Authority, Legitimacy and Epistemic Accounts of Democratic Law:

Estlund vs. Habermas

Jonathan Trejo-Mathys

ABSTRACT: David Estlund has recently articulated the most sophisticated form of epistemic proceduralism available in Anglophone legal and political philosophy, one that grants that there are normative political truths and that some possess superior knowledge of them but defeats the claim that superior knowledge can successfully confer legislative authority. In the course of his argument, he charges Habermas' rival version of proceduralism with 'political nihilism'. The first two sections of this paper show that this charge is false and based upon a misunderstanding of Habermas and that it obscures the fact that Habermas provides a potentially stronger form of epistemic proceduralism that defeats the claim that those who know should rule by undermining the very idea that there could be some who possess superior normative political knowledge. The third section discusses a weak point in Habermas' justification of coercive democratic law that is revealed in the course of the defense and suggests a Kantian way of remedying it.

Epistemic Accounts of Democratic Authority and Legitimacy:
Estlund vs. Habermas[1]

Epistemic views of democracy defend its legitimacy, or stronger, its authority in light of the epistemic virtues that the democratic process displays with respect to tracking some normative standard such as justice or the common good, however those are understood. They involve the belief that the truth about such matters, and in particular a sufficiently strong and effective tendency to pursue and discover it, is a necessary part of the justification of political authority and the legitimacy of the norms issued by such authority. This raises the question whether some other way of organizing political power and authority might do better on this score than democracy. If the truth about justice or the common good is the appropriate normative standard to use when justifying political authority or assessing the legitimacy of its exercise, why not connect these things more directly by giving authority to those who know such truths better? Why not let the knowers, or in more classical terms, the wise rule? The rule of the wise, or those who know (better), would be epistocracy. By letting normative truths serve as epistemic standards for the justification of authority, epistemic views create room for an epistocratic challenger that, according to David Estlund, it must defeat before it can claim to have vindicated the (epistemic) claims of democracy.

In this paper, I want to explore the differences of the theories of democracy that Estlund's epistemic proceduralism and Habermas's discourse theory offer us, in particular regarding the question of the authority and legitimacy of democracy. Since they both offer epistemic accounts, both face the challenge of showing why democracy rather than epistocracy is justified. Such a comparison proves worthwhile for three reasons. First, due to a misunderstanding of Habermas's view, Estlund dismisses it as a form of political nihilism,

and hence unfortunately does not undertake the kind of examination I do here, despite acknowledging other respects in which the views are quite close.[2] Second, the discourse theory uses a different, more radical strategy to undermine epistocracy than epistemic proceduralism and this alternative path may offer other advantages. However it also turns out that if this path is taken, the issues of authority and legitimacy cannot be separated as they are in epistemic proceduralism and many other current approaches. Indeed, following Kant, the order of explanation between these two things turns out to be just the reverse of the usual treatment, and the implications of this are worth exploring.

I. *The Charge of Political Nihilism*

So what exactly is political nihilism and what is Estlund's evidence for accusing Habermas(ians) of it? Political nihilism is defined as the view that "there are no appropriate standards (not even minimally true ones) by which to judge political decisions".[3] It would be, to put it mildly, very surprising if this were true of Habermas (most of all to him). What is Estlund's evidence? For such a strong charge, he does not in fact present very much. I reproduce two key passages, including some of the few (and as we will see, misleadingly selective) quotations of Habermas that Estlund uses:

> Deliberative democratic theory claims to employ only purely procedural standards for the public employment of reason. "The notion of a higher law," Habermas urges, "belongs to the premodern world." There are no standards that loom over the political process, policing its decisions, not even any standard of reason itself. "We need not confront reason as an alien authority residing somewhere beyond political communication." The only normative standards that apply to political decisions are noninstrumental evaluations of the procedures that produced them—in particular, standards of "procedural rationality" based on the power of reason in public political

---

[2]Indeed, according to Estlund, "the moral challenge for the epistemic conception of political authority is to let truth be the guide without privileging the opinions of any putative experts", *Democratic Authority: A Philosophical Framework* (Princeton: Princeton University Press, 2008), p. 102. Hereafter DA. The part after "is" could serve as a motto for most of Habermas's work.

[3] DA, 25. Cf. DA, 29, where Estlund here even says that Habermas is motivated by a 'no-truth' (in politics) argument.

discourse. Any imposition (in theory or practice) of substantive political standards would preempt the ultimately dialogical basis upon which Habermas thinks political normativity must rest. There is an echo of Arendt here: politics is the site of discursive contestation, so politics cannot begin with conclusions.

Here is where the price of the nihilist view is evidently too high to pay, or so I will argue. No appeal to good outcomes is permitted on this view, there supposedly being no such thing.[4] … The task for deliberative democratic theory, then, has become the odd one of explaining the central importance of substantive public discussion of the procedure-independent merits of possible political decisions, without ever granting that there actually are any procedure-independent standards.[5]

The two quotations of Habermas come from two very different contexts in *Between Facts and Norms*, the first being an introduction of Habermas' own conception of law by means of a discussion of its relationship to morality and the second being a normative account of judicial review. With regard to the first, Estlund misleadingly leaves out a parenthetical remark that might have clued readers into the fact that the metaphors of height in Habermas' text are not functioning in the way he suggests. The full sentence runs "the notion of a higher law (i.e., a hierarchy of legal orders) belongs to the premodern world".[6] Estlund is thus misleadingly general when he says, in accordance with the height metaphor, that *no* standards "loom over" the political process, "policing it", "not even *any* standard of reason itself".[7] If he had instead said that Habermas holds that there are no genuine *legal orders* that transcend positive systems of law, "loom over" politics, etc., he would have been right, but it wouldn't have sounded like something worth accusing someone of, since to many it would appear to be common sense. However, formerly it *was* a very widespread view that a kind of legal order did so 'loom'. For the Stoics, the norms of this system were inscribed within the ordered whole of the cosmos whose rational ontological structure was graspable by right reason. For Aquinas, this view was then ramified into distinctions between eternal law, natural law,

---

[4] DA, 27.

[5] DA, 30.

[6] *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy* (Cambridge, MA: MIT Press, 1996), p. 106. Hereafter BFN.

[7] My italics.

human law and divine law.[8] Even for Locke, the natural law (or at least its legal force) was rooted in the will of a divine lawgiver, while for Kant it was still (arguably) rooted in a 'higher' moral law, and so on.[9]

In all these cases, a normative order that was not merely metaphorically conceptualized as a *legal* one was both hierarchically superior and such that systems of positive law should in a certain sense imitate, embody or reflect it.[10] *This* idea, and the intuition behind it, is being rejected by Habermas, not the idea that political decisions can be criticized using non-political and non-procedural standards. In fact, this ought to have been clear from what is said just three sentences prior to the passage Estlund (partly) quotes. Habermas notes there that nevertheless "this intuition is not entirely false, for a legal order can be legitimate only if it does not contradict basic moral principles".[11] Since this claim clearly sets out a necessary condition for the *normative* legitimacy of a legal order, and laws are the outcomes of (certain) democratic political procedures, it is clear that Habermas acknowledges standards that bear critically on procedural outcomes.[12] Indeed, he elsewhere explicitly states that morality is "a standard for legitimate law".[13] So Habermas is definitely not concerned to "free politics from the supposed despotism [!] of normative standards for political decisions".[14]

In the second of Estlund's quotations, Habermas is rejecting a different though partly related alternative posed by Frank Michelman's writings on republicanism and constitutionalism, namely, the false choice between communitarian-republican forms of

---

[8] *Summa Theologiae*, I-II, q. 91, aa. 1-4.

[9] As will be seen below, I think such a view of Kant's legal-political philosophy is at least questionable.

[10] This is what Habermas calls a "Platonic intuition" that remains even in Kant's noumenal/phenomenal distinction (BFN, ibid.).

[11] BFN, 106. If Estlund had included the parenthetical remark, readers might have hit upon this sentence while in search of the needed clarification of the reference to a "hierarchy of legal orders".

[12] I leave aside the cases of customary or common law norms that do not stem from legislative procedures. However the fundamental constitutional norms of modern democracies, whose content generally trumps that of customary or common law, do stem from positive ratification.

[13] BFN, 206.

[14] DA, 34.

legitimation that require a thick prior consensus on the values of a shared way of life or collective identity and a "transcendent" authority residing, or coming from, somewhere outside the social and political life of the community (hence 'alien').[15]   Again, Estlund leaves out a crucial first clause as well as the important context of the previous sentence. Here is the already quoted statement with this context restored.

> A consistent proceduralist understanding of the constitution relies on the intrinsically rational character of a democratic process that grounds the presumption of rational outcomes. Under this description, reason is embodied solely in the formal-pragmatic conditions that facilitate deliberative politics, so that we need not confront reason as an alien authority residing somewhere beyond political communication.[16]

Note two things about this passage. First, the intrinsically rational character of a (deliberative) democratic process of lawmaking only grounds a *presumption* of a normatively legitimate outcome (this is what saying the outcomes are 'rational' amounts to here). It does not allow us to infer that the outcome *is* normatively legitimate (or just or correct) simply because it is the outcome of the process. A properly constituted procedure is thus (at most) a necessary, not a sufficient, condition of a correct (or at any rate justified) outcome.[17] It does not collapse 'good outcome' into 'outcome of the correctly run procedure'. This is strong evidence that Habermas understands democratic procedures in an epistemic fashion just as Estlund does. Second, reason is embodied in "formal-pragmatic conditions" that make a deliberative politics possible. What that phrase means will hopefully be cleared up a bit below, but for the moment it is important to understand that for Habermas the formal-pragmatic presuppositions (as he more usually says) of communication and argumentation are what give discursive processes like that of 'deliberative politics' the power to even generate a presumption that their outcomes are correct (or rational or just or legitimate). It is only insofar as the democratic process *as a whole* adequately or sufficiently satisfies these

---

[15] Precisely the false choice in Michelman's writings that Estlund notes and dissolves using a minimalist conception of truth in "Who's Afraid of Deliberative Democracy? On the Strategic/Deliberative Dichotomy in Recent Constitutional Jurisprudence", *Texas Law Review*, v. 71 (1992-3): 1437-1477: cf. here pp. 1463-73.

[16] BFN, 285.

[17] Hence it is an 'imperfect' procedure, in Rawls' influential terms.

conditions that it has an 'intrinsically rational character'.[18] It is perhaps more important, as we will see, to keep in mind that these presuppositions operate *wherever* attempts are made to support, justify or criticize a claim by means of reasons. Hence there is nothing special about "political normativity" in this respect, as Estlund might be taken to suggest above. In Habermas' view, *all* normativity rests on an "ultimately dialogical basis", and science, morality and art criticism are just as much "sites of discursive contestation" as politics is.

It is possible that such a serious misunderstanding could have arisen otherwise than by lack of sufficient attention to an admittedly demanding text. In particular, I think Habermas' own very liberal use of "procedure" might encourage the kind of reading Estlund gives. How this term is understood greatly affects whether or not it is right to say that Habermas does not allow any procedure-independent standards to play a role in evaluating procedural outcomes and whether this is an objectionable aspect of his political theory. In addition, there are at least two different ways of using "standard" whose differences have similar implications.

First, let's take a look at Habermas' uses of "procedure".[19] Then we'll see whether he fails to allow for any "procedure-independent" normative standards for evaluation of political outcomes. For our purposes it will help to distinguish three uses (though there are others). Habermas sometimes uses a very general sense of "procedure" in which the major streams of post-Hegelian, post-grand-metaphysical-systems philosophy are said to converge on a theory of *rationality* that is 'procedural' in giving particular importance to the theory of argumentation.[20] He has various arguments for why the rationality that inhabits everyday or routine forms of action and communication must ultimately refer the practice of arguing or 'discourse'. These justify its central theoretical role because they demonstrate the simultaneously cognitive or rational and social or conflict-resolution functions that it serves.

---

[18] This claim can be compared to the use Estlund makes of the notion of ideal epistemic deliberation in justifying the legitimacy (and authority) of democratic procedures, cf. DA, ch. 9.

[19] I will not in fact list all of the types of discursive procedures that Habermas distinguishes, but only three major types that are crucial for the purposes of the argument here.

[20] *The Theory of Communicative Action*, vol. 1 (Boston: Beacon Press, 1984), p. 2.

In a more specific way, *moral* argumentation can be called a "procedure" insofar as it involves the operationalization of a particular rule of argumentation for the rational resolution of moral conflicts, namely, (U) or the principle of universalization.[21] However, the use of this rule in moral discourses should not be understood to function as a decision procedure in the strict sense. Either a rationally motivated agreement emerges that a candidate norm satisfies (U) or it doesn't. If not, the issue remains open or unsettled. Lastly in a more familiar way, a *political* process of legislation involving, say, a structured debate followed by a majority vote is also clearly a procedure. This time there *is* a clear mechanism for *deciding* an issue, say, whether abortion will be legal (a slim majority votes yes), even if it has not been *settled* by the achievement of a rationally motivated agreement (fierce public debate still rages).[22]

Given these distinctions, we can see that, in addition to the exegetical reasons given above, there are systematic reasons why Habermas can accommodate the use of standards in normatively evaluating the outcomes of *political* procedures. Any currently accepted standards (of whatever kind) and any standards justified by non-political procedures can be used to criticize political outcomes.

At this point it may help to clear up a possible ambiguity in "standard". We can distinguish between the two following different kinds of standards: first, there are standards that anything *which can count as* a rational discursive process of justification must meet and second, there are standards deployed *within* a process of justification (i.e., of some particular claim). The pragmatic presuppositions of communicative action (and hence discourse, argumentation, justification, etc.) are standards – though the term is a bit awkward – of the first kind. They represent the 'must' of a "weak transcendental necessity" and they constitute

---

[21] Here is one formulation of it: "A norm is valid [i.e., morally right] when the foreseeable consequences and side-effects of its general observance for the interests and value-orientations of *each individual* could be *jointly* accepted by *all* concerned without coercion", *The Inclusion of the Other: Studies in Political Theory* (Cambridge, MA: MIT Press, 2000), p. 42.

[22] The settled/decided distinction comes from Simone Chambers, *Reasonable Democracy: Jürgen Habermas and the Politics of Discourse* (Ithaca: Cornell University Press, 1996). For dramatic contrast, compare the issue of slavery, now both decided *and* settled. There is no fierce debate raging about this as a political matter.

the "counterfactual basis of an actual practice of reaching understanding".[23] Yet they do not have the sense of the prescriptive 'must' of normative rules of action (whether deontological, teleological/axiological, or empirical, i.e., regarding technical or instrumental efficacy).[24] If they prove not to be satisfied – say, it turns out that you are actually sleepwalking and sleeptalking, and hence not an accountable interlocutor, or we find out the others were merely pretending to enjoy a dinner at the restaurant with us (supposedly to plan a couples retreat) while their accomplices robbed our house – then we were simply not engaging in communicative action (or discourse), but rather doing something else.[25]

The demonstrable violation of these pragmatic conditions will significantly (and often entirely) impair the epistemic contribution any discursive process (rational, moral, or political) can make to generating a presumption that its outcome is correct in the relevant sense (true, just, right, legitimate). In the case of legally constituted procedures such as democratic deliberation and legislation, violation of the conditions that enable the procedure to claim its 'rational' character, for instance, through the exclusion of most adult blacks from the enjoyment of their political rights, will immediately cut against any claim to legitimacy on the part of a law the procedure generates (e.g., 'Jim Crow' laws of racial segregation). So there is certainly room for the kind of "retrospective" criticism of outcomes Estlund mentions.[26] However, even here I think Estlund misconstrues the role of the basic rights in

---

[23] BFN, 4. The notion of a counterfactual *basis* may be misleading insofar as it suggests something like an *unreal* basis for a *real* practice. The counterfactual presuppositions are *actually* made but their *content* is typically, and taken strictly perhaps always, not actually instantiated. We are not in fact fully free of internal and external forms of coercion, deception or distorted understanding and communication, nor do all participants have precisely the same opportunities to participate or access to information, nor are all affected actually included, and so on. But without the *presumption* that these conditions are sufficiently or approximately met, we cannot even take ourselves to be involved in genuinely rational discourse.

[24] Ibid.

[25] There was no real communicative exchange of reasons for and against various proposals regarding the couples retreat and no real exchange of binding commitments to accept the victorious one. There was in fact an extended (strategic and deceptive) diversionary tactic. Nor was there a shared moment appreciating my recital of Wordsworth's great "Ode on Intimations of Immortality" in the moonlight, one you validated by saying, in a dreamy sort of way, "Yes", when I asked if you liked my performance – you really were dreaming.

[26] DA, 89.

Habermas's theory. He seems to view them as external standards for an *ex post* evaluation of outcomes such that "destruction of the relevant liberties would be illegitimate even if it had been decided by the proper procedure" (88). If we understand the relation between 'basic liberties' and the legitimating power of the democratic procedure in terms of the abstract *categories* of rights involved in Habermas's argument for the co-originality of basic rights and popular sovereignty, then Estlund's description of the situation here is simply incoherent. For in this case, if the basic rights that secure the communicative enabling conditions of politically legitimate legislation and state coercion are 'destroyed', then the relevant procedure simply did not exist. This 'outcome' is not illegitimate *ex post*.[27] It is *ex ante* illegitimate. In view of the destruction of the procedure-enabling conditions, there simply *was nothing that could count as a legitimating procedure*, and hence no outcomes relevant to the assessment of the procedure. On the other hand, if we understand 'basic liberties' as the concrete, legally specified ones we find in actual systems of national law, then Estlund's description is both coherent and correct regarding Habermas' theory, but there will typically be many other normative elements of the legal system and the political tradition that can be used to criticize the violation or abrogation of specific liberties by the legislative process or by administrative decrees. Estlund's misunderstanding of Habermas here is thus likely rooted in a failure to take sufficient heed of this distinction between the 'abstract' categories of rights used in his explication of the 'normative core' of constitutional democracy and the concrete or 'saturated' specific rights and liberties of existing polities.

Now let us consider standards understood as considerations deployed *within* a procedure of deliberation, justification, or decision-making. Is it true on this understanding of "standards" that in Habermas' theory there are no appropriate standards to use in judging political decisions? I think not. In fact, I think both moral and political standards will be

---

[27] The main place for a retrospective ex post criticism of procedurally proper outcomes in Habermas' theory is when the anticipated consequences of a decision turn out to be (sufficiently) different than expected.

available here. First, *any* currently accepted moral standards can be used to criticize the outcome of a political procedure. Second, prior political decisions or outcomes – i.e., existing legal or political norms – can be used to criticize a current (or even a looming) one. For instance, various states in the U.S.A. are currently suing ('criticizing') the federal government over the recent health care reform bill by using (rightly or wrongly) a pre-existing constitutional norm. Thus it should be clear that at the level of political justification there is no question of Habermas's being a political nihilist.

II. *Epistemic Proceduralism and the Challenge of Epistocracy*

Since my discussion proceeds comparatively by way of Estlund's theory, let us first make clear how he defines the key terms 'authority' and 'legitimacy'. According to Estlund, an agent X has *authority* when it has the moral power to issue a command or prohibition to another agent (or set of agents) and the issuing of the command or prohibition creates a moral requirement (of some weight or other) for that agent (or set of agents)(DA 42, 118). The statement that 'X has authority' says nothing yet about the permissibility of exercising authority, of issuing the authoritative command. This is a separate issue. It just entails that X's commands can count as reasons for action on their own if one is under X's authority, or in other words, that one is under a duty to obey *if* commands are issued (DA 119, 127). The agent has *legitimacy*, on the other hand, if it has moral permission to coercively enforce (certain) commands (DA 41). This applies to both acts and threats of coercive enforcement. In addition, locutions like 'a legitimate law' are held to be derivative uses of 'legitimate' entailing that the state (lawgiver) is permitted to coercively enforce the law owing to the law's procedural source (ibid.).[28]   Now for a precise definition of the philosophical thesis of epistocracy.

---

[28] According to Estlund there can definitely be authority without legitimacy, yet he is uncertain about, though willing to entertain, the possibility that there can be legitimacy without authority (DA 134).

Estlund characterizes epistocracy in terms of three tenets:

Truth tenet – there are true procedure-independent normative standards by which political decisions ought to be judged.

Knowledge tenet – some (relatively few) people know these normative standards better than others.

Authority tenet – the normative political knowledge of those who know better is a warrant for their having political authority over others.[29]

In order to defeat epistocracy, an epistemic defense of democracy has to establish the falsity of one of these tenets. Estlund believes that denying the truth tenet is too high a price to pay, since it results in what he calls political nihilism—something which, as we have seen, he mistakenly claims results from Habermas' theory. He also believes that the knowledge tenet (given the truth tenet) is very difficult to deny. Hence he concentrates on refuting the authority tenet and demonstrating that it is fallacious -- the 'expert/boss fallacy' (cf. DA 40).[30] Though we have already established above that Habermas is not a political nihilist, this also follows directly from the fact that Habermas accepts the truth tenet since there are all manner of normative standards that are independent of specifically political procedures that ought to be used to judge political decisions. Nevertheless, another reason why Estlund might have found it easy to believe that Habermas denies the truth tenet may lie in the fact that Habermasian political theory, as I understand it, would likely choose a different strategy to deal with epistocracy, namely, it would attack the knowledge tenet. Since Estlund thinks it is very hard to grant the truth tenet without granting the knowledge tenet, it is natural for him to think that any view that denies the knowledge tenet must deny the truth tenet too, indeed, as natural as modus tollens. The tenor of the discussion of Habermas and Arendt in the early parts of his book suggests this is why he views them as 'nihilists'. Thus in the end the *real* challenge that Estlund presents to a Habermasian theory of democratic authority is to explain

---

[29] DA, 30.

[30] The expert/boss fallacy involves a mistaken inference from "S would rule better (knows better)" to "S is a legitimate or authoritative ruler". Estlund's argument then becomes a way to flesh out the thought wittily expressed in a phrase he attributes to Nomy Arpaly: "You may be right, but who made you boss?"

how one can provide reasons to reject the knowledge tenet *without* undermining the truth tenet and thereby falling into a dreaded political nihilism. Since the purpose of this paper is also comparative and exploratory, I will develop the outlines of a Habermasian view by means of an exposition of Estlund's favored strategy that reveals several possible points at which the Habermasian strategy might have certain advantages.

Here, in brief, is Estlund's argument to establish the normative or moral superiority of democracy to epistocracy. He defines democracy as "citizens collectively authorizing laws by voting for them, and/or for officeholders who make them" (DA 65). While in this quotation and many other places he only mentions voting, he does make clear at certain points that his argument for the epistemic virtues of democracy depends upon there being a prior period of deliberation among citizens who are each focused on the normative truth of the matter at issue before a decision is arrived at by voting.[31] Next, he introduces the core idea of epistemic proceduralism, namely, "procedural impartiality among individuals' opinions, but with a tendency to be correct", or in other words, "the impartial application of intelligence to the moral question at hand" (DA 107). He also relates this thought to deliberative democracy, which he understands as involving the view that "political authority depends on a healthy application of practical intelligence in reasonably egalitarian public deliberation" (87). The basic thesis is then that "from a normative (or moral) point of view, democracy is the best epistemic strategy of governance among those that are generally acceptable (though it is not necessarily better than every alternative) (DA 42).[32] As the parenthetical remark makes clear, the argument does not establish that democracy is epistemically the best political arrangement, but only that it is the best among those that are generally acceptable.

---

[31] DA, chs. 7-8. That citizens aim at the normative truth of the matter at issue, as opposed to successfully satisfying their own preferences or pursuing their interests, is crucial for the argument. For powerful arguments against the coherence of interpretations of voting as the expression of desires or preferences, see Estlund's "Democracy without Preferences", *Journal of Philosophy*, v. 99, n. 3 (July 1990): 397-423.

[32] As he puts it elsewhere in the context of the argument for the legitimacy of democracy, the theory "requires that the procedure can be held, in terms acceptable to all qualified points of view, to be epistemically the best (or close to it) among those that are better than random" (DA, 98).

So two things must be established. First, that democracy has epistemic virtues that recommend it as a way of organizing political authority that is superior to ways of making decisions about political matters that may have valuable moral, but no epistemic, virtues; for instance, by flipping a coin or some other randomizing method, which would be fair but obviously have no tendency whatsoever to track any normative political truth. And second, that out of the class of epistemically valuable ways of organizing political authority, democracy is the one to be chosen, and not some other, in particular not epistocracy.

In order to accomplish these tasks we require two things: a way of assessing the epistemic performance of democracy and a way of showing that democracy is, from a normative political point of view, more choiceworthy than epistocracy.[33] The first requirement is met by a list of 'primary bads', things which presumably all agree are political evils of great weight. Estlund's list of primary bads comprises famine, hunger, genocide, political collapse, economic collapse and war.[34] To establish the epistemic choiceworthiness of a given political decision-making procedure, or mode of exercising political authority, one must at least show that it does better than choosing at random among available alternative decisions. The primary bads thus serve as the epistemic standard in light of which one must show that deliberative democratic decision-making will do far better than a randomizing procedure.[35]

---

[33] Ideally one might want to have arguments showing that democracy is the best political procedure, period. This would require arguments showing exhaustively that of any pair of possible political procedures that includes democracy and some other procedure, democracy is superior. It is naturally an extreme understatement to say that that would require a very, very long book. Instead, given the epistemic nature of the justificatory strategy, one simply tries to show that it can defeat what looks like the most salient competitor, epistocracy.

[34] DA, ch. 9.

[35] They are particularly appropriate to serve this function because one can safely assume widespread agreement, perhaps even universal consensus, on their badness and because they are political evils of very great weight. It would not be impressive if one could show that a political decision-making procedure generated the normatively correct answer with great frequency on relatively trivial or at any rate less important matters. Such a procedure might do worse than random on things we very much want to avoid, e.g., the primary bads. We would not take the impressive epistemic performance on trivial matters to justify the authority or legitimacy of a procedure that did worse than random on the most weightiest forms of political evil or harm (cf. DA 163-66). Yet if we can show that democracy does far better than random on very weighty issues, then we may hope this result will carry

The second requirement is quite a bit more complicated. This is at least in part because Estlund grants the knowledge tenet, according to which there are a (relatively) few persons who have superior knowledge about the political normative standards (justice, common good, etc.). Since we are aiming to establish the authority of democracy from an epistemic point of view, this fact strongly suggests the challenge, "Why not let those who are admittedly in an epistemically superior position rule, if getting the political truth right is the point after all?" Here Estlund makes use of a familiar liberal thought to rule out epistocracy. In order for a form of political authority to be legitimate, it must be acceptable to the reason of those who are subject to it. It must be possible for them to understand and endorse the reasons why it can legitimately (permissibly) enforce its commands or law with coercive means. The particular form this constraint takes in Estlund's argument is that of a *qualified general acceptability requirement* according to which no political procedure may legitimately coerce a population unless it can be accepted by all 'qualified' views within that population. This requirement states a necessary condition for legitimacy that entails that the rejection of a candidate procedure by *any* qualified view suffices to eliminate the procedure from contention. The criteria for 'qualification' are left vague by Estlund, who notes that particular claims about the qualification or disqualification of this or that view will often be very controversial. However, one presumes that there will be no doubt that many views, for instance, one according to which no form of government is legitimate that does not execute randomly selected brown-eyed persons on the fifth of every month, are disqualified for the purpose of political justification. The idea is similar to the widespread use of a criterion of 'the reasonable' by theorists influenced by Rawls. The exact location of the boundaries of the class of qualified views is unimportant if there is agreement that *any* view that suffices to put epistocracy out of contention falls within them. This indeed proves to be case.

over for a wide range of less important ones (DA 160).

With the qualified acceptability requirement in hand, Estlund holds that epistocracy is defeated by the reasonable rejectability of the most powerful and plausible attempt to defend the authority tenet of epistocracy, namely Mill's proposal that the educated should have a greater share in political power (in the form of more votes per person). The decisive defeater for this proposal is the "demographic objection", according to which, "even though we must all grant that a better education (somehow conceived) improves the ability to rule wisely, it is not unreasonable or disqualified to suspect that there will be other biasing features of the educated group, features that we have not yet identified and may not be able to test empirically, but which do more epistemic harm than education does good" (DA 222). What is important, however, is that the beliefs that ground these doubts need not be true. It need not actually be the case that a given group of epistocrats will in fact be hampered in the way those who doubt believe. It is enough that this worry be a reasonable or 'qualified' one. The only thing in this argument which has to be in fact true, or at least asserted as such, is the qualified acceptability requirement itself.[36] Thus the qualification of the demographic objection effectively establishes that epistocracy cannot be a legitimate mode of political rule, since it knocks out what is arguably the most impressive and plausible form of epistocracy. Despite the admission that there are some who know the political truth better than others (the knowledge tenet), it does not follow that they ought to rule (the authority tenet).

Yet a further step is needed to justify the *authority* of democracy (moral power to require or forbid actions), since qualified acceptability only confers legitimacy (permission to coercively enforce commands). Estlund accomplishes this by means of a democracy/jury analogy argument that aims to establish a 'normative consent' theory of democracy's

---

[36] See Estlund's, "The Insularity of the Reasonable: Why Political Liberalism Must Admit the Truth", *Ethics*, v. 108, n. 2 (Jan 1998): 252-275, for sophisticated and, I think, sound arguments that political liberalism must at least assert the ("minimal") truth of the liberal principle of legitimacy itself in order to avoid incoherence. The same arguments establish that the general acceptability requirement Estlund uses must itself be asserted as true for the purposes of political justification (whether it must actually be true is perhaps a separate matter).

authority.[37] He first asks us to imagine a society, Prejuria, in which there is no established judicial system but everyone agrees on rules forbidding certain behaviors and on the appropriate punishments for violations of the respective rules. He then describes the undesirable moral consequences that result from the lack of an acknowledged judicial authority who determines who is guilty of what kind of infraction and hence is officially liable to which kind of punishment. Among these consequences are unreliable and biased judgments about guilt and repeated punishments of an offender for a single violation by different individuals who are unaware of each other's uncoordinated, spontaneous acts of rule enforcement. If we agree on the serious moral disvalue of these and other consequences, then Estlund thinks we will agree that if a workable public jury system for adjudicating purported rule violations arises, even if at first only through the unilateral initiative of a subset of citizens, then all members of Prejuria would have a moral obligation to consent to abide by its decisions, given the significant epistemic advantages of jury systems over decentralized, self-help arrangements. The power of the idea of normative consent comes from the fact that, if sound, it establishes that one is obligated to obey any authority to which one would be morally wrong to *refuse* to consent, if one were given the chance.

The general moral background here is filled in by the notion of basic humanitarian duties that all persons have to make fair contributions to the prevention or remedy of moral disasters.[38] Prejuria's lack of a publicly authoritative judicial system is supposed to lead to just such a disaster. The argument is then generalized to the political case as a whole, and the epistemic merits of (deliberative) democracy are analogized to those of the jury system. Political-juridical anarchy is such an evil that one has a humanitarian duty to make a fair contribution to its solution. Epistocracy has already been ruled out of contention by the

---

[37] As defined by Estlund, this is equivalent to establishing that the subjects of the democratic authority have obligations of compliance or obligations to obey the law, i.e., what are often referred to as political obligations (DA 151-156). In addition, the argument for the epistemic better-than-randomness of deliberatively democratic procedures employing the primary bads also serves as a background for the 'normative consent' argument.
[38] For the details of the argument, see DA, chs. 7-9.

legitimacy argument. And the epistemic sufficiency of democracy for solving the problem is established by reflecting on how it would do regarding the primary bads. So one is obligated to obey since it would have been morally wrong not to promise to obey had one been given the chance. One would have failed to make one's fair contribution to solving a pressing humanitarian problem, for democracy is an available solution that is (unlike epistocracy) both generally acceptable and epistemically sufficiently good. This ends the presentation of Estlund's epistemic proceduralist theory of the authority and legitimacy of democracy.

Before attempting to sketch out a Habermasian strategy for defeating epistocracy and justifying the legitimacy and authority of democracy, I would like to note some features of Estlund's theory that stand out from a Habermasian perspective. The acceptance of it by addressees depends at two points upon actual agreement: first, regarding the question of qualification, and second, regarding the question of standards for the epistemic evaluation of democratic procedures. Now insofar as there actually is agreement upon these questions, there is no reason why Estlund's arguments are not compatible with Habermas's discourse theory. In fact, if there is agreement among Estlund's addressees that the circumstances of Prejuria are such that it would be morally wrong to refuse to consent to the authority of a jury system and/or democratic institutions, then in this context a Habermasian can even go along with the theory of normative consent. In this respect, Estlund's theory and the discourse theory can be made logically consistent and complementary, instead of competing, accounts.[39] However, one possible advantage of the discourse theory is worth mentioning here. The question of qualification, or in Rawlsian terms, of the reasonable, is and has been a deeply political one, both in the historical development of constitutional democracies and in recent debates in democratic theory. The abolitionist, feminist and civil rights movements offer

---

[39] Asserting them at the same time would require modulating the way Estlund argues regarding the knowledge tenet. Instead of simply accepting it, as Estlund does, one would have to embed it in a conditional and show that epistocracy would not follow even *if* the knowledge tenet were true. Otherwise the strategies are not logically compatible since the Habermasian argument *denies* the knowledge tenet.

abundant examples of claims which only achieved recognition as 'qualified' or 'reasonable' views through active social struggle and contestation. For instance, one writer of the then leading literary journal, *The Edinburgh Review*, responded to Wollstonecraft's arguments in favor of female suffrage by saying that soon people would be demanding that their dogs get the right to vote. (The writer was a man, of course.) Naturally this claim is not best read as a counterargument. Rather it expresses a judgment that Wollstonecraft's view is disqualified or unreasonable and not worthy of serious rebuttal. When such views are deeply entrenched among socially dominant groups it can take a long time and much social struggle to unseat them. This fact is explicitly built into the discourse-theoretical view of the democratic process and indeed of the evolution of the constitutional state. Hence the discourse theory treats the qualification question in a reflexive way by building it into its account of politics.

III. *The Discourse Theory and the Challenge of Epistocracy*

As I have already suggested, the Habermasian account levels its sights at the knowledge tenet, according to which some (relatively few) know the normative standards of politics better than others. It thereby directly undermines epistocracy by denying the existence of 'experts' in political morality. Yet as we have seen Estlund thinks it is very hard to deny the knowledge tenet, so it is worth trying to shed some doubt on its plausibility before going on.[40] He even claims at one point that "it is certain there are subsets of citizens that are wiser than the group as a whole" (DA 41), where their 'wisdom' pertains to specifically normative claims about what a polity ought to do, i.e., what kinds of laws and policies are just or right. This may sound obvious in the face of disagreement about laws and policies, since if any group's views are the right ones, then all those with whom they disagree are wrong, and

---

[40] It is important to keep in mind that the knowledge tenet concerns the superior knowledge of the few, and not the existence of knowledge per se.

those who are right are presumably wiser than those who are wrong. But let me say something to weaken the plausibility of this claim.

Note first that we are supposed to believe that, without regard to any specific issue or subject matter, there are subsets of citizens who are wiser than the group *as a whole*. This sounds true in the abstract. That is, if we consider, say, nuclear physics, then obviously the subset of citizens comprising all citizens who are nuclear physicists is "wiser" on this issue than the citizenry as a whole. But here a question arises. Is the cognitive division of labor of the group constitutive of the wisdom of the group-as-a-whole or is it rather that, for the purposes of assessing the collective wisdom of the group, we simply as it were sum up the knowledge of each separate adult in the group? If the members of the group acknowledge the legitimacy of a cognitive division of labor, then it might well be the case that the "wisdom" of the nuclear physicists *is* the wisdom of the group as a whole with respect to the subject of nuclear physics. In this case, if you ask each separate individual member of the group to give you all of his or her wisdom regarding nuclear physics, each individual who is not a nuclear physicist will say something like, "Not my area: ask the nuclear physicists", and each individual who is a nuclear physicist will offer up whatever wisdom they have about nuclear physics. If the group is *that* kind of a group – and aren't existing societies more or less precisely this kind of group? – then it would be at least misleading, and perhaps outright false, to say that there are subsets of citizens that are wiser than the group as a whole, since the wisdom of the group as a whole (at least on the issues for which they have developed a division of cognitive labor) will simply be the wisdom of the subset that specializes on that issue. But the implausibility of the knowledge tenet on this construal of the wisdom of groups should be obvious. Who exactly is supposed to be the proper subset of the citizenry that comprises the "normative" or "moral" or "political morality" experts? Here Estlund's arguments concerning the failure of the epistocracy of the educated to be generally acceptable seem roughly as applicable as they were in the context of the argument against the authority

tenet. Any form of education that supposedly led to this subset's superior knowledge of what we ought politically to do – and how else would they have come by this knowledge? – would be subject to reasonable suspicions of having fatally biasing features.[41] This line of thought becomes even stronger if, as Habermas does, one views morality as designed for the impartial resolution of conflicts among social actors and holds that there is an irreducibly intersubjective element in moral justification.   This leads to the conclusion that on fundamental moral questions there is no similar cognitive division of labor, or at any rate there is only the limiting case of a division of labor in which in principle everyone involved has an equal role to play. This will become clearer through a discussion of the Habermasian treatment of authority.

Since our target is a Habermasian account of democratic authority and legitimacy, let's recall that authority is, on the Razian understanding, the normative power to require or forbid action, or put otherwise, to be a source of binding commands, while legitimacy involves the permissibility of coercively enforcing such requirements, prohibitions or commands.[42] With this in mind, let's consider what the grounding of authority is in the discourse theory. At the beginning of *Between Facts and Norms*, Habermas says that his approach "offers a guide for reconstructing the network of discourses that, aimed at forming opinions and preparing decisions, provides the matrix from which democratic authority emerges".[43] As one might expect from this, insofar as the account involves social theory, it is social theory of a 'reconstructive' kind that seeks to bring to light the "intuitive knowledge of subjects", that is, what is implicit in the "the know-how of subjects who are capable of speech and action" and thereby to "provide an account of the pretheoretical knowledge and the intuitive command of rule systems that underlie the production and evaluation of … symbolic

---

[41] This is true even if the bias results not from the content of the education but from the factors that select persons to receive the education.

[42] This is an inclusive 'or'. The three terms are not mutually exclusive.

[43] BFN, 5.

expressions and achievements".[44]   In addition, the locution "emerges" suggests that democratic authority is, in an important sense, an emergent property, and this hints at a difference between Estlund and Habermas in the way authority is seen.[45]

Estlund's argument purports to show that a jury system, and by extension democratic procedures, would have 'original authority' in the sense that they have an authority to require action that (a) does not depend upon acts of explicit consent and (b) is not traceable to the authorization of any prior political procedure, thereby avoiding the threat of an infinite regress.[46] The thought that some form of authority must be original is plausible and it would be natural for Estlund, once convinced of the mistake of reading Habermas as a political nihilist, to ask where one might find 'original' authority in Habermas's theory. If democratic authority emerges from anything, it ought to be something of this kind. And in fact there does seem to be something in the discourse theory that meets the two conditions of original authority as defined: namely, "the intersubjective authority of a common will".[47] This answer requires some explanation.

The idea of authority has been so far expressed in terms of the notion of a normative power of requirement. A norm of action, say, a prohibition of killing, can understood as embodying such a requirement. Its content can be expressed in indicative statements of either impersonal ('It is wrong/not permitted to kill') or second-personal types ('You are not allowed to kill') or in imperative statements ('Do not kill!'). This content can also be

---

[44] Jürgen Habermas, *Moral Consciousness and Communicative Action* (Cambridge, MA: MIT Press, 1990), p. 31. Hereafter MCCA. Cf. Habermas' statement, during his attempt at a 'transcendental-pragmatic' justification of the universalization principle, that "I must appeal to the intuitive preunderstanding that every subject competent in speech and action brings to a process of argumentation" (ibid., 89).

[45] As a matter of fact, the German original would be translated differently in a more literal rendering. But, firstly, the claim that Rehg's rendering expresses is perfectly consistent with the broader argument of *Between Facts and Norms*, and secondly, he states that Habermas "had a considerable hand in the translation, in some cases adapting and rewriting the text for the Anglo-American audience" (BFN, xxxvi). So I think there is nothing wrong with my placing some weight on a suggestive connection between "emerges" and the idea of an emergent property.

[46] DA, 136, 157.

[47] MCCA, 160.

expressed in the language of duties and rights, i.e., a duty not to kill and a right not to be killed. One can ask where such a norm gets its 'authority', i.e., its binding or 'requiring' nature. As analyzed by P.F. Strawson, the existence of reactive attitudes such as resentment and indigation testifies to the deeply embedded presumption that certain expectations concerning human interaction are not simply like expectations about how physical objects will behave, but rather embody norms held to be valid precisely in the sense of being authoritative or requiring that we act or not act in certain ways.[48] Hence their violation requires a justification rather than a causal explanation. The same point reveals that such norms pervasively structure social life and human interaction, since the reactive attitudes can clearly be activated in any kind of social context. It also suggests that these attitudes and norms are part of a web of relations of intersubjective recognition in which, as Strawson shows, actors mutually acknowledge each other as accountable, responsible persons that demand a basic form of respect. This in turn implies that cases of conflict, or an 'upset' of the balance of relations of recognition between actors, can arise that do not necessarily stem from the violation of a prior mutually accepted norm, but rather center on whether an existing (or proposed) norm is one that should be (deserves to be) seen as authoritative (valid, right, worthy of recognition).[49]

> Conflicts in the domain of norm-guided interactions can be traced directly to some disruption of a normative consensus. Repairing a disrupted consensus can mean one of two things: restoring intersubjective recognition of a validity claim [i.e., a norm] after it has become controversial or assuring intersubjective recognition for a new validity claim that is a substitute for the old one. Agreement of this kind expresses a *common will*. … What is needed is a "real" process of argumentation in which the individuals concerned cooperate. Only an intersubjective process of reaching understanding can produce an agreement that is reflexive in nature; only it can give the participants the knowledge that they have collectively become convinced of something.[50]

---

[48] Cf. MCCA, 45-50. For Strawson's classic discussion, see his "Freedom and Resentment", in *Freedom and Resentment and Other Essays* (London: Routledge, 2008).

[49] Habermas more commonly uses the terms in parentheses, but I use 'authoritative' here to maintain the connection with authority.

[50] MCCA, p. 67. It would be instructive to figure out why "real" is in quotation marks in this passage.

This ultimately means that if the disturbed relationships of recognition are to be repaired or the question concerning the authority (validity) of a norm is to be resolved, the actors concerned must arrive at a solution they all jointly affirm for what they take to be good reasons, where "good reasons" is just a placeholder for whatever participants themselves, in an appropriately structured process of argumentation, find convincing.[51] A norm that meets this test has earned the title of being valid or having authority. Since it is effectively moral norms that are in question in the justification of democratic authority, they are the only ones being considered here for the moment. Remember that even the 'original' authority of a jury system or of democracy in Estlund's argument was based on a presupposed background of moral standards and values. Insofar as rational moral discourse turns out to be the basis of that power to require or forbid action that valid moral norms have, i.e., of their authority, moral discourse is more original than the 'original' authority of Estlund's jury/democracy analogy. In sum, the only original authority that Habermas could ultimately recognize is that which is intrinsic to the communicative rationality of discursive justification. Paradoxically, this root of political authority thus resides in what he elsewhere describes as the 'anarchy' of communicative freedom.[52] If the discourse theory provides an adequate explication of this even more original authority, it would demonstrate yet another possible advantage over

---

[51] This is reflected in the 'formalism' of the (U) principle that Habermas offers. To recall, it reads, "A norm is valid [morally right] when the foreseeable consequences and side-effects of its general observance for the interests and value-orientations of *each individual* could be *jointly* accepted by *all* concerned without coercion" (see fn. 21 for source). This principle provides no real guidance concerning which reasons are good reasons. To say that any reason that provides significant support to a norm all can accept in light of the interests and values of each is a 'good' reason will not help pick out any specific reasons. One must provide content for the other variables: interests, values, consequences, side-effects, proposed norm, etc. Only those actually involved are in a position to do this. In addition, the descriptions of each of these elements are themselves subject to critical revision in the course of argumentation, so they are not fixed inputs that can be univocally related to a set of possible outputs by means of any kind of function or algorithm. Hence, (U) is not a criterion that can be well used, independent of actual discourse by those affected, to determine the validity of a norm.

[52] Cf. the linkage of anarchism and communicative power in "Popular Sovereignty as Procedure", BFN, 480-483. The phrase is also the title of collection of essays on the theory of communicative action in international relations theory and Habermas' recent writings on cosmopolitanism, *Anarchie der kommunikativen Freiheit*, eds. Peter Niesen and Benjamin Herborth (Suhrkamp: Frankfurt am Main, 2007).

Estlund's epistemic proceduralism.[53]

Here, however, Estlund would be in a position to reformulate his charge of political nihilism, but now at a much deeper level as a more general procedural or moral nihilism. He might say, "according to you, then, the power to require or forbid action, i.e., authority, does ultimately rest in a procedure, so I was right about that. But now it is supposed to be not a *political* procedure but some other rather airy 'moral' procedure. But if there are no procedure-independent standards with which to judge the outcomes of *this* procedure, then you *are* a kind of nihilist, since you are saying that at the *moral* level we can't distinguish between 'good outcome' and 'outcome of the properly performed procedure'!" In reply to this challenge, two points can be made.

In the first place, part of the motivation for the (hypothesized) objection may lie in another misunderstanding of the discourse theory. Estlund falsely assimilates it to a contractualist model for the justification of a system of moral norms as a whole. On his reading of contractualism it would be circular for participants in the contractualist procedure to be already "armed with views about justice or rightness", since the point of the procedure is to *define* which norms or rules are just or right. Hence they have to focus on "less impartial reasons" than those of justice and be armed with individual veto powers to be protected "against the pursuit by others of their own personal concerns".[54] If having views about what particular norms or actions are just or right counts as being "armed with views about justice or rightness", then it is false that participants in moral discourse as conceived by Habermas lack such views. So is the discourse theory guilty of a vicious circularity? No. It is not

---

[53] I only say "possible" because there is another sense in which it could be a disadvantage because discourse theory sticks its neck out, so to speak, on the contentious matter of moral justification or moral validity as such, whereas epistemic proceduralism simply assumes (plausibly) that its addressees agree on certain moral matters without exposing any vulnerable flanks on the issue of the validity or justification of those agreements. In this way, epistemic proceduralism at least appears to lower its burden of proof. On the other hand, if anyone wonders about why and whether such background moral norms are valid in politics, i.e., can ground any form of *political* authority, then some explicit account of the authority of moral norms and why it extends to other contexts of action may be required. Here discourse theory stands ready.

[54] DA, 172.

contractualist in Estlund's sense.[55] It aims at explicating the moral point of view as it unfolds from the standpoint of participants in argumentation about whether a disputed norm is worthy of recognition or just or right. The participants are assumed to have a pre-theoretical, intuitive grasp of what it is for something to be seen as a morally valid (or invalid) norm, of ideas like justice, rightness, or wrongness, and of what it is to engage in serious rational argumentation, among other things.

Furthermore moral discourses of justification are envisaged as taking place against indefinitely extensive background agreements that, as I understand it, *already include* various moral norms, moral ideas, and wider value orientations (which may overlap with moral ideas). Discourse ethics is not designed for Moral Year Zero situations in which participants are supposed to construct moral norms with the spare resources of putatively non-moral elements like individual interests narrowly understood or desires. In other words, the standard situation is one in which, when the justification of a disputed norm is at stake, other moral norms, as well as values or interests non-morally understood, are available for deployment in making arguments pro and con.[56] We can dispense with the phantasmagoric scenario in which we are confronted with the question, "Suppose the validity of *all* the moral norms you currently accept (along with the semantically embedded values which are inseparable from them) were suspended, what moral norms would you be able to justify then?" Such fantasies may excite foundationalists,[57] but here the pragmatist point is well taken: we can question any

---

[55] I leave aside the question whether Rawls, Scanlon or Barry are contractualist in this sense either.

[56] The debates around abortion are a good example. Broadly speaking, the universally accepted norm expressed in a human right to life is used by opponents of abortion, while on the other side similarly recognized norms expressed in rights to physical integrity or rights to privacy are invoked.

[57] See Joseph Heath's "Foundationalism and Habermas' Discourse Ethics", *Philosophy and Social Criticism*, v. 21, n. 1 (1995): 77-100, for a criticism of passages in Habermas' presentations of discourse ethics, in particular his first presentation of the (U) principle, that easily give rise to the impression that his own ethical theory has a problematic foundationalist structure of justification. Heath also provides strong arguments, both internal and external to the theory, for why this is undesirable and avoidable. For a survey of the issues here in moral philosophy more generally and a defense of a broadly Kantian, or at least counter-Humean, non-foundationalism, see his "Foundationalism and Practical Reason", *Mind*, v. 106, n. 423 (1997): 451-474.

belief or norm we have, but we cannot question all our beliefs or all our norms at once.[58]

Two things follow from this. First, the outcome of any instance of moral argumentation can be criticized from the standpoint of any other currently accepted moral norm. Second, one need not envision each currently accepted norm as the 'outcome' of a moral 'procedure' (whether actually carried out at some point or hypothetical). Discourses ethics elaborates a view about the justification and validity of norms, not their genesis. Hence there may be many moral norms which are 'procedure-independent' in a genetic sense.[59]

In the second place, we must address an additional question that the objection raises. This one relates not to misunderstandings relating to foundationalism and contractualism, but to something which is in fact a genuinely open question within the camp of discourse ethics. It is also one possible construal of Estlund's worry about procedure-independence. The question here is whether rational acceptability under ideal discursive conditions is to be understood as *constituting* normative or moral rightness or as simply the best (perhaps ultimately the only) *epistemic* support for such claims. Habermas' official position is that ideal rational acceptability constitutes rightness.[60] But one can find passages supporting both

---

[58] For the *locus classicus*, see Charles S. Pierce's classic essays, "The Fixation of Belief" and "Some Consequences of Four Incapacities", in *Philosophical Writings*, ed. Justus Buchler (Dover: Indianapolis, 1955), pp. 5-23 and 228-51 respectively.

[59] Even with respect to their justification or validity, there is a sense in which a given moral norm may be procedure-independent from the standpoint of Habermas' theory. This results from an ambiguity in "justification". The "justification" of a norm may refer to a process of justification or to a status of being justified. It is certainly possible for a moral norm to have the status of being or counting as justified (or valid) independent of any actual procedure of justification. This follows directly from the nonfoundationalist or broadly pragmatist nature of Habermas' view of justification, according to which justification is not more basic than agreement, since all justification takes place on the basis of background agreements. However, once such a norm comes into dispute it has by definition lost this status, and its justification can then only mean that it survives argumentative testing. On the nuances of the epistemological questions, see (again) Joseph Heath, *Communicative Action and Rational Choice* (Cambridge, MA: MIT Press, 2001), pp. 200-5.

[60] Habermas also long defended a reading of truth as ideal rational acceptability but in his 1996 essay, "Rorty's Pragmatic Turn" (in *The Pragmatics of Communication* [Cambridge, MA: MIT Press, 1998], pp. 343-83), he abandoned this after criticism from writers such as Albrecht Wellmer, Hilary Putnam, and Cristina Lafont, among others. Lafont has pressed her case for abandoning it with respect to rightness in *The Linguistic Turn in Hermeneutic Philosophy* (Cambridge, MA: MIT Press, 2002), ch. 7. However, she frames the issue in terms of realism versus anti-realism. This move is questioned by Heath, *Communicative Action and Rational Choice*, ibid., and, as I understand him, Rainer Forst, *Recht auf Rechtfertigung* (Suhrkamp: Frankfurt am Main, 2007), cf. pp. 9-100. Habermas makes clear in his response to Lafont that he is *not* changing his view on rightness, cf.

readings in his texts.[61] In our context, the relevant issue is whether the 'procedures' of moral justification are to be understood as 'pure' or 'imperfect' in Rawls' terms.[62] In other words, if moral argumentation regarding a disputed norm has been carried out by all affected under ideal discursive conditions and each accepts the norm as worthy of recognition, can we still meaningfully ask whether it is morally right?[63] If the answer to this is no, then it looks like we make ideal moral discourse into a case of pure procedural justice and collapse 'good outcome' into 'outcome of the properly run procedure', and hence fall under Estlund's suspicion of procedural 'nihilism'. But then it is very difficult to see how one can preserve the cognitivist and fallibilist intuitions that Habermas clearly wants to preserve.[64] For it certainly seems that human beings could get things morally wrong even under ideal discursive conditions (where those conditions are not made so ideal that the participants no longer look human).[65] While there is no space to argue out the complicated details here, in the dialectic

"Rightness versus Truth", in *Truth and Justification* (Cambridge, MA: MIT Press, 2003), pp. 237-77. Lafont has replied, to my mind convincingly, and further clarified the issues at stake in "Procedural Justice? Implications of the Rawls-Habermas Debate for Discourse Ethics", *Philosophy and Social Criticism*, v. 29, n. 2, pp. 163-181, and "Moral Objectivity and Reasonable Agreement: Can Realism Be Reconciled with Kantian Constructivism?", *Ratio Juris*, v. 17, n.1 (March 2004), pp. 27-51.

[61] Lafont, "Procedural Justice?", ibid.

[62] Rawls, *A Theory of Justice*, rev. ed., (Cambridge, MA: Belknap, 1999), pp. 74-77.

[63] Lafont, "Procedural Justice?", ibid., p. 172.

[64] Cf. Habermas, "Rightness versus Truth", ibid., p. 245, where Habermas speaks of a "double fallibilist proviso".

[65] Briefly, Habermas's double fallibilist proviso sets out two sources of fallibility in moral discourse. One relates to whether the presuppositions of discourse (e.g., inclusion of all affected, equal argumentative rights, absence of deceit and coercion) are fulfilled or at least sufficiently approximated. The other relates to whether the consequences and side-effects of observing a proposed norm turn out to be as expected. If the *actual* fulfillment or sufficient approximation of the presuppositions of discourse is built into the notion of ideal rational acceptability, then it is true by definition that genuinely valid or right norms are those and only those that were justified in (sufficiently) ideal discursive conditions. If in hindsight we realize those presuppositions were not adequately met when we justified a given norm, then it immediately follows that it was not justifiably accepted in the first place. But when we turn to the question of foreseeing consequences matters are different. If we build fully accurate foresight into the definition of ideal discourse, we cease to have an idea of something we can imagine human beings doing, and the idealization is no longer relevant to *us*. However if we rightly allow for fallibility regarding consequences, then in cases where discursive conditions were ideal but the consequences turn out to be so different than we anticipated that the adopted norm is clearly not equally in the interest of all involved, the constitutive reading of rightness in discourse ethics leads to a fatal paradox. For it yields the conclusion that the norm was right (since it was ideally rationally acceptable), but also not right (since it was not equally in the interest of all). The participants themselves would judge the norm to have retrospectively failed to meet the conditions embodied in the (U) principle that discourse ethics says is the rule of argumentation that

with the epistemic proceduralist it seems better to follow a strategy like that which Lafont suggests and present discourse ethics as offering a purely procedural notion of *rational acceptability*, but a non-procedural or 'procedure-independent' notion of moral rightness (as that which is equally good for all or equally in the interest of all).[66] In other words, even ideal moral discourse is an imperfect epistemic procedure that only offers strong support for believing that norms surviving argumentation are right.[67] But it is one for which we have no functional equivalents, hence it is the best kind of epistemic support we can ever get.[68] There is no threat of procedural nihilism at the moral level.[69]

At this point, the epistemic proceduralist may say, "Now, hold on. If you've made anything plausible by this point, it's only some supposed 'authority' of norms worthy of recognition in (sufficiently ideal) moral discourse. These have whatever original authority communicative reason has. But this just sounds like a description of a 'horizontal' form of authority relationship in which norms that *everyone* accepts in the (supposedly) symmetrical

---

operates in moral discourse. The option of temporalizing the rightness predicate – saying the norm was right until its consequences and side-effects were clearly not in the equal interest of all and thereafter not right – is not available to Habermas unless he gives up a key analogy between truth and rightness: namely, that the rightness of norms, like the truth of propositions, is not a property that can be lost. This is part of what is entailed by saying they are 'universal' or 'context-transcendent' validity-claims that have a moment of 'unconditionality'.

[66] Lafont, "Procedural Justice?", ibid., p. 169. If it is objected by the epistemic proceduralist that the latter construal of what rightness or justice means is too controversial for political justification, perhaps we can leave its definition open and hold on to a thin set of agreed contents like Estlund's primary bads and the pure procedural understanding of rational acceptability. However this might threaten to collapse the discourse theory into epistemic proceduralism.

[67] Whether the resulting position is a kind of moral realism, as Lafont argues, can be left open for our purposes.

[68] "The fact that there are *no alternatives* to these rules of argumentation is what is being proved; the rules themselves are not what is being *justified*" (MCCA, p. 95). What Habermas says of communicative action in general is true of moral discourse in particular: "there is no other equivalent medium in which these functions can be fulfilled" (MCCA, p.102; cf. the arguments on, ibid., p. 100). For discourse ethics, it is crucial to maintain the *necessity* of discourse as an epistemic means to moral knowledge and as a conceptual explication of moral impartiality. For arguments that it is not necessary to the explication of impartiality and that the moral reasons discourse tracks are just reasons of fairness, see Christopher McMahon, "Discourse and Morality", *Ethics*, v. 110 (April 2000): 514-36, and "Why There is No Disagreement Between Rawls and Habermas", *Journal of Philosophy*, v. 99, n. 2 (2002): 111-29. For a discourse-theoretical response, see William Rehg, ""Grasping the Force of the Better Argument: McMahon versus Discourse Ethics", *Inquiry*, v. 46, n.1 (March 2003): 113-33.

[69] In addition, it is worth noting that the position that results is much closer to Estlund's own use of a model of 'ideal epistemic deliberation' than the misleading caricature he criticizes as political nihilism. On that model, see DA, 172-3.

conditions of discourse have the power to require or forbid action. But there must be a 'vertical' authority relationship at some point. We are, after all, talking about politics. Where is the state apparatus and the formal decision-making bodies that claim to give laws that are binding on all citizens and subjects regardless of whether they accept any particular law or not?"

It is a noteworthy fact about the work of Raz, the source for the definition of authority that Estlund uses, as well as the work of those who follow him such as Green, Morris, Simmons and McMahon, that democracy is effectively absent from the basic framework used for justifying political authority.[70] From the standpoint of the Rousseau-Kant tradition in which Habermas stands, the image presented is a rather reified and rigid one that portrays a relationship between two separate parties: a ruling authority standing over against an independently conceived body of ruled persons. The authority is literally the author of directives and the subjects are simply the addressees of those directives. The question is then asked, "How can we justify subjects taking these directives as pre-emptive reasons for action?" One then naturally seeks some properties of the putative authority that could justify such a subordination of subjects' own judgment to it. As Habermas points out, though, "Rousseau and Kant, by contrast, did not conceive of popular sovereignty as the transfer of ruling authority from above to below or as its distribution between two contracting parties. For them, popular sovereignty signified rather the transformation of authority into *self-legislation*" ("Citizenship and National Identity", BFN 496). Hence it turns out that Habermas paved the way for criticisms of the Razian picture by recent democratic theorists:

> As Samuel Hershovitz puts it in his critique of the Razian approach to the legitimacy
> of authority, in a democracy there is no sharp division between the "binders" and the

---

[70] This is noted by Thomas Christiano, "The Authority of Democracy", *The Journal of Political Philosophy:* Volume 12, Number 3, 2004, pp. 266–290: 277n10. In Raz's recent attempt to address criticisms of his view of authority, he still fails to make any mention of democracy even though he touches upon the issue of whether the rigid distinction between ruler and ruled can be blurred. The only way he envisages this being able to happen is through psychological forms of identification. See Joseph Raz, "The Problem of Authority: Revisiting the Service Conception", *Minnesota Law Review*, 90 (2006), pp. 1003-1044, esp. pp. 1040-44.

"bound". The political authority of the democratic assembly is entailed by some account of the conditions under which citizens may legitimately exercise coercive power over one another and vice-versa. For this reason, political authority, in a democratic context, does not exist independently of legitimacy.[71]

I think Habermas' political theory is best seen as following this order of explanation. We must start with the question of legitimacy, or the conditions for the moral permissibility of coercion, before we can fully answer questions about political authority. Some independent support for this reading comes from Arthur Ripstein's excellent and provocative recent work on Kant, in particular the way he makes very clear the difference between Kant's approach and the dominant approach in the Anglophone literature on political authority, legitimacy and political obligation.[72] The dominant approach deals first with the question of when and in what respects states are "entitled to tell you what to do" (the problem of authority) and only later asks when they are allowed "to force you to do as you are told" (the problem of coercion or legitimacy).[73] Kant's approach reverses the order of explanation in political philosophy and starts with the conditions of possibility of legitimate coercion. It views coercion as legitimate when it prevents or rectifies the subordination of one person's will to that of another. It turns out that this requires that any coercion be grounded in a "reciprocal authorization", i.e., in a general or universal will.[74] In other words, norms which are enforced by legitimate coercion must have the following property: all of their addressees must also be able to conceive of themselves as their authors. This is the idea of self-legislation that seemed to take the place of Razian authority in the last citation from Habermas. Translated into the terms of democratic theory, "the permission for legal coercion must be *traced back* to the *expectation of legitimacy* connected with the decisions of the legislature", but the

---

[71] Fabienne Peter, "Political Legitimacy", *Stanford Encyclopedia of Philosophy*, April 2010: here section 3.3. The work cited by Peter is Scott Hershovitz, "Legitimacy, Democracy and Razian Authority", *Legal Theory*, 9 (2003), 201-220: here 210f.

[72] Arthur Ripstein, "Authority and Coercion", *Philosophy and Public Affairs*, v. 32, n. 1 (2004), pp. 2-35. For a fuller elaboration of his interpretation of Kant, see *Force and Freedom: Kant's Legal and Political Philosophy* (Cambridge, MA: Harvard University Press, 2009).

[73] Ripstein, "Authority and Coercion", ibid., p. 2.

[74] Ripstein, "Authority and Coercion", p. 10-11, 27, 32-5.

legislative decisions of this formal body can only adequately embody a form of *self-legislation*, and hence truly authorize legitimate coercion, insofar as *they* can be traced back to the 'original' authority (or non-authority from a Razian viewpoint) of the "jointly exercised communicative freedom of citizens", which supposedly "can assume a form that is mediated in a variety of ways by legal institutions and procedures".[75]

Yet something seems to be missing from the picture. Even supposing that it can legitimate specific acts of coercion, why should the joint exercise of citizens' communicative freedom result in the need for coercion at all? Where does the justification for the specifically coercive element to be applied 'vertically' by a state apparatus (as opposed to the purely horizontal, intersubjective justification of the authority of moral norms) come from? Kant provides a relatively clear answer to this question, but at first glance Habermas seems to simply evade it. It will help to keep in mind that for Habermas, as for Kant, coercion is an intrinsic feature of law, or put otherwise, the concept of coercive enforcement is inherent in the concept of law:

> One can leave open the further question of whether there are *moral* grounds for entering a legal order in the first place—the problem that rational natural law posed as the transition from the state of nature to civil society. The positive law that we find in modernity as the outcome of a societal learning process has formal properties that recommend it as a suitable instrument for stabilizing behavioral expectations; there does not seem to be any functional equivalent for this in complex societies. Philosophy makes *unnecessary* work for itself when it seeks to demonstrate that it is not simply functionally recommended but also morally required that we organize our common life by means of positive law, and thus that we form legal communities.[76]

In the first sentence, Habermas seems to be clearly rejecting the Kantian construction of a moral duty to exit the state of nature and enter into a civil condition of public law (a state).[77] In the last sentence, I think there is an indication that this rejection rests on a misunderstanding, for the Kantian argument purports to show, not that we are *directly* morally

---

[75] BFN, 33.

[76] BFN, 460.

[77] This claim is grounded in the third universal duty of right, the *lex iustitiae* (*Doctrine of Right*, 6:236-7). It is clearly said to be a command of practical reason in the remark to §44 of the same work.

required to create legal communities or to make use of positive law, but rather that we are morally required not to subject others unilaterally to our own will and it turns out that positive law is a necessary *means* to avoiding this. To make this clear I will present (with criminal brevity) Kant's account.

The bare concept of right tells us that an action is right if it accords with a law deriving from reciprocal and equal restrictions on the freedom of each. According to Kant, it follows from this idea that coercively preventing actions that unilaterally attempt to restrict freedom is right.[78] But this abstract concept leaves us with problems of indeterminacy and assurance if we try to act on it as individuals in an uncoordinated fashion, *even if we all intend to act rightfully*. If we try to implement our own individual interpretations of what is right here and now we are simply imposing our own practical will on others who have equal title to implement their interpretations.[79] Additionally, if the interpretation that gets enforced in such a situation is simply the one backed by the greatest individual strength, then again we have subordination and power in the place of right. Hence, only the establishment of a public determination of the content of right and a public coercive power that equally subjects all to those determinations can achieve the desired result: "both the use of official force and the claim of states to tell people what to do are justified because, in their absence, arbitrary individual force prevails, even if people act in good faith".[80]

On the other hand, when Habermas states his 'functional explanation' for why postconventional morality requires law, he mentions three kinds of 'demands' of which persons in modern society have to be 'unburdened' (again for social-functional reasons): the *cognitive indeterminacy* involved in applying abstract moral norms to complex circumstances, the *motivational uncertainty* arising from lack of assurance that others will follow them too, and lastly the organizational problem of accountability or how to impute which obligations to

---

[78] For Kant, "right" and "entitlement to coerce or compel" *mean* the same thing (*Doctrine of Right*, §E, 6:232).
[79] Cf. *Doctrine of Right*, §8.
[80] Ripstein, "Authority and Coercion", p. 3.

whom in conditions of a complex *division of sociomoral labor* (e.g., with regard to duties of aid or social welfare functions).[81] As a result, "in complex societies, morality can become effective beyond the local level only by being translated into the legal code".[82] But the first and the second factor seem clearly to be forms of just those indeterminacy and assurance problems that appear in Kant's argument, and the third is arguably simply another aspect of the indeterminacy problem. The difference between such a 'functional account' of why law is a necessary 'complement' to morality and the structural features of Kant's own argument is hard to see. Indeed, in an earlier essay Habermas mentions precisely the same factors and says that together they constitute "a moral reason for law in general" and even a "normative justification for the transition [!] from morality to law".[83]

In light of this, it is perhaps telling that Habermas ends the long passage cited above with something that seems very close to a restatement of the Kantian point: "the philosopher should be satisfied with the insight that in complex societies, law is the only medium in which it is possible reliably to establish morally obligatory relationships of mutual respect even among strangers".[84] Firstly, it is unclear why this justification, such as it is, has to wait for modern societies to come into being, given that the force of the indeterminacy and assurance problems does not seem to depend on the presence of recognizably modern conditions. Once certain thresholds of scale and complexity are reached, the three problems requiring law to functionally complement morality emerge. It seems quite plausible that these thresholds were crossed long before modernity. Secondly, if the brief presentation of Kant's general line of

---

[81] Habermas seems to reject a 'normative justification' in this context because "the legal form is not a principle one could 'justify'" (BFN, 111-2). But the Kantian argument does not justify 'the legal form', it provides an account of the legitimacy of coercion that grounds the authority of a public system of law to determine what will be coercively enforced while preserving the freedom of all. Law as such is not 'justified'. It is a functional means for the expression and implementation of a public will, much as it is in Habermas' account.

[82] BFN, 110. If part of the point of morality is that, put crudely, we are supposed to do as it says, and if morality can only be effective (i.e., we can only do as it says) when we translate it into law, then doesn't it follow that morality says we should translate it into law?

[83] *Justification and Application* (Cambridge, MA: MIT Press, 1994), p. 88.

[84] BFN, 460. Translation amended.

thought above is correct, it is precisely because law is the only (available) medium to solve these problems that we are required to use it. This suggests that in principle the structure of Kant's argument could be reformulated within the discourse theory, modulo various adjustments having to do with certain fine details of metaethics.[85]

If this is right, we would have a neat way to present the Habermasian alternative strategy for dealing with epistocracy and providing an account of the authority and legitimacy of democracy. As we saw earlier, the intersubjective account of moral justification (and hence moral knowledge) of the discourse theory undercuts the knowledge tenet of epistocracy, thereby defeating it without having to move on to tackle the authority tenet. It also points toward an intersubjective, 'horizontal' account of normative authority that bursts the bounds of the (Razian) image of authority that shapes Estlund's discussion. In order to recapture the 'vertical' dimension of authority essential to polities, we had to see that the discourse theory, following Rousseau and Kant, both refuses to separate the issue of legitimacy and coercion from the issue of political authority and reverses the predominant order of explanation, deriving the justification for 'vertical' state authority from the 'horizontal' conditions of legitimate coercion. I argued that this seems to require Habermas to have an argument like Kant's and tried to show that all the basic elements for such an argument are already present in the discourse theory.

There are many interesting issues left unaddressed here, such as the respective differences between epistemic proceduralism and the discourse theory regarding the problems of political obligation, the minority democrat or civil disobedience, to name just a few. I hope to have shown, however, that the differences between Estlund's theory and that of Habermas

---

[85] Habermas also feels he must reject the idea that positive law must 'mirror' the moral law which he still sees in Kant. But it seems to me that the indeterminacy problem alone at least greatly diminishes this worry. For how could positive law in a state reasonably be expected to 'mirror' precisely those abstract moral ideas whose indeterminacy is its own reason for being? Positive law is precisely there to specify their unworkably abstract content and 'mirroring' is not an appropriate metaphor for this endeavor. So if the mirror idea is still present in other aspects of Kant's legal and political philosophy, it is doubtful that it affects the specific argument presented here.

are both more subtle and more interesting than Estlund's misguided dismissal of the latter as a political nihilist allows. In particular, Habermas offers an intriguingly radical way of dispensing with the threat of epistocracy and thus "making truth safe for democracy".[86] Whether one favors the strategy of epistemic proceduralism or that of the discourse theory, however, will depend, among other things, on one's assessment of the respective merits of an intersubjective account of justification, in particular of moral justification, or more open-ended appeals to standards of qualification or reasonableness, as well as the extent to which one believes that keeping a close connection between the concepts of normative theory and the social sciences is an advantage.

AUTHOR CONTACT INFO:

Jonathan Trejo-Mathys
Assistant Professor of Philosophy
Department of Philosophy
Maloney Hall, 21 Campanella Way
Boston College
Chestnut Hill, MA 02467
jonathan.trejo-mathys@bc.edu

---

[86] David Estlund, "Making Truth Safe for Democracy," in Copp, Hampton, Roemer, eds., *The Idea of Democracy* (Cambridge: Cambridge University Press, 1993).