathematical notation that would have been distracting in the present context, so I have edited without ellipses in a few places, without changing the meaning materially.
Strictly on poker terms, the rationale for Truman’s order is unintelligible, if not “unethical.” Like other social practices, poker has its own norms and virtues. But they are very peculiar norms and virtues. A big part of what makes the game distinctive and enjoyable is that many of the norms that regulate most social interaction are not only relaxed, they are positively inverted. Truman’s moralistic critics struck upon a grain of truth – in a certain sense the game is antisocial. In tournament play it is considered unethical, not merely imprudent, to try to help another player. Similarly, telling the truth about one’s cards can be sanctioned, while lying is protected. And similar social inversions emerge in friendly pick-up games as well. Had Churchill known that the other players were trying to assist him, rather than manipulate and defeat him, he would not have felt grateful, but rather embarrassed or insulted.²

Indeed, if mundane human beneficence is a prominent factor in the game, you may be teaching someone to play poker or trying to help an honored guest save face. But when players are intentionally folding with flushes and three of a kind, there is an important sense in which they have ceased playing poker, no matter what they seem to be doing with their hands, and cards, and chips. And they have ceased playing poker precisely in virtue of substituting some other set of norms and goals for the norms and goals of poker – i.e., trying to maximize one’s

² I happen to enjoy poker myself, and consider the peculiar sense in which it is antisocial harmless in most cases. On the point about lying and truth telling: consider the following reply (from a poker discussion board) to a post expressing incredulity about the way this rule was applied in a particular tournament: DCJ001: “If you sincerely don’t understand how a player can lie about his holding, but if he tells the truth he is violating a Tournament Directors’ Association rule, you are either new to poker, or …you have not read and made an effort to understand the TDA rules...If a player truthfully announces the cards that he holds, it’s the same as showing his cards, which is against the rules. And, if he is lying, he has not shown his cards. Not showing one’s hand is not against the rules.” http://www.pokertda.com/forum/index.php?topic=219.15 In my view, the real oddity about this rule is that, on the grounds in the epigraph from von Neumann and Morgenstern, enforcing it ends up conveying more information when a player speaks, since the other players can confidently infer the negation of the speech. I should also note that part of the injunction against helping other players is to prevent collusion. But that is only a part of the rationale. One need not intend to benefit, actually benefit, or potentially benefit personally to run afoul of the rule.
individual gains, positively excluding regard for others. Absent such norms and goals, “trying to define ‘good,’ ‘rational’ playing,” becomes much more than just “a delicate problem.” Without seeing the folded cards and knowing a fair bit about poker, a third party might have had difficulty even knowing that Truman’s game had changed utterly.

Since von Neumann and Morgenstern’s 1944 classic, *Theory of Games and Economic Behavior*, poker has served as perhaps the paradigmatic case to motivate intuitions and provide examples for game theoretic concepts. And that makes good sense. It represents something close to unambiguous zero-sum strategic conflict, with well-defined rules, and consistent, knowable motivations under normal circumstances. But, as noted, poker is really quite peculiar vis-à-vis most other forms of social interaction. Nonetheless, poker (along with chess) is forever being dragooned into serving as a metaphor for politics. Indeed, Churchill’s “Iron Curtain” speech contains such echoes, and Truman’s press release announcing the Hiroshima bombing refers to the technological race against the Germans as having wagered “two billion dollars on the greatest scientific gamble in history — and won.”

And yet the comparisons are often inapt in crucial respects. Real political problems are typically not well defined, interpretations of the structure of the game vary across players and change over time, the operative social norms and individual motivations tend to be quite different from poker, etc. None of this is to say that simple game theoretic models cannot

---

3 In fact, Truman gradually informed the press corps about the Manhattan Project and the impending attacks on Japan during a series of poker games aboard the U.S.S. Augusta, while on his voyage to the Potsdam Conference (Ferrell, 1994:204). The press release had been written before he left, but was announced while he was en route home. To my ear (and with the benefit of historical distance), the connection to poker creates a whiff of glibness in the triumphalism that otherwise might have been forgivable given the scale of the conflict that was beginning to wind down. Secretary of State James Byrnes is reported to have had related concerns. A reporter remarked that Truman “was running a straight stud filibuster against his own Secretary of State” (McManus, 2009).
capture important parts of more complex phenomena, nor that rational choice cannot accommodate altruistic preferences, expressive values, inequality aversion, a taste for telling the truth, and the like. But there is an elegance and aptness to the way that game theory captures the heart of poker without ad hoc emendations. In principle, one could represent things like Churchill’s vivid appeal to the shared identity and values of the “English speaking peoples” or Truman’s orders to switch from playing poker to playing gracious host. But doing so is not likely to be very enlightening without a partnering theory to help us understand when and how we come to be playing the games that we actually play. Indeed, it may sometimes add to our confusion, rather than resolve it.

As a case in point, game theory, social choice theory, and democratic theory have developed an unnecessarily fraught relationship over the last few decades. Formal theorists have developed an impressive set of results that many believe severely constrain the ambitions of normative theories of democracy. Though there are antecedents, this project got going in earnest with Kenneth Arrow’s famous analysis of social welfare functions (e.g., voting rules). In recent years, this literature has grown to include formal accounts of specifically deliberative theories of democracy as well. Though there are many complications, it is fair to summarize the take-away as generally critical and negative: normative theorists seem to consistently underestimate the obstacles and limits involved in realizing their normative ambitions. For the most part, deliberative democrats have responded by trying to show that the formal results are largely irrelevant, by lowering their normative standards, or by trying to fortify specific institutions against the brunt of the critique.
William Riker, for example, argued that his “confrontation between the theory of democracy and the theory of social choice” ends in a decisive victory for social choice. Popular democracy is at best “meaningless” and at worst leads to “hatreds and oppression” (Riker, 1982: 253). All that is left for democracy is the bland and periodic function of throwing elites out of power. Many others agree that social choice theory is fatal to democratic theory’s ambitions. And, since the social choice results are formal proofs, some take the conflict as settled permanently. Democratic theorists responded to this challenge from social choice theory, but the early responses tended to take the form of a wholesale rejection of the rational choice paradigm. Yet, given the existence of reasonable disagreement, modern complex democracies will need recourse to voting, and social choice theory has a lot to teach democratic theorists about voting. Social choice theory without democratic theory is normatively blind, but democratic theory without social choice theory is institutionally deaf.

There has been a whole host of responses to Riker, but none more forceful than Gerry Mackie’s exhaustive rebuttal in Democracy Defended. Rather than evading the criticisms, Mackie shows that Riker’s conclusions do not follow, largely on their own terms. He concludes that many social choice theorists have systematically misinterpreted the meaning and applicability of the theorems that they have proven.4

I am highly sympathetic to Mackie’s arguments, and think that he is largely successful in the task that he sets himself – i.e., to show that democratic minimalism does not directly follow from the social choice results. He marshals a remarkably powerful and thorough defense of our

---

4 This section draws on Mackie (2003) and my own similar, though much less comprehensive, discussion of Riker in Neblo (2000).
folk understanding of voting, and puts the burden squarely back on democracy’s critics. And that is no small accomplishment. That said, his arguments offer us little guidance for going forward with respect to either social science or political reform in practice, beyond suggestions to increase the use of the Borda rule, among a few other ideas. Moreover, his arguments are vulnerable to a second round of formal critique, at least vis-à-vis specifically deliberative versions of democracy.

For many theorists, one of the main attractions of deliberative democracy was that it seemed to provide a way to side-step the whole setup of this debate. Democracy is not primarily about preferences and preference aggregation, so any purported problems in meaningful preference aggregation were moot. As a simple empirical matter, though, real political deliberation over controversial political issues rarely issues in anything like a full consensus. Whether because of a deep pluralism of values, differences in our culture and socialization, or because of simple human finitude in our cognitive or volitional capacities, we rarely reach consensus, and have to recur to some other decision rule such as voting. Thus deliberative theories rapidly found themselves back in the same boat as aggregative theories, trying to withstand the force of the social choice and game theoretic critiques of their coherence and workability. Mackie’s arguments would seem to provide significant comfort for deliberative democracy on this front, and vice versa – i.e., deliberation may be one of the ways that we can induce the similarity in preferences that leads to stability.

Ironically, however, by attending to the specifically deliberative concerns of democratic theorists, it becomes apparent that Mackie’s arguments only pass the buck down to “the means by which the majority becomes the majority.” Any gains in vindicating the normative
meaningfulness of voting are utterly dependent on successfully mounting a similar defense of deliberation per se. Upon reflection, this insight should not be too surprising. But it should alert us to the fact that what was supposed to be an escape route can also serve as a second entrance for the original trouble.

In a more recent wave of research, formal theorists have shown that many of the same incentive and coherence problems that threatened the meaningfulness of voting appear endemic to the process of exchanging, assessing, and aggregating information and reasoning as well (Austen-Smith and Feddersen, 2006; Meirowitz, 2007). That is, deliberation is not immune to the formal critique. Even if we all want the same thing – to convict the guilty and exonerate the innocent, for example – strategic dynamics can interfere with our ability to reason together productively under a remarkably wide range of circumstances (Landa & Meirowitz, 2009). The situation only degenerates further as partially conflicting interests begin to come into play. By the time that we reach predominantly conflictual situations, cheap talk abounds (Calvert, 2006, Meirowitz, 2006). We find ourselves stuck in a democratic Babel of belief.

These results threaten to throw us back into the Hobbesian interpretation of voting. The whole point of the Arrow conditions, of course, is to build modest normative features beyond Hobbesian stability into democratic voting. Pareto efficiency is a kind of minimal welfare condition. Non-dictatorship guards against being subject to arbitrary power. Independence of irrelevant alternatives embodies a kind of respect for the separateness of persons in excluding interpersonal comparisons of utility. And unrestricted domain guarantees extensive room for pluralistic values.
However, if we do not know whether the majority became a majority via passably good deliberation, then there is little reason to think that the outcome promotes even these modest goods. We would have little warrant, for example, to believe that the choice really is Pareto efficient unless we are willing to take any old preference report at face value. As we saw in the previous chapter, however, a large and powerful literature suggests that it is more the exception than the rule for people to reliably know their own minds about politics ex ante. So absent good deliberation it would be question begging to appeal to the sovereignty of preferences in establishing the welfare improving properties of democratic votes. Similarly, if my fellow citizens form their preferences in a haphazard or power-laden way, then non-dictatorship hardly protects me from being subject to arbitrary power. Formal equality is cold comfort if my coequals are coercing me on the basis of vague impulses, or worse. And if we are dependent on the way that elites frame politics, then any goods flowing from the independence of irrelevant alternatives go out the window, since invariance under framing practically defines the criterion. Finally, the prospect of unrestricted domain is positively alarming in the presence of demagogues – radical pluralism becomes a pretext to pander to or draw out the worst in us.

Thus, democratic stability, agreement between voting rules, and the other apparent goods that Mackie shows that we can get with some reliability reduce down to Hobbesian decisiveness, absent any evidence of reasonably good deliberation. So we are left without a theoretically unified account of why instability is not ubiquitous, whether the same problems end up re-entering through the back door of deliberative quality, nor much in the way of advice on how to improve. We do get several reasons to think that things are not likely to be that bad. Natural human sympathy means that people will take some regard for each other. The civilizing force of
hypocrisy will constrain most people’s worst motives. A certain degree of cultural similarity will induce concomitant similarity in their preference profiles, etc. One might think that we should not let a focus on abstract possibilities magnify the problem out of proportion when common sense tells us that things are mostly OK.

As it happens, I do not believe that things are *that* bad. But serious people present arguments that the strategic implications for deliberation are that bad, and that an assessment that does not let our hopes tutor our intuitions will reveal it so. Other serious people argue that we do not really know, and so we should like a more general explanation of why things go better (if they go better) than we might think. And in any case, a more integrated theoretical account, beyond its academic value to both social science and normative theory, would be valuable in helping us to do better as a matter of practice, even if we are already doing reasonably well.

Mackie conducted his battle with formal theory as something of a scorched earth campaign, leaving us in a poor position to re-engage in order to assess what the extent theories can tell us. None of the original authors of the seminal papers (e.g. Arrow) thought that they were showing that democracy was meaningless, even if some of their epigones interpreted them that way. Indeed, they generally hoped to develop research going forward to help democracies function better. An inferentialist account of deliberation holds out the promise of providing a more integrated theoretical account of how and why voting, deliberation and other democratic processes manage to avoid the dismal implications of rampant instability predicted by some formal critics of democracy.

2. Buying into the Game
In poker, one typically has to “buy in” to a game. In games with a betting limit, the buy-in is substantial, usually ten times the maximum bet. And once in a game, many poker forms require that players “ante up” or make “blind” bets for each individual hand as well. The idea is that players should have a significant stake in the larger game before the local action even gets started. They cannot dip in and out of the game at will, and must pay some cost to do so for individual hands. These sorts of rules are necessary to keep the players meaningfully engaged, induce the right kind of motivations, and reward skillful play by spreading the vagaries of chance over a larger number of hands.

Without trying to push the analogy too far, I want to suggest that something like the concept of a buy-in can illuminate the way that deliberation helps to stabilize political interaction, and to invest it with normative content. The philosopher Robert Brandom has described “the core of discursive practice as the game of giving and asking for reasons” (1994:159). Like other games, the game of giving and asking for reasons has rules (even if, again, like other games, they are sometimes implicit), and requires that we apply and interpret those rules in the course of playing the game. That is to say, discourse is a normative practice. The social meaning of the words that we speak or write and the actions that we take or eschew emerge out of this practice.

The poker analogy operates at both the level of buying in to the larger game and at the level of making blind bets or antes. That is, except in extreme cases – sociopaths and people with severe autism – all adults have already bought in to the larger game of giving and asking for reasons. I hasten to add that this claim does not imply that there is no variance in the level of buy in among adults. It only means that, excepting rare cases, everyone’s buy-in to the larger
game is well above zero. Children buy in gradually as they become socialized, and thus exhibit variance as well. Such variation across adults and over-time in children means that there are testable implications of the claim about game-level buy-in. It is not a *deus ex machina* invoked to redeem us from political perdition. Similarly, most particular “hands” of the game of giving and asking for reasons call for an ante or a blind bet. And here, we have much more variation – among individuals, social situations, the inter-action between the two, over-time, across cultures, “objects” of the reasoning game, etc.

Though Brandom’s ideas developed via paths internal to philosophy, the broad outlines of his approach have received substantial support from a remarkable range of independent sources, including evolutionary anthropology, developmental psychology, and cognitive linguistics. Developmental psychologist Michael Tomasello and his colleagues (2005), for example, show that our ability to understand other people’s intentions is not enough to explain our capacity for language, mutual coordination, and a whole range of cultural accomplishments. Contrary to previous assumptions, nonhuman species (e.g., chimpanzees) are quite capable of understanding the intentions of other actors (675).

Similarly, people with autism were once thought to be lacking in such capacities, but it turns out that our previous understanding was subtly mistaken (686). For example, while it is true that children with autism tend not to follow the gaze of other people spontaneously, if they are specifically asked to assess what others are attending to, they are quite able to do so. The difference between children with severe autism and those without lies less in their ability to understand intentional action than in their motivation to and practice with sharing experiences and emotions with others. From a very early developmental stage, most children evince a
remarkably strong drive to *share* psychological states with other people – that is, to interpret others’ internal states (to understand), to express their own internal states (to be understood), and to bring the two into alignment (to mirror or share experiences and intentions). This process is essential for language acquisition and more generally for successful socialization, which is why children with autism typically experience difficulty with both. And what starts out as a relative lack of drive develops into a lack of skill in managing shared intention in complex social environments.

What such deficits throw into relief, however, is the powerful and pervasive legacy of the developmental path that the vast majority of us take into the social world. Tomasello et al. describe the main achievement of this process as an older child’s ability to manage what they call “dialogic cognitive representations” (2005: 675). In effect, the game of giving and asking for reasons is the adult manifestation (and vehicle) of this continuing drive toward shared meaning, experience, and intention.⁵

This is not to say that adults cannot kick the ladder out after they climb up, so to speak. Part of becoming socially competent involves learning how to deploy social meaning for more narrowly strategic purposes. But for most of us, both the medium itself and the legacy of emerging into the social world via this path means that we typically have at least minimal normative buy-in to individual “hands” in the game of giving and asking for reasons. Indeed, one might aptly distinguish sociopaths as people who have completely “kicked the ladder out.” That is, they have mastered the social skills surrounding shared intention – they can understand

---

⁵ These developmental findings comport remarkably well with Habermas’s notions of an “orientation toward understanding” and “the weak motivating force of good reasons.” Similarly, Habermas anticipated important elements of Brandom’s project.
and effectively manipulate social situations – but do not give the typical normative buy-in any intrinsic weight in their individual deliberations, no matter the situation. In particular, sociopaths seem unresponsive to socially articulating emotions. They may understand how various emotions typically map into social situations, but they exhibit no motivation to align them with others (Blair et. al., 2005). Since emotions often serve as affective markers of evaluative judgments, sociopaths end up wildly out of alignment regarding norms of feeling remorse, indignation, and other socially evaluative emotions.\(^6\) Happily, though, few of us are sociopaths. So in most situations we retain strong motivations to coordinate our actions (at least in part) by socially articulating the reasons for and against those actions. That is, we seek to justify ourselves to each other.

3. Rational Expectations

Game theorists have developed a substantial literature analyzing the potential for strategic dynamics in such deliberative attempts at mutual justification, much of it at odds with the setup and claims of normative theorists of deliberation. Landa and Meirowitz (2009) lament the lack of fruitful interaction between the two camps, attributing it to the fact that “game theorists tend to exert little effort toward making their results accessible to a less technical readership … [and that] the normative literature, with very few exceptions, takes essentially no account of the presence of the game-theoretic work on deliberation and ignores the fundamental incentive problems that surface in nearly all relevant game-theoretic studies” (428). In an

\(^6\) As a clinical matter, the older terms “sociopath” and “psychopath” have been more or less absorbed into the DSM-IV diagnosis of Anti-Social Personality Disorder (ASPD). I use the older term both because it is less clumsy, and because I want to evoke a phenomenon that shades out of the clinical into more general experience.
important and ambitious project, they aim to summarize and translate this body of results into non-technical terms in order to “make the case for the relevance of the existing game-theoretic analysis of deliberation to the development of deliberative democratic theory” (428).

Landa and Meirowitz motivate their central complaint against deliberative theory by noting that “A key issue in game-theoretic models of policymaking is whether it is reasonable to expect those participants who possess valuable information to reveal it to others, and whether those others have good reasons to believe it” (431). Despite the implied contrast, however, we can also aptly render one of the key issues in deliberative conceptions of policymaking in exactly the same terms: whether it is reasonable to expect those who possess valuable information to reveal it to others, and whether those others have good reasons to believe it. However, in the deliberative version, the significance of the key words in these sentences changes subtly from the way that game theorists typically use them: “reasonable,” “expect,” “possess,” “valuable”, “information,” “good,” “reasons,” and “believe.”

For example, note the ambiguity between the predictive and the normative meanings of “reasonable” and “expect.” I teach my daughters that it is “reasonable” (in the normative sense) to “expect” (in the normative sense) them to use good manners, even though it is not “reasonable” (in the evidentiary sense) to “expect” them (in the predictive sense) to do so consistently while they are still young. And yet, over time, it is reasonable to expect (in the predictive sense) that they will come to use good manners consistently, precisely because they will come to see that such behavior is a reasonable social expectation (in the normative sense). At first their motivations may be rooted in a desire to please me or to avoid sanctions. But unless something goes wrong, they will eventually come to recognize that they typically have good
reasons to refrain from rude behavior, even when they feel a desire to engage in it and do not anticipate offsetting social sanctions.  

So it is emphatically not the case that normative theorists are unconcerned about game theorists’ key issue regarding what it is reasonable to expect in political communication. Rather, normative theorists implicitly argue that the two senses of “reasonable” and “expect” are often internally related in a way that most extant game theoretic models of political talk do not capture adequately. Indeed, for present purposes, I will define a norm as a pattern of action that is reasonable to expect of most people (in the predictive sense) largely because most people believe it to be a reasonable expectation (in the normative sense).

To illustrate their point about reasonable expectations, Landa and Meirowitz note that, “In the game of poker, which has some strategic properties in common with the example of the decision process regarding the invasion of Iraq, expecting truthful revelation of someone’s face-down cards is unreasonable” (431). Invoking the poker analogy here usefully illustrates one of the central problems with extant models. Far from serving as a sensible baseline for most social

---

7 Though I will admit that one of the peculiarities of raising a child in the midst of “the terrible two’s” is that it often seems all too plausible that they will never become socialized adults, despite the reassurances offered by low baseline rates of adult sociopathy. I should also note that the process can work in the other direction as well, when adults unlearn socialization to unjust norms. For example, adult males might subconsciously expect (in both senses) that a woman will defer to them in conversation because of implicit sexist norms common when they were socialized, and only somewhat less so now (Schegloff, 2000). If so, then they may react more negatively to being interrupted by a woman than they would a man. Yet they may not be willing to defend those norms as reasonable if they are made consciously available to them through political thematization, and the unconscious reaction may lose its force because the gap in one’s commitments makes it to conscious attention, or is subject to criticism via political articulation.

8 This definition of norms is a bit more restrictive than one might want in that it cannot capture implicit norms, for example, about invasions of personal space. That said, it does a reasonable job capturing the main class of norms relevant for my discussion here, and highlights the cognitive, rather than just behavioral, dimension necessary to make sense of them in the context of deliberation. In another potential abuse of terminology, I include under the term “game theory” (and, mutatis mutandis, its cognates) any mathematical model of rational choice or bounded rational choice. So, in addition to game theory proper, I mean to include social choice theory, mechanism design, implementation theory, evolutionary game theory, and computational and agent based models. Similarly, I use normative theorist, deliberative theorist, and deliberative democrat interchangeably here.
processes, poker is an aggressively limiting case that positively *inverts* otherwise ubiquitous norms and motivations. In the same way that a chess match would not count as chess if rooks could move diagonally, the Churchill anecdote illustrates how the meaningfulness of poker as a distinct social practice (i.e., a game) depends on excluding pervasive social norms and motivations.

Thus, arguing that poker shares “some strategic properties” with the Bush administration’s putative deception about weapons of mass destruction in Iraq, while technically true, is quite misleading. When we failed to find such weapons (i.e., when “the cards were revealed”) opponents of the administration did not gamely congratulate them on “winning the hand” via deft strategic manipulation. Indeed, to this day, former members of the Bush Administration deny that it engaged in any deception, and describe the events as a mundane failure caused by human finitude (i.e., in a straight game of giving and asking for reasons). Perhaps their account is true, perhaps officials are sincere but self-deceived, or perhaps this episode is a case of the tribute that vice pays to virtue. The important point here is that a significant part of what makes game theoretic models interesting is that their implications do not typically generalize in simple, foreseeable ways when games only share some properties. And if my argument about the ubiquity of deliberative buy-in, antes, and blind bets is right, then models that do not take account of them are liable to go wrong in important ways, both as explanatory accounts and for purposes of a normative theory of institutional design.9

---

9 Even a shift to implementation theory and mechanism design will not address this version of the sensitivity problem; the whole class of models will exclude norms. And even if the Iraq war example is a case of the tribute that vice pays to virtue, such tribute still points to the role of social norms. If the administration’s opponents and the public were all just playing their role in a larger strategic game without norms, then one of the key papers in this literature, Austen-Smith and Feddersen (2006), implies that if they are right, the pretense of deliberation would be odd and inexplicable, since deliberation would be uninformative (without, presumably, being completely costless).
Landa and Meirowitz do address themselves to one important facet of deliberative buy-in, and it is worth analyzing their response in some detail. They argue against the reasonableness of presuming that many people default to a basic normative commitment to honest deliberation in a wide variety of situations, arguing that “[I]n a complex environment with private moral values and noncommon veridicality, it must presuppose not simply abnegation of self-interest, but an extraordinarily high degree of abnegation of moral instrumental behavior as well” (2009: 441). On an inferentialist account, private moral values, noncommon veridicality (i.e., differing standards for judging the truth), and raw self-interest all end up looking structurally similar: they are all examples of an agent’s proposed reasons for a collective choice that prove unpersuasive to others in deliberation. What makes private moral values private is that they cannot be justified on public grounds; what makes noncommon veridicality noncommon is that my standards of judgment cannot compel general assent; and what makes raw self-interest both raw and selfish is that it cannot claim grounds beyond my mere assertion of will. In the game of giving and asking for reasons, all three tend to be relatively weak reasons. Now, I grant that people typically act out of self-interest, but I also claim that they typically experience significant motivation to justify themselves to each other.

Landa and Meirowitz’s more pregnant claim, then, is about “moral instrumental behavior.” What does it mean to say that a norm of honest deliberation presumes an extraordinarily high degree of abnegation of moral instrumental behavior? Contrary to what they suggest, I would argue that most social and political situations start with a rather strong presumption against “moral instrumental behavior,” since that is just another term for a familiar concept – paternalism. I know what is best for you (or for society), and even though I cannot
convince you, I do not hesitate to strategically manipulate the situation for your (or the public’s) own good.

Sometimes, of course, we do behave paternalistically, and, on occasion, justifiably so. For example, I routinely find myself incapable of reaching agreement with my three year old daughter. In some of those cases, I may withhold information or otherwise strategically manage our discussion out of instrumental altruism. But the set of circumstances in which I would think myself justified in doing the same to my wife or my fellow citizens is much smaller and more peculiar. All of this is just to say that a defeasible presumption against paternalistic manipulation is a generic part of the buy-in to the game of giving and asking for reasons. If so, then it hardly presumes “extraordinary” restraint, and the burden of proof runs in exactly the opposite direction. We need specific reasons to think ourselves reasonable in acting paternalistically, and most of us default to not doing so in a wide variety of situations. If so, insisting that we must always “earn the sincerity by reconstructing it as equilibrium behavior rather than assuming it by default” will tend to lead us to less accurate models by sending us on a search for deep, structural explanations that do not track the real dynamics commonly at play, thereby generating misleading “normative argument[s] for institutional design” (442).

To see why, consider one of Landa and Meirowitz’s more engaging examples. They astutely point out that, “One of the central issues of contention among deliberative democrats is the expectation of consensus following deliberation,” concluding with an ironic twist that “the very fact that the scholars of deliberative democracy disagree on this point strongly suggests that the assumption of common values is untenable” (435-6). But as we saw above, the word “expectation” is ambiguous, even in ordinary language. Recall that Habermas referred to the one
right answer thesis as a “pragmatic presupposition.” He certainly intends “expectation” in the normative rather than the predictive sense. Just as poker is no longer poker if we are not trying to win, rational discussion is no longer rational discussion if we are not trying to find the right answer. The “expectation of consensus” is not an empirical claim about the distribution of agreement after deliberation, but rather a conceptual claim about how to define different phenomena properly. The fact that scholars of deliberative democracy disagree about that claim is, therefore, hardly dispositive for their point.

For example, Dennis Thompson, an eminent deliberative theorist, strongly disagrees with Habermas’s analysis of the consensus criterion (Gutmann and Thompson, 2004). But what is the import of this fact for judging Thompson’s future contributions to scholarly deliberation about deliberation. Suppose that he were asked to review a paper that develops some new bit of social-psychological research suggesting that the consensus criterion may be better supported than we previously thought. In this hypothetical scenario, his judgment is that the research is sound, but he also expects that the broader scholarly community will give the new evidence more weight than he believes is appropriate, persuading more people to Habermas’s (for Thompson, importantly mistaken) view. Would he write a dishonest review so that the evidence is suppressed or appears in a less prominent venue? Would he knowingly misrepresent Habermas’s position about the philosophical grounds and meaning of the consensus argument so as to confuse the issue among non-experts?

It seems quite reasonable to expect (in both the normative and predictive senses) that

---

10 To be clear, this example is entirely hypothetical, and intended to illustrate how jarring it is for us to have to consider such scenarios. Thompson has authored a book on professional ethics, and has an impeccable professional reputation.
Thompson would not manipulate the review. Moreover, those expectations are not grounded merely in thinking that he might worry about getting caught and paying a steep price in his professional reputation. Instead, I expect that he would not lie in his review because I believe both that he should not do it, and that he would recognize that he should not do it. This shared expectation articulates a professional norm against such behavior in doing academic reviews – one that is rooted partly in a presumption against paternalism. So I do not think that his restraint would evince any extraordinary abnegation of moral instrumental behavior.

On the contrary, it would be quite ordinary. The academic enterprise relies on individuals exercising their own judgment within a community of scholars. Paternalism subverts the process by which others form their judgments, and substitutes one’s own judgment for the community’s via strategic manipulation. The game of giving and asking for reasons, as played in the academy, would implode if such restraint were anything but mundane – which is not to say that it is perfectly realized. To the extent that I think that the answer to many important questions is procedurally defined (at least in part) – e.g., whether justice was served by “good” jury deliberations in a liability case – strategic manipulation can be logically self-defeating. I cannot promote some good via means that subvert the original logic of what I use to judge that good.

Yet a very large swath of the formal critique of deliberative democracy implies that Thompson’s (presumed) candor should not be mundane – that his restraint would be mysterious, if not irrational (Austen-Smith, 1990; Austen-Smith and Banks, 1996; Austen-Smith and Feddersen, 2006, 2009; Calvert, 2006; Feddersen and Pessendorfer, 1998, 1999a, 1999b; Gerardi and Yariv, 2007; Meirowitz 2006, 2007). Even if we are purely altruistic (or we all want the
same thing), on this view, whatever our best assessment of the public good is, we should be expected (in both senses) to pursue it with full paternalistic abandon. Austen-Smith and Banks (1996), for example, rely on a setup that is nearly identical to the situation in which I imagined Thompson, since journal reviewers can be conceived of as a kind of jury. Positing a simple taste for the truth or against paternalism – adding a constant to people’s utility functions – will not really resolve the issue, since we would then just act with “meta-paternalistic” abandon on the new assessment of utility, with no way out of a regress.

Of course, the logic of such a regress is part of why game theorists believe that the problem is harder than deliberative theorists have recognized. But the problem is rooted in an unmotivated consequentialist bias. Landa and Meirowitz argue that “the existing game-theoretic analysis of deliberation” should influence “the development of deliberative democratic theory” and guide institutional design (428). Yet I cannot imagine that they or any of these authors actually teach (or expect, in either sense) their graduate students to make a habit of writing journal reviews in a way that is consistent with the deliberative logic of the models: strategically withholding key information, consciously distorting assessments, voting “as if pivotal,” etc. To the extent that scholars do engage in such behavior (which, I grant, may be substantial), it seems much more likely that they do so out of simple self-interest rather than out of a desire to serve the greater good of the scholarly community. If so, nearly all of these papers emerged out of a deliberative process ill captured by the models that they contain.11

11 Indeed, I take it as a corollary of Austen-Smith and Feddersen (2006) that peer review would unravel if their model did capture its dynamics. None of this is to deny that reviewers can be biased, venal, lazy, self-interested, and paternalistic. But if such considerations completely swamped a basic tendency to report our candid assessments, the entire academic enterprise would collapse. Since I do not think that the enterprise has collapsed, it is hardly naïve to believe that basic academic norms play too important a role in deliberation to be ignored. In my view, bias is
Existing models cannot make sense of deliberative democrats’ concern over legitimate procedures. Landa and Meirowitz foresee this objection, and try to pre-empt it:

[I]n order for the argument [from legitimacy] to have bite, it must effectively endorse the following two claims: (1) interest in the perception of the outcome as legitimate overrides whatever interest [the actor] may have in revealing less information; and (2) perception of the outcome as legitimate is responsive to how much information [the actor] reveals...[and] that it is known when [the actor] has relevant information and refuses to share it (441).

They argue that the first claim runs into their objection regarding “moral instrumental behavior.” But, as we have seen, there is nothing extraordinary in believing that people regard paternalistic manipulation as a costly and, in this context, even a logically incoherent mode of default behavior. Indeed, the burden of proof may run in the other direction for a large range of cases.

Regarding the second claim, they argue that we will not typically know whether someone has acted strategically, and thus it is implausible to think that acting paternalistically will decisively affect perceptions of legitimacy. This response evinces a deep misunderstanding of the argument from legitimacy. The idea is not merely that we want outcomes to be perceived as legitimate (whether or not the perceptions are rooted in lies and manipulation). Instead, we want the outcomes to be perceived as legitimate because they really are legitimate. Again, we seek an internal connection between the two senses of our “expectation” of legitimacy. And we regard the test of that legitimacy as largely procedural.

probably the most important issue, and paternalism the least important in peer review, whereas these formal results on deliberation focus mostly on the latter.
If this all seems too abstract, consider the following example. As part of my research, I sometimes observe deliberative field experiments, silently sitting in on small group discussions. Toward the end of one session, the group could not reach consensus, and decided to vote instead. A somewhat cantankerous participant who had been defending the minority position ended up voting with the majority, much to everyone’s surprise. The group member who had argued the most with him during the discussion asked why he ended up supporting the group’s position. He explained: “If I were king, I would still choose [the other policy]. But I’m glad not to have that responsibility…[pause] …And after spending the day together I bet you’re all glad I’m not king too!” Everyone laughed, but he was making a serious and important point. He was not the sort to be bullied easily into mere conformity, but neither was he the sort to think that his individual judgment, removed from the context of the larger deliberative process, should be the relevant standard of social choice. If so, then paternalism – even if an actor can secure perceived legitimacy – will rarely appear as a sensible (or perhaps even coherent) principle of action.\textsuperscript{12} [Link to democratic interpretation of \textit{sophrosyne} in both the sense of restraint in pursuing one’s own appetites and in the sense of epistemic humility.]

This fundamental tendency to conflate legitimacy and perceived legitimacy is common in the formal literature. For example, Patty (2008) develops an “argument based” theory of collective choice that connects preferences over outcomes to rationales for those outcomes. On \textsuperscript{12} In my experience, the sort of behavior and the rationale offered for it by this participant is quite common in deliberative forums among regular citizens. After pressing their cases vigorously, participants routinely exhibit a kind of humility about their own judgments (or perhaps trust or investment in the procedures) for purposes of making a group decision that seems quite different than mere social conformity. Fishkin’s Deliberative Opinion Polls do not provide as much scope for the phenomenon to emerge, because his small groups are never called on to make group decisions per se. In my view, this phenomenon is of enormous potential significance, and deserves greater theoretical and empirical scrutiny. In preliminary work, I have proposed linking it to the ancient cardinal virtue of \textit{sophrosyne}, which is often translated as “temperance” or “sound-mindedness,” though the concept really does not have a clear modern referent (Neblo, n.d.).
its own terms, the argument is novel, ingenious, and generative. And on the surface, the theory appears to be a promising way to formalize the way that deliberative democrats seek to promote legitimate outcomes by connecting them to processes of public reasoning. Yet, rather than developing a theory of deliberative choice, the model really amounts to a pure theory of sophistry. In the model, actors choose their rationales with no regard at all to their internal persuasive force or logical entailments to the outcomes, but only because they prove effective in helping them achieve fixed, pre-given goals. This setup is practically the definition of sophistry, in that sophists actually practiced switching sides in an argument, and otherwise training their skills at manipulating arguments to serve external goals (Jacquette, 2007). To the extent that there is any inferential articulation in this setup, its goal is purely formal and pre-determined. In a related project, Patty and Penn (2010) use structurally similar logic to develop “a social choice theory of legitimacy.” Principles “legitimate” the processes by which social choices are made, except the connection between principles and processes is purely formal, rather than operating in any substantively “principled” way. This setup equivocates on the meaning of “legitimate” as deliberative theorists use the term. It is hard to see why legitimacy so defined would be anything more than mere legitimation (achieving the perception of legitimacy by any means). Indeed, it is not clear how the process would even generate perceptions of legitimacy: why should I feel respected by a process that offers me public rationales that I know are always mere rationalizations?

4. Why Rules Are Rules
Much of this might seem familiar by now, but it is important to appreciate fully both the complications that mundane normative processes cause for extant formal models of deliberation, and the way that words and concepts fundamentally shift meaning when we move back and forth between the two literatures (Steiner, 2008). Within standard models, one could of course just stipulate that agents have an aversion to lying, and build a large cost for doing so into games with communication. But in addition to the regress problem discussed above, we would not have “earned” sincerity, which was supposed to be the whole point in modeling deliberation. As the great game theorist John von Neumann quipped: “With four parameters I can fit an elephant, and with five I can make him wiggle his trunk,” which he intended not as a boast, but as a methodological caution (Dyson, 2004: 297).13

Even if we decided that building in some fixed cost to lying that was not circular and worth the loss of parsimony, doing so would not take account of the shifting contexts in which norms of truthfulness get activated across and even within games. Positing a fixed cost to lying does not explain why we have wildly different expectations about truthfulness in a poker tournament, in a confessional, in talking to our children about Santa Claus, in writing referee reports, or in contract negotiations. On an inferentialist account, most norms are not merely behavioral ticks, but reasons for action subject to dynamic social application and articulation.

13 The context of the quote is a bit complicated, but all the more interesting for it. As a young physicist, Freeman Dyson went to meet with the older and more eminent Enrico Fermi to show him results from Dyson’s lab that connected a novel theory to measurements close to those that Fermi had observed independently. Fermi asked him how many arbitrary parameters he had used to get his calculations. When Dyson reported “four,” Fermi replied with the elephant quote, prefacing it with, “I remember my friend Johnny von Neumann used to say…” Dyson recounts that the quip was a hard blow and a conversation stopper at the time, but that it ultimately “saved us from several more years of fruitless wandering along a road that was leading nowhere” (297).
Reasons (as opposed to mere information) are offered with a specifically articulated set of suggested connections for their significance, rather than merely leaving the listener to figure out their relevance, or to apply the information in a way different from the speaker’s intention, which is not to say that reasons cannot trigger different articulation. Public reasons do the same thing, but via paths that are regarded as articulable in common for the polity. Part of the distinction between reasons and information can be captured by the distinction between “hard” and “soft” information, as those terms are used in game theory. Hard information is something that is intrinsically verified upon inspection – say some mathematical proposition. Soft information relies more upon the credibility of the speaker and the institutional environment for its persuasive force. Cheap talk models mostly operate along the limiting case of soft information, whereas, if the inferentialist account is right, much deliberation effectively deals with various grades of “harder” information.\textsuperscript{14}

But why should normative talk ever be anything but cheap? In the face of countervailing desires, norms and reasons can retain some motivational traction in much the same way that our other beliefs and commitments do. Note that having true factual beliefs is not always strategically beneficial on a standard rational choice account. For example, we might derive both direct psychic benefits as well as strategic advantages from being overly optimistic in some situations. Yet we (and most game-theoretic models) consider it the exception, rather than the rule, to credibly maintain to ourselves and convey to others false belief and irrational intent in

\textsuperscript{14} In practice, the distinction is probably more of a spectrum than a dichotomy, and varies by actor, context, and resource constraints. In principle, I could learn to assess climate science in a way that made it “hard” for me, but as a practical matter, I cannot take years to get a PhD in climatology. One might be tempted to think that “hard” information is necessarily factual, but on inferentialist terms, a norm or any other reliable inference can be “hard.” Or rather, it is something whose support, once placed in an articulated context, is such that it compels acceptance, at least as a reason. Normative claims can have that property as well.
the face of evidence and common knowledge. All of us have had the familiar experience of “not wanting to believe” something, but feeling compelled to do so by new evidence. Similarly, we have all experienced not wanting to do something, but feeling compelled to do it by recognizing some obligation or responsibility. Just as we cannot trivially turn on and off our responsiveness to the evidence of our senses and the laws of logic (norms of reasoning) as local expedience suggests, neither can we trivially turn on and off our responsiveness to social norms, roles, and meanings, nor accountability via public articulation. Our larger normative buy-in is not globally optional as a practical matter, and is locally “sticky” in most cases as well.15

Thus, the call to be “reasonable” and meet those expectations will at least sometimes be experienced as “good reasons” by all but sociopaths. Whether or not they end up being decisive, they become at least a motive for action, without necessarily becoming a desire in the strict sense. Again, we may feel obligated to do things that we do not really want to do in any straightforward sense. This distinction between a desire and a normative reason for action is difficult to capture in extant game-theoretic models, yet the peculiar properties of norms vis-à-vis other kinds of motives reside in this distinction.

Now I hasten to emphasize that I do not believe that self-interest, private (though perhaps not selfish) values, different standards for judgment, and even paternalism are typically marginal

15 It is important to distinguish between motivated reasoning per se, which is quite common, and strategically choosing to engage in motivated reasoning in a way that is credible to others or to ourselves (Elster, 1983). Moreover, while motivated reasoning is quite common indeed, nearly all of the evidence for it is based on deviations from a baseline. That is, we do not typically process reasons and evidence with utter impunity, but rather a certain degree of bias. And accountability to others, the hallmark of deliberation, tends to constraint that bias even further (Mercier and Sperber, 2011). Standard game-theoretic models create a sharp distinction between beliefs and desires (especially with respect to being subject to articulation) by methodological fiat. Minozzi (n.d.) proposes a model allowing beliefs to be motivated by preferences, loosening the distinction in a way that comes from the opposite direction of what I am discussing here. One might think of schizophrenia and sociopathy as representing the limiting cases of beliefs and desires being unconstrained by physical facts and social norms, respectively.
in political discussion or politics more generally. Clearly there are fundamental incentive problems with which any adequate theory of deliberation must grapple. Moreover, some of the very social processes that I have been discussing (e.g., norms and procedural accountability to each other) can actually complicate the incentive problems. Just as Truman’s game continued to look like poker after it stopped really being a game of poker, political discussion about going to war in Iraq may have looked something like rational public deliberation without actually meeting its minimal presuppositions.

Indeed, regular citizens routinely articulate their frustration with elected officials and other elites by accusing them of “playing politics” with public discourse – allowing their behavior to be dominated by sophisticated rhetoric, phony posturing, and veiled self-interest. The implicit contrast is with political deliberation understood as playing the game of giving and asking for reasons. Democratic citizens express frustration because they expect better of their public officials. But here, again, we run up against the ambiguity in the word “expect” in its normative and predictive senses. In effect, realist critics of deliberative democracy claim that we should jettison the normative sense of our expectations about playing politics. “Expecting better” in either sense is unrealistic and immature, and will only lead to frustration when those expectations go consistently unmet.

Realist critics make this claim either because they deny that the game of giving and asking for reasons can be usefully distinguished from playing politics, or because they think that any internal connection between the two senses of expectation is so weak as to be ignorable. Unlike expecting my daughters to behave politely, expecting more of politicians will not lead to better deliberation or policy in the short run or over time. As I have argued above, though, if
there is truly no useful distinction between the games, then we are back in something very much like Hobbes’s world, however much it might look like we are playing at something more appealing on the surface. And the empirical claim seems wildly unsupported if for no other reason than that we have no relevant evidence for the counterfactual. Even authoritarian regimes maintain the pretense of public justification. And a cursory glance across cases suggests that there are hardly obvious benefits to regimes where the population has accommodated itself more thoroughly to mere pretense.

5. Citizen Competence as a Moral Problem for Democracy

In the summer of 1932, a large plurality of German citizens voted for the National Socialist German Workers’ Party. The Nazis garnered more votes than the next two largest parties combined, and other extremist parties also fared well. Combining the votes for the Nazis, the Communists, and the like, a solid majority of citizens voted to elect representatives who were avowedly hostile to liberal democratic constitutionalism. Though the Nazis never formally replaced the Weimar constitution, soon after assuming office they passed the Enabling Act, investing the Chancellor with temporary plenary power. In the ensuing years, the Reichstag dutifully renewed the Enabling Act at the appointed intervals, evincing a concern for democratic form that might seem quaint or comical were the context not so grotesque. A year later, Hindenburg, the German president, died, and the Reichstag called for a plebiscite to ratify a proposal to combine the powers of the Chancellor and the President under Hitler’s leadership. With ninety-five percent of registered German voters turning out, the proposal won ninety percent approval. Over thirty-eight million German citizens voted to invest Hitler with
dictatorial power, with only four million dissenting. On paper, the Weimar constitution seemed like a reasonable blueprint for liberal democracy, but the democratic elements were mobilized to eviscerate the liberal elements, whose demise, in turn, destroyed democracy itself.

Given this context, it is not hard to appreciate why postwar social scientists regarded citizen competence as a topic of overwhelming importance. The initial catastrophic sequence of political events in Germany occurred with chilling rapidity and broad popular support in “what was once regarded as the citadel of Western civilization” (Adorno et. al., 1950: vii). So it became a matter of real urgency to understand how it came to pass “that in a culture of law, order, and reason … great masses of people [could] tolerate the mass extermination of their fellow citizens” (vii).

In their massive study, The Authoritarian Personality, Theodor Adorno and his colleagues sought “to develop and promote an understanding of social-psychological factors which have made it possible for the authoritarian type of man to threaten to replace the individualistic and democratic type prevalent in the past century and a half of our civilization, and of the factors by which this threat may be contained” (xii). The whole point of the project was to integrate normative, scientific, and practical concerns. The argument starts from an openly normative conception of good (democratic) and bad (authoritarian) citizenship, proceeds to develop operational measures of latent dispositions germane to those conceptions, their causal antecedents, and then to the prospects for remedy via policy interventions in educational and political institutions.

Adorno developed the “F Scale” to measure people’s dispositions toward authoritarian or democratic patterns of citizenship. The scale consisted of several factors: authoritarian
submission (being submissive and uncritical toward in-group authorities); authoritarian aggression (a tendency to see those who deviate from in-group ideology as enemies); conventionalism (rigid adherence to dominant in-group values and norms); anti-intraception (opposition to the subjective and imaginative, and thus an unwillingness to consider things from another’s perspective); stereotypy (thinking in rigid categories, intolerance of ambiguity); superstition (mystical sense of causality and hostility to science); hierarchical power (thinking in terms of dominance-submission, strong-weak, leader-follower); projectivity (a tendency toward conspiracy theories and threat sensitivity).

It is easy to see how most of these factors might work against good notions of democratic citizenship. For example, uncritical submission to in-group authorities is precisely the kind of democratic abdication that contributed to Weimar’s fall. Similarly, being unable or unwilling to take up another’s perspective on politics facilitates regarding those who disagree with us as threatening outsiders who should be marginalized. The whole syndrome might make peaceful co-existence in large pluralistic democracies more difficult. That said, some of the other criteria proved controversial, especially given the way that Adorno et. al. operationalized them. Critics accused them of focusing exclusively on right-wing authoritarianism to the exclusion of left-wing authoritarianism. (Stalin’s Russia was beginning to look scarcely less threatening than Nazi Germany had been.) Moreover, main-line conservatism was implicitly characterized as a soft form of authoritarianism on these measures. Milton Rokeach (1960), among others, tried to develop more general scales that did not run into these problems, and several of the ensuing 16

I have reorganized the factors a bit for purposes of my presentation. In the original, superstition and stereotypy are inexplicably combined. The original scale also contained a sex factor (exaggerated concern with sexual “goings-on”) and destructive cynicism (generalized hostility; vilification of the human), both rooted in Freudian theory.
controversies are still matters of active scholarly research today. But the debate became strangely detached from the question of civic competence, and another paradigmatic approach to citizen competence displaced the theoretical line initiated by Adorno and his colleagues.
6. Citizen Competence as an Informational Problem for Democracy

Adorno thought of citizen competence in terms of adopting a peculiarly democratic political morality: tolerance, intellectual independence, flexibility, openness to persuasion and fair compromise, etc. In his seminal paper “The Nature of Belief Systems in Mass Publics” Philip Converse (1964) stepped back from this morally infused setup and focused on political knowledge as a precondition for any effective citizen action, whatever an individual’s goals may be. Converse demonstrated that most citizens had remarkably low levels of political knowledge, and that the gulf in knowledge between nearly all citizens and political elites was vast.

Converse’s analysis hinged on what he called “belief systems,” a concept which, in turn, hinges on the notion of “constraint.” Belief systems are systematic precisely insofar as the beliefs that they contain exhibit constraint; that is, they fit together in relations of entailment (whether as a matter of logic or convention). These relationships induce correlations between beliefs and opinions at any one time, and underwrite a degree of stability in individual beliefs and opinions over time. Converse argued that the knowledge gap between elites and citizens was so large that it was implausible to think that citizens could organize their political beliefs and opinions on their own terms. The best that they could hope to do was pick among the belief systems on offer from elites. Indeed, large swaths of the public appear to be too ignorant even to manage even this modest task. They exhibit very little constraint in their beliefs, and end up expressing “non-attitudes” – effectively random guessing, which can serve no real democratic goal. On Converse’s setup, then, constraint – understood largely in terms of committing to either a liberal or conservative ideology – became the touchstone for citizen competence.
At first glance, Converse’s standard may seem like an obvious way to provide a more neutral conception of citizen competence by stepping back from the strong moral content built into Adorno’s approach. Whatever goal a citizen might want to pursue, she will at least need a rudimentary map to orient herself in political space to achieve that goal. Faced with a forced choice between political illiterates and partisan ideologues, Converse chooses the ideologues, since they can at least make their preferences felt in democratic choice.

Upon reflection, though, this seemingly neutral choice ends up being much more substantive and objectionable than it might appear, since it defines citizen competence in a way that leaves citizens completely unaccountable to each other. In his celebrated book, *The Nature and Origin of Mass Opinion*, John Zaller extends Converse’s empirical theory and draws out the normative implications of its elite-centered analysis with brutal frankness:

[My theory] makes no allowance for citizens to think, reason, or deliberate about politics: If citizens are well informed, they react mechanically to political ideas on the basis of external cues about their partisan implications, and if they are too poorly informed to be aware of these cues, they tend to uncritically accept whatever ideas they encounter. As normatively unappealing as this implication of the model may be, it is consistent with a large body of theory and research concerning political persuasion (1992:5).

On this account, the normative content of Converse’s conception of competence nearly dissolves into a descriptive category. And to the extent that it retains normative content, it comes close to turning Adorno on his head. For Adorno, mechanically accepting the exhortations of in-group elites and mechanically rejecting those of out-group elites is one of the foundations for authoritarian citizenship.
The rationale for preferring ideologues to political illiterates relies implicitly on an appeal to a market-model of democratic politics. If we accept this model, mechanical sorting on the basis of ideology need not seem so threatening. Indeed, it may drive efficiency. Preference satisfaction is the main goal, and partisan ideology is a decent proxy for individual preferences, so there is no real need for citizens to think, reason, or deliberate about politics independent of party elites. The “liberal” part of liberal democracy sets up institutional guard-rails within which the “democratic” part can run its course without the need to worry about moralistic notions of citizen competence. Strong partisan ideology appears to help average citizens relate their preferences to the choices on offer from elites.

This implicit appeal to a populist interpretation of the market model explains why Converse’s conception of democratic competence seemed so attractive. But it also reveals that his move is at least as substantive and contestable as Adorno’s more openly normative conception. I have already discussed reasons why deliberative democrats want to move beyond the minimalist, market conception of democracy. To those, we might add the reason that seemed so looming and urgent to social scientists in the aftermath of Weimar’s rapid, unforeseen slide from parliamentary republic into authoritarian barbarism. In the end, liberal rights depend on the democratic practices that they structure. On a long view, it seems like whiggish speculation to be confident that institutional guard rails can do all of the work in containing inevitable crises. And even if one feels justified in being whiggish about liberal democratic institutions, the enormity of
being wrong counsels against too easily accepting partisan ideologues as the only alternative to political illiterates for a realistic aspiration of democratic competence.\footnote{Berelson (1952: 317) actually argued for something like a mix of ideologues and illiterates: “A sizable group of less interested citizens is desirable as a ‘cushion’ to absorb the intense action of highly motivated partisans.” Others have deployed revealed preference arguments to suggest that apathy, ignorance, and disengagement from politics are actually signs that most people are basically satisfied with political processes and outcomes.}

7. Attempts to Redeem the Public

Though the Converse-Zaller approach to citizen competence has exerted enormous influence, counter-streams of research rapidly emerged. Early responses attempted to rebut Converse’s empirical conclusions, or at least his interpretation of them. For example, Achen (1975) and Erickson (1979) argued that measurement error could account for much of what seemed so troubling in Converse’s account. More commonly, though, dissenters do not dispute the facts of the case (i.e., that average political information among citizens is low and its variance high), nor really the standards of competence (i.e., being able to translate one’s personal preferences into left-right ideological space).\footnote{V. O. Key Jr. (1966) anticipated this line of research in *The Responsible Electorate: Rationality in Presidential Voting, 1936-1960*. The most important books that sparked the revival of interest in this theme are: Page and Shapiro’s *The Rational Public: Fifty Years of Trends in Americans’ Policy Preferences* (1992); Sniderman, Brody and Tetlock’s *Reasoning and Choice: Explorations in Political Psychology* (1991); Samuel Popkin’s *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns* (1991); and Erickson, MacKuen, and Stimson’s *The Macro Polity* (2002). This empirical line of research was bolstered by rational choice theoretic explanations of why and how voters effectively use information short cuts. This has become a huge literature, but see especially McKelvey and Ordeshook, “Information, Electoral Equilibria, and the Democratic Ideal” (1986), Lupia and McCubbins, *The Democratic Dilemma: Can Citizens Learn What They Need to Know?* (1998), and William Minniti, “A Jamming Theory of Politics” (2011).} Rather, critics made the startling claim that democratic regimes do not necessarily need anything but the most minimally attentive, thoughtful, and knowledgeable citizenry to function well.\footnote{For example, Lupia and McCubbins write: “In this book, we concede that people lack political information. We also concede that this ignorance can allow people of ‘sinister designs’ to deceive and betray the uninformed.” (1998:1). See also Erickson, MacKuen, and Stimson (2002: 428).}
These minimalist claims come in four major varieties. The first group points to the efficiency of decision heuristics, claiming that individual voters can use information shortcuts to arrive at the same decisions that they would make if they were more informed. The second group argues that average citizens process information online – i.e., on the fly – storing its significance for choice affectively, and discarding the factual content itself. Thus, even though they do not “have” much political information, their choices are nonetheless based on much more information than Converse’s analysis would suggest. The third group is somewhat less minimalist in that it purports to show how emotional processes can explain how we efficiently sort cases for which heuristics and online processes suffice, and those that require more concerted attention. Finally, the fourth group appeals to the logic of aggregation, claiming that the electorate can be collectively rational and informed, even if individual voters are not.

In the aftermath of World War II the problem of citizen competence assumed an urgency and substantive inflection that was qualitatively different from similar inquiries going back to Aristotle up through Mill, Tocqueville, and Dewey. In the U.S. and other Anglophone countries, however, the problem of citizen competence got transformed from a moral problem to one of information and efficiency. The general ignorance of mass publics raised anew questions about the fitness of the masses to rule, and underwrote retrenchment under the aegis of democratic minimalism and elitist theories of democracy. Deliberative theories appear especially naïve and aspirational from this perspective, pushing democratic expectations in precisely the wrong direction. But the case against the competence of the demos turned out to rest largely on the same question begging presumptions about the market model and partisan ideology that confused matters on the motivation to deliberate as well (see Neblo et. al. 2010). The public’s input to the
deliberative system cannot be assessed adequately without investigating deliberative quality directly. Yet neither is it the case that deliberative mini-publics can solve the problems of mass democracy on their own. The dialectics of deliberative legitimation mean that there are no closed, simple solutions apart from assessing the quality of deliberation in many forms throughout the deliberative system. Deliberative democracy is not an engineering problem to be solved once and for all, but rather an open-ended search for a better way to live together.

8. *Sophrosyne & Epistemic Democracy*

Skeptics of deliberative conceptions of democracy often point to the rather demanding expectations imposed on citizens in the deliberative account. With some justice, critics argue that it is utopian to believe that sufficient numbers of citizens in mass democracy will dramatically improve their knowledge about politics and sophistication in reasoning about policy. I argue that the key characteristic required of deliberative democratic citizens is not so much political knowledge or sophistication in reasoning. Rather the key trait to develop is something akin to what the ancients called *sophrosyne*, often translated as temperance or moderation (and perhaps better, sound-mindedness). *Sophrosyne* is alone among the ancient cardinal virtues in lacking a clear modern referent, and it has been relatively neglected in updated accounts of politics and the virtues. Yet, I argue that *sophrosyne* is the most needful virtue in the citizenry if a deliberative conception of democracy is to fulfill its purpose of allowing us to steer between technocratic elitism, vulgar populism (Adorno’s problems), and an anomic politics-as-market (Converse’s problem). Unlike producing vast increases in information and political sophistication, fostering *sophrosyne* in the context of democratic deliberation is a plausible goal.
Thus it should be regarded as a kind of low-hanging fruit for hopes of democratic reform.

In “Charmides” Plato runs through four definitions of sophrosyne (quietness, modesty, minding one’s own business, and knowledge of what one knows and does not know). Despite the aporetic ending of the dialogue, these definitions, taken together, sketch a syndrome of deliberative habits that is remarkably well tailored to ameliorating the main vulnerabilities of deliberative politics in an imperfect world. The main barriers to warranting an epistemic interpretation of deliberation are not rooted in individuals being ignorant or naïve, but rather in mobilizing the knowledge and judgment latent in varying sectors of the public appropriately.

Thus, we have a duty as citizens to foster reasoned discourse about public matters in a manner congruent with sophrosyne. Many people, however, treat political choices like impulsive consumer choices and political discourse like a call in show on sports radio: “McCain sucks...Obama rules!” or vice versa. If we are going to use our political power to pass laws affecting our fellow citizens, though, we owe each other reasoned explanations in a way that we do not about what sports teams we support or what toothpaste we buy. Political choices are different from consumer choices and sports loyalties because laws are enforced by people with guns.

That said, discourse governed by sophrosyne is not the same as polite, unemotional discourse. Democracies sometimes need passionate protest, and civil disobedience can actually be a duty in extreme cases. The tricky part is knowing the difference between gross injustices that cry out for redress and deep, but reasonable, disagreements that people in a diverse society cannot avoid. Who is to decide which is which? The first amendment to the U.S. constitution, wisely, says that it cannot be the government itself – the people with guns. So the only ones left
to decide are you and me and our fellow citizens. That is one of the main reasons why *sophrosyne* is so important to good deliberative citizenship.

****

I will eventually permute the following for the ensuing discussion:

1st Dimension: *sophrosyne* as a stand-alone virtue that promotes 1a) one’s own good; 2) others’ good in private life; and 2) tempers the direct pursuit of justice in public life.

2nd Dimension: examples of vices of 1) excess, and 2) deficiency

3rd Dimension: across different domains/objects (bodily impulses, self-assertion of one’s own putative good in a wider sense, self-assertion and expansion of one’s own judgment even when altruistic – i.e., a lack of epistemic humility).

4th Dimension: the four main definitions in *Charmides* (*quietness*: necessary for listening in deliberative exchange, *modesty*: necessary for restraint in epistemic instrumental altruism, *minding one’s own business*: necessary for the proper relation to expertise, and *knowledge of what knows and does not know*: necessary for both the proper relationship to relative expertise and a proper assessment of risk and uncertainty – along with the dream of a world in which no one pretends to be something other than he is: necessary for managing what I have called deliberation’s legitimation crisis).

****

Anachronism problem for my usage: the facets and definitions of *sophrosyne* in *Charmides* represent a kind of athropolotical genotype that must be understood in terms of its phenotypical expression in new mass democratic environments.

Links between dramatic context (headaches, future tyrants, Socrates’s attractions, etc.) on both epistemic humility and self-assertion of goods directly; minding one's own business as a judge of the other's good and basic privacy/individualism, but also unintended consequences for oneself, others, and systemically (as well as other's interpretation of likely consequences). Develop the analogy to and irony of the U.S. prohibition example.

Links back to game theoretic literature, Adorno, Converse, etc.: in order for deliberation to realize the virtues attributed to it, citizens must have the virtue of *sophrosyne* – e.g., in order to avoid market manipulation or magnification of power, unraveling via “instrumental altruism” a la the game theoretic critique, vulgar populism, etc.
Tyrants: raw thirst for influence (in either the market self-assertion sense, or via paternalism); Quietness; Modesty. Minding One's Own Business (Critias comes in; most forcefully rejected; doing good for both oneself and others, but Socrates argues that people often do not really know which of their actions will accomplish this – i.e., against over-self-confident, paternalistic instrumental altruism).

Self-Knowledge (what one knows and not; knowledge of knowledge is too abstract, but it can powerfully guide future learning, and keep us in mind of what we yet do not know: a kind of Deweyan moderation of Burke). Socrates dreams of a world in which no one pretends to be something he is not (against jamming, preference falsification, etc.; if conceived of in too-abstract, intellectual terms, though, sophrosyne cannot lead to concrete good).

Aporia here is partly sincere, but also partly performative: sophrosyne involves admitting what one does not know, and so the elusiveness of a full definition for such a practical, contextual concept is illustrated by the fact that we have made progress, but have not stuck the landing. More generally, the refutation mode illustrates developing knowledge of what one does not know (and knowing negation, positive); provides a model of epistemic humility.

*****

The perfect state is “nowhere to be found” – inducing a kind of humility with respect to those who think that they should rule by fiat and will (Critias and Charmides), but also the problem that causes for those who are more humble, but made part of the ruling (especially in a democracy – though not for Plato).

Need for citizens to actually engage the process as at least analogous to truth-revealing, rather than merely an assertion of will. In deliberative democracy we ask people not merely to assert themselves by any means available; and because equality creates opportunities for either hubris (i.e., inefficient use of social learning) or manipulation of the process that cannot be all institutionally enforced without great cost, we need the individual level restraint (e.g., because the deliberative mechanisms actually provides potential levers of increased influence, control, and manipulation either of an authoritarian kind or technocratic).

Someone who is purely instrumental nearly all of the time is a sociopath. Sophrosyne helps avoid disinhibition (hence sobriety as avoiding the near occasion, etc.); impulsivity, social convention, time-indexing of risks and hyperbolic discounting, etc. for those who are not sociopathic.

*****

The problem of projecting a notion of the common good and public reason requires that individuals behave in certain ways: that we do not unilaterally assert ourselves while others are trying to cooperate; that even in “non-common values” and “non-common veridicality” problems
we are slow to try instrumental altruism – arrogantly substituting our judgment for those of others, which is both a meta-bad instrumentally (in destroying a valuable social practice, akin to epistemic trust), we are apt to be wrong, even when we think that we have good reason. Moreover, there is a theoretical problem with contradicting the assumption that gets the social practice going. Yet we often (though not always) do not have particularly good ways of tracking people’s sincerity, whether their secret ballot matches their public professions in reasoning, whether they are self-asserting selfishly, or via attempts at instrumental altruism, etc. Moreover, despite most of us often being bad at attempting to emulate public spirited reasoning monologically, we have to rely on it to some extent, just out of scale and time problems, etc.

So, *sophrosyne* in democratic deliberation hinges on whether to say anything, when, why, what, etc. Restraint and self-knowledge v-a-v those public processes. Being relatively good at self-interrogation, unbiased reasoning, i.e., doing relatively well monologically. Also, not being quick to self-assert when we do not have a good claim from social/LW/impartial point of view. Also being hesitant to substitute our own judgments for other, especially the group or those who are likely to be more knowledgeable. Being good at spotting those who deserve our deference b/c of knowledge, expertise, trustworthiness, etc., and a disposition to act upon it. Don’t act impulsively, don’t think exceptional (unless you really are), and don’t altruistically manipulate except in exceptional circumstances, etc.

****

Two main components in its democratic/political manifestation (as opposed to the auto-normative notion of managing one’s own bodily desires for the purpose of one’s larger individual good (e.g., exercise to avoid future pathology): Plato blurs the distinction because he thinks that despite its temptations – like ice-cream or physical indolence – tyranny is ultimately not good for the individual soul nor the city – argument that I am not making a great leap in just applying to modern mass democracy.

1) Ability/willingness to restrain expansive self-interest in order to get the normative project (in inferentialist terms) off the ground (related to Rawls’s notion of reasonableness as a willingness to propose fair terms of cooperation: but not so much a one time thing with little variation in people to set up basic terms, but an ongoing capacity that varies between individuals and situations and can be cultivated to stabilize and grow institutions, applications and ordinary legislation and ongoing political life, etc.)

2) Recognition of the burdens of judgment (and perhaps even pluralism/reasonable disagreement, for liberal component of liberal democracy: which does not extend to repressive tolerance: again bivalent, not lazy relativism) and one’s proper role in an epistemic democracy both generally and locally (again: not pure deference and humility as an excuse, but neither ignorant, arrogant, know-nothing self-assertion or paternalism).
3) These two come together in the context of vigilance about improperly motivated reasoning (the desire part distorting the epistemic part, and perhaps sometimes vice versa) in oneself, both locally (trying to form good judgments) and as a global disposition (remaining humble even when we cannot correct in the moment: second-order epistemic restraint). Relates to being open to honest persuasion via social articulation, and restraint when such persuasion fails. (The extremes are Claro! versus lazy or obsequies deference and lack of epistemic vigilance.) “It is difficult to get a man to understand something when his salary depends upon his not understanding it.” (Upton Sinclair, 1935/1994: 109) Link the twin vice forms to the four parts of my sketch of deliberation’s legitimation crisis.

****

Sophrosyne as a bivalent virtue (mean between the extremes relative to you): thus we shift focus somewhat from pure character trait to a situational reinterpretation of authoritarianism (or being anti-democratic in the specific way that lacks sound-mindedness) components: how to support or prompt intraception (Adorno et. al.), for example. We all do it, and so I do not deny that there is considerable individual variation in standing dispositions, but suspect that the ratio of variance explained tends toward the situational (fundamental attribution error), and in any event, more easily altered than people’s personalities.

Time discounting (specifically the way that hyperbolic discounting can generate dynamic preference instability in a way that exponential cannot, though the former is more typically an accurate model (of behavior). Helps explain pre-commitment, intemperance, temptation phenomena, and sometimes correcting our own past mistake about our (benign) future preferences (e.g., variety in snacks, ex.) or preferences for pain meds in childbirth, etc. This approach admits that it is locally and individually rational, but that the same person can be reasonably said to have different preferences that can be respected and acted upon taken in a larger context of reflective judgment. Policies that reflect our non-hyperbolically discounted preferences are also democratic and perhaps represent “our better selves.” Sophrosyne is the virtue that helps defeat perversities of hyperbolic discounting and its attendant preference reversals and time inconsistencies (like “rational” addiction, etc., at least with respect to the kinds of things that generate regret). We can also think of this in reverse: e.g., we should not make rash commitments.

Distinction between “accepting” and “believing” may be characterized partly by their inferentialist implications and relation to other nodes: what can be inferred in/out; “for these purposes…” analogous to a behavioral disposition: religion; nature of proximate warrant (e.g., trustworthy authority versus internal grounds)(and for the purposes follows from this); also related to sophrosyne and epistemic humility (as the mean between two extremes relative to you:
“I’m just a cave man” citizenship or submission to authoritarianism versus arrogant over-confidence, hubris, and self-assertion.

*****

**Sophrosyne:** in political / democratic terms, linked to efficacy (internal, external, related to knowledge, and varied by context and field). Between hedonistic consumerism/self-indulgence/self-assertion versus asceticism, quiescence, quietism, renunciation, effacement, submission (surprisingly linked to a big part of authoritarianism). Not a first order technique or art, like playing an instrument or even something like medicine, but still requires a “knowledge of what” specification. Any first order good can be misused, but we want an art of goodness (gesturing at the Kantian insight of efficiency not being equivalent to goodness): art of temperance, science of sciences, how to gain practical knowledge of good and evil. (Raises the partial circularity of the expertise problem: we need to know medicine to distinguish physicians from quacks.)

Contrasts with hubris – so it contrasts with over-confidence and willingness to substitute one’s own judgment and will (either self-assertion or paternalism), and gestures toward a kind of mild Burkean humility and mindfulness, scaled to a “relative to you” and one’s own business and knowledge when we really do know (i.e., a balance, while recognizing that the more common fault is hubris or self-assertion and self-confidence). Hubris is also related to motivated reasoning and self-confirming bias without the humility to recognize that we have a tendency in that direction: sophrosyne is the opposite tendency without tipping into paralysis or self-dissolution: trusting oneself enough to act/judge with due attention to such knowledge and prudent steps to guard against it, perhaps as simple as acting freely in democratic context, but renouncing unilateral decision power like the guy from the VA study.

Kant: a second-order judgment that finds a person’s appropriate attitude toward herself as a fragile, fallen creature dependent on human society, and yet, a rational agent with the dignity and responsibilities entailed by that station. This middle-way assessment operates both in a general way as to our rational agency, and in particular applications, where our capability, rationality, dependence, and corruptibility will vary greatly.