

Motivational Facts, Legitimacy, and the Justification of Political Ideals

Lior Erez¹

Accepted: 23 June 2023 / Published online: 20 July 2023 © The Author(s) 2023

Abstract

Should facts about motivation play a role in the justification of political ideals? Many theorists argue that political ideals should be tailored to the limitations of human nature—'taking people as they are'—while others maintain that facts about motivation should be excluded. This article offers a critical intervention in this debate: the important question is not so much whether people can motivate themselves, or whether they are capable of being motivated, but what social mechanisms would be required to motivate them, and whether these mechanisms are legitimate. Reframing the question of motivation as a political question of legitimacy, I argue that if people could only be motivated to act through illegitimate use of power, the ideal in question cannot be fully justified.

Keywords Feasibility · Legitimacy · Methodology of political theory · Motivation · Normative constraints · Political realism

Introduction

Over the last two decades, a growing number of political theorists have been interested in questions of methodology, aiming to explicitly defend or challenge the formerly implicit assumptions of the field. One such underlying assumption is that political ideals should be tailored to the limitations of human nature, endorsing the Rousseauvian maxim of 'taking people as they are' (Rousseau 1762). If people are too self-interested for socialism, too aggressive for anarchism, too tribal for cosmopolitan democracy, then these utopian ideals cannot be justified—or so goes the argument.

My main aim is to offer a critical intervention in this debate, and defend an argument which has so far been largely neglected. After laying out the prominent positions of the debate, I will argue that there is a strong reason for political theorists to

Department of Politics and International Relations, University of Oxford, Oxford, UK



[☐] Lior Erez lior.erez@politics.ox.ac.uk

be concerned with motivational facts as part of the justification process of political ideals. As I demonstrate below, both sides of the debate are unduly concerned with what individuals are capable of bringing themselves to do, or with the non-compliance of others as a background condition for the individual ability to act. These considerations, while important, sideline a central normative problem of political power: making others do what they otherwise would not do.

The normative question of power, I argue, is not so much *whether* people could be motivated to act in certain ways, but what social mechanisms would be required to motivate them, and whether these mechanisms are legitimate. Taking the problem of political legitimacy into account, I argue that if people could only be made to act through the illegitimate use of political power, the ideal in question cannot be fully justified. However, I also maintain that critics of motivational constraints are right to argue that motivational facts, by themselves, cannot relieve individual agents of their moral duties. But, as I argue below, this issue is marginal to political life: motivational facts matter specifically to the justification of *political* ideals given a background theory of legitimate means. By reframing the question of motivation in this way, I suggest a way for political theorists to move beyond the impasse in the methodological literature.

Whose Motivational Facts? The State of the Debate

By 'political ideal', I mean a vision or model for a desired political state, which provides a benchmark for evaluating the compatibility of various normative principles with that ideal. This ideal can take many forms, from a utopia that represents the perfect society to a more realistic target for political reform of a particular practice (cf. Wiens 2015, p. 448). A political ideal is distinct from a principle in this sense; it cannot be reduced to a principle or a set of principles but is rather the institutions and social practices that manifest these principles in the world. 'To each according to his needs, from each according to his ability' is a principle. 'A Communist Society' is a political ideal where this principle is manifested, including the institutional, social, and dispositional assumptions underlining it. My account in this article seeks to examine the role that motivational facts play in the debate over justification of political ideals, understood in this way.¹

This debate is a broad and messy one, and while my aim in this section is to map the different positions on offer, I do not claim this to be an exhaustive survey.² Instead, I want to highlight two ways of thinking of the problem of motivation prominent in the existing literature that cut through orthogonal debates on ideal/non-ideal theory, realism and moralism, and the like. The first, and more prominent, is the focus on the ability of agents to bring themselves to act, and the second is the

² For good overviews, see (Erman and Möller 2015b; Gilabert 2012; Sleat 2016; Valentini 2012).



¹ Note that my definition of what constitutes a 'political ideal' differs from more abstract interpretations, where 'ideal' and 'principle' may be interchangeable. See for example (Cohen 2008; Hamlin and Stemplowska 2012).

view of others' motivations as background conditions for one's normative duties. For simplicity's sake, I will henceforth refer to the former as 'the self-oriented view' and to the latter as 'the background conditions view'. I wish to demonstrate that these approaches, while valuable, are of limited use when considering the problem of political power.

The self-oriented view asks whether the agent is able to bring himself to act. If 'ought' implies 'can', then one cannot be morally obliged to do the impossible; and if we assume that psychological impossibilities exist—that some actions are at odds with human nature, for example—this applies to motivation as well. James Griffin, for example, argues that '[s]ince ought implies can, what lies outside natural human motivation does not even enter contention for being a moral requirement. Norms are not fashioned in a vacuum; they are cut to fit agents like this' (Griffin 1990, p. 128 [my emphasis]; cf. 1998, pp. 87–93; 2015, pp. 23–43).

The success of this kind of argument depends on our definition of ability, which in itself is a contentious issue. In his influential account, David Estlund argues that if a normative theory sets standards which are within the ability of individuals and institutions to follow, and it is merely unlikely that they will follow, this does not render the theory hopelessly utopian but merely *aspirational* (Estlund 2008, pp. 263–269). In Estlund's formulation, even if we accept that 'ought implies can', it is false that a lack of ability to muster the required motivation is 'requirement-blocking' (Estlund 2011, pp. 230–231). This is because Estlund's definition of ability is conditional: 'a person is able to (can) do something if and only if were she to try and not give up, she would tend to succeed' (Estlund 2011, p. 212). Importantly for my purposes, this claim is applicable not just to fundamental political principles (e.g., equality), but also for what he calls institutional principles, which are 'institutional arrangements as part of a broader prescription or proposal, even if the described arrangement itself is not proposed or prescribed' (Estlund 2011, p. 218).

In their account, Pablo Gilabert and Holly Lawford-Smith draw a distinction between 'hard' and 'soft' constraints, classified according to their malleability, and the probability of overcoming them. Different kinds of facts play different roles in the justification of a normative theory, depending on whether we consider them to be 'hard' or 'soft'. Hard constraints are constant features of the world, and thus their effect on feasibility is binary—no agent, under any circumstances, is able transcend them, and they are categorically ruled out of the feasible set (Gilabert and Lawford-Smith 2012, pp. 812–816). For example, logical impossibilities, or actions that violate the laws of nature, are hard constraints. Soft constraints, on the other hand, are scalar; because they are changeable and malleable, they affect only the probability, not the possibility of an action. Institutional, cultural, or economic facts are examples of this second type. It is clear that within this framework, with their questionable status as soft constraints, motivational facts cannot affect the desirability of

³ Griffin is not the only philosopher endorsing this strong line of argument—see for example (Flanagan 1994)—but he is sufficiently representative.



the ideal—only the likelihood that it could be feasibly implemented (Gilabert and Lawford-Smith 2012, pp. 813–816).⁴

It is important to notice that these accounts are not merely interested in developing a particular conceptual account of feasibility, but are guided by moral considerations as to not excuse morally wrong actions. As Estlund writes, 'agents' abilities and inabilities to muster their will are subject to moral evaluation in their own right' (Estlund 2011, p. 207). Thus, for example, if we accept that people are naturally partial, and as such 'requirements to be otherwise are specious and false', we must—implausibly—accept the structurally similar argument when faced with the motivational fact that people tend to display some degree of cruelty (Estlund 2011, p. 224). Similarly, Lawford-Smith writes that 'we don't want to let agents off the moral hook. The fact that a person won't do something isn't enough for us to retract an imperative that she ought to' (2013, p. 254f.; cf. Wiens 2016, p. 336).

These accounts attracted many critics. David Wiens argues that Estlund's definition of ability already assumes motivational capacity and is insensitive to the distinction between mere lack of motivation—which would not be requirement-blocking—and motivational failures in good faith (Wiens 2016). Brian Carey argues that if there are, in fact, unalterable facts about motivation, we should accept them as hard constraints, although there may be epistemic constraints on evaluation that would lead us to adopt a rule against this accommodation (Carey 2016). Moving away from the conditional account of ability, Zofia Stemplowska argues that it overcounts actions as feasible. Instead, she proposes an incentive account of ability, according to which an action is feasible for an agent if there is an incentive that, when presented, would make the agent likely to act (Stemplowska 2021).

Despite their disagreements, these critics share a focus on the agent's ability to motivate herself. Even Stemplowska's account, which brings forward the role of incentives, is primarily interested in distinguishing mere unwillingness and motivational inability. This is a problem, in my view, as even in the narrower accounts of motivational ability, the number of cases in which the agent is truly unable to be motivated to act will be close to zero: only pathological or extreme cases would meet the requirements. Once the agent passes this low threshold they are held to be able to 'bring themselves to act'.

One may argue here that considering whether one can or cannot bring oneself to act does not simply depend on evaluating the probability of such motivation, but should contain a moral element as well. In David Miller's example, a normative theory that would require the United Kingdom to resemble North Korea should be rejected not only because the end-state would be undesirable, or because it is unlikely to happen; it should also be rejected because the people on whom this new

⁶ For similar critiques, see (Elford 2015; Southwood 2016).



⁴ It is not entirely clear where Gilabert and Lawford-Smith place motivational facts on the spectrum from Hard to Soft. See (Gilabert and Lawford-Smith 2012, p. 824, footnote 4), and, for skepticism over whether motivational facts could even count as soft constraints in the self-oriented view, (Lawford-Smith 2012, pp. 661–662).

⁵ It should be said that some accounts of feasibility, which I do not discuss here, remain focused on the conceptual definition and avoid the link to justification (e.g., Hamlin 2017).

order is imposed are likely to find it morally unacceptable (Miller 2013, pp. 36–37). This, indeed, seems to be the motivating thought in Griffin's argument as well: while he recognizes that the absolute sense of 'cannot' is unavailable, Griffin nevertheless believes that there is a sense of 'cannot' which is available and is conditional on certain ethical restrictions on the production of motivation, for example excluding motivation by autonomy-eroding fear (Griffin 1998, pp. 88–90).

As persuasive as this line of argument may be, it cannot be sufficient. Consider Gilabert and Lawford-Smith's distinction between 'actual' moral costs and 'believed' moral costs of transition (2012, p. 817). Actual moral costs of the ideal are relevant to its desirability. They may, all things considered, be outweighed; but they maintain their *pro tanto* force. On the other hand, believed moral costs, or 'positive morality', affect only considerations of feasibility; they make a certain normative theory more difficult to implement, as it is less likely that people would endorse it. The response to believed moral costs, they argue, should be as with all other soft constraints: strategic, not concessive. After all, positive morality is prone to error, subjected to different biases, and changes over time; it should not, therefore, be treated as more than a soft constraint. Motivational gaps, then, constitute in this view no more than believed moral costs, unless supported by a further moral argument.

Estlund, similarly, maintains that '[w]hat people can bring themselves to do is, often, relative to an environment of threats and incentives', and that these motivational incapacities 'can be overcome by structuring those threats and incentives' (Estlund 2019, p. 130). Clearly, not all forms of inducement would be permitted. Yet, he argues, it is wrong to assume that 'the fact that the oppressive threat or inducement should be forgone somehow erase the injustice of the behavior that will result in that case' (Estlund 2019, p. 130). More strongly, as I have already shown, Estlund argues that even deeply entrenched features of human nature cannot let people off the moral hook, unless accompanied with an additional moral argument (Estlund 2019, p. 136). Even if this conclusion is too strong, its driving force is compelling: it seems perverse to bend morality to brute facts about motivation. A mere lack of will does not relieve one of their moral duties.

To address this issue, some authors have turned to an alternative way in which motivational facts can be taken into account: focusing not on the ability of the agent to bring herself to act, but on the behavior—specifically, the non-compliance—of other agents. Lawford-Smith argues that 'the right way to deal with the motivation question... is to say that the motivation of other people is part of the context in which an agent acts, and therefore properly a soft constraint on whether her action will succeed' (Lawford-Smith 2013, p. 256). David Enoch argues that 'once there is more than one relevant agent, whether or not one agent will (or is likely to) act as it ought may be very relevant indeed to what another agent ought to do' (Enoch 2018). The expected behavior of third parties, regardless of whether it is morally correct or not, is relevant as a 'mere circumstance' to the obligations of the agent in question.

This is an important step in the right direction, as it rightly moves from the focus on the obligations of the individual agent to the question of multiple agents and the interaction between them. Moreover, I agree with Enoch that this formulation of the question is more useful to *political* philosophy. As he writes: 'In political



philosophy, [unlike moral philosophy], the multiplicity of agents is a crucial part of the problem. Political philosophy is *essentially* about multiple agents' (Enoch 2018, p. 4). The background conditions view nicely captures the insight that political agency is, in Sharon Krause's terms, non-sovereign and socially distributed, and deeply dependent on the actions of others (Krause 2015). However, perceiving the actions (and inactions) of others as static background conditions is still an incomplete account of the place of motivational facts in normative theory. Beyond the fact that the actions of others, and specifically, their failure to act, alter the content of the agent's moral duties and set of feasible actions, it is also relevant for their ability to change others' behavior. This is the question of political power, to which I turn next.

Reframing the Question: The Limits of Political Power

My argument is that the self-oriented and the background conditions accounts, despite their valuable insights, obscure a central feature of political action. Politics is not primarily about individuals motivating themselves to action, nor is it merely about moral agency and the content of duties constrained by the actions or inactions of others. In my proposed conception of politics, it primarily involves actors who are attempting to alter the behavior of others. Political theory, in its action-guiding role, aims to provide political actors with ends and values to orient their actions. Political actors include, most evidently, those in positions of power, but our understanding of politics should not be so restricted: activists, ordinary citizens, and 'everyday' political action all exemplify attempts to change others' behavior. Doing so, these political actors use different mechanisms of social change—rational persuasion, negative and positive incentives, compelling rhetoric, threats, brute violence, to name a few—to get other actors to behave as they would not have otherwise. Evaluating these mechanisms requires knowledge of people's psychology and their tendency to behave in certain ways. In other words, the question of motivation in political theory concerns the use of political power.

In a recent contribution, Naima Chahboun advances a related argument in her conceptualization of the 'institutional constraint' on the concept of justice. As she puts it, the institutional constraint holds that 'if a proposed principle demands that which we, even under ideal circumstances, are unable to make others do (through the creation of coercive institutions), it must be rejected as a principle of justice' (Chahboun 2017, p. 438). My proposed account differs in two ways. Unlike the binary nature of Chahboun's institutional constraint, I emphasize the comparative efficiency of different social mechanisms (which I take to be a broader concept than 'institutions'). More importantly, I want to maintain that the evaluation of political power has a related, and necessary normative component, which relies on a background theory of legitimacy. Let me unpack these two ideas in turn.

Some mechanisms of social change are better than others. First of all, some are better in terms of their *comparative efficiency*: they may be more reliable at

⁷ Chahboun claims that this kind of constraint is endorsed primarily by realists (Chahboun 2017, p. 437).



delivering results, or their implementation is more cost-effective, or they may be more robust against abuse and error. For example, Emily McTernan (2014) has argued that given the findings of situationist social psychology, a reliance on education for civic virtue as a means to secure certain behaviors should be replaced with the creation of social norms. This view is shared even among those who reject the status of motivational facts as normative constraints. As Gilabert and Lawford-Smith write, 'the fact that people do not want to do something does not mean that we should think getting it done is infeasible, it just means we should think about how to change incentive structures and thereby change people's desires' (Gilabert and Lawford-Smith 2012, p. 813; cf. Estlund 2019, pp. 130–131). Facts about human motivation are clearly important for evaluating the comparative efficiency of different social mechanisms, which is a straightforwardly empirical question.

Some mechanisms of social change may be better than others in a second sense: they may differ in terms of their permissibility, or, as I will refer to this criterion hereafter, in terms of their *legitimacy*. My use of 'legitimacy' here is restricted to the means-legitimacy, and is therefore distinct from the broader question of what makes political authority legitimate: a legitimate ruler may act in ways that are, in themselves, illegitimate. But of course, the two senses of the concept are related: what makes an action illegitimate is that it undermines, or is in contradiction with, the grounds of legitimate authority. To render this concept more concrete, consider the case of political corruption. One individual act of corruption—taking bribes, for example—is means-illegitimate, but does not necessarily undermine rule-legitimacy. Systemic corruption, on the other hand, undermines legitimate authority—it demonstrates a contradiction between action and normative foundations (cf. Philp and Dávid-Barrett 2015).

Two important clarifications are in order. First, this sense of means-legitimacy holds independently of how one conceives of legitimacy as a normative principle. This allows me, arguably, to remain agnostic on whether legitimacy is a moral principle applied to politics, or a principle derived from a distinct political normativity, as some political realists may argue (Cross 2021; Fossen 2021). Second, insofar as the distinction between sociological and normative legitimacy is productive (Horton 2012), my discussion here focuses on the latter; that is, on what a particular theory of legitimacy prescribes, not on what people happen to believe is legitimate.

In either case, considerations of legitimacy are of course not entirely distinct from those of comparative efficiency; one does not need to go as far as John Dewey in suggesting that comparative efficiency is 'the only question which can be raised about the justification of force' (1916, p. 367), to see that the use of excessive force in the pursuit of a normative end is unjustified. Torture may be an effective way to secure compliance, but it is not a legitimate way to motivate people to vote, since there are less violent means to effectively achieve this goal. And while assessments of legitimacy have a comparative element, most conceptions of legitimacy consider some social mechanisms to be beyond the pale: many, for example, would consider torture to be illegitimate even if necessary for the achievement of a narrow political end, because it is a practice inherently at odds with the underlying values of a legitimate rule.



This way of conceptualizing the legitimacy of social mechanisms differs, in my view, from viewing it through the lens of 'moral costs' (Buchanan 2003, pp. 60–61; Raikka 1998; cf. Southwood 2019). It has no *necessary* connection to the affected agents' moral beliefs about desirability, or to a second value such as autonomy, justice, or welfare. In this view, the standards according to which social mechanisms are evaluated are internal to the background theory of legitimacy itself. This theory of legitimacy *may* dictate that subjects' moral beliefs are taken into account, or that legitimate rule must be just or morally good in some other way. In those cases, the distinction I draw between my view and the self-oriented and background conditions views may be less clear, as considerations of legitimacy are reduced into other values. But importantly this is not generally the case for all theories of legitimacy.

To illustrate, consider two possible theories of legitimacy. Some liberal political theorists—characteristically committed to the value of personal autonomy and to the idea of consent as the source of legitimacy—consider some social mechanisms of influencing behavior, such as rational deliberation and economic incentives, as categorically preferable to rhetorical persuasion and violent coercion (e.g., Nozick 1974). As rational deliberation and incentives are not considered to be forms of power in this view, they are often bracketed away when evaluating the legitimacy of political action. Instead, as we have seen above, they are folded into the question of whether one is able to bring oneself to act. In comparison, realist theories of legitimacy identify the source of legitimacy in the ability to maintain order under conditions of disagreement (e.g., Burelli 2022). The legitimacy of means under this conception will ask whether the actions of state contribute or undermine this function.

This reinterpretation of the problem of motivation is preferable to the false dichotomy implicit in the methodological debate. It is simply false that we should either think that motivation is fixed and unchanging, or that, because it is malleable, it is only relevant for pragmatic reasons. History and experience teach us that people can be persuaded, inspired, manipulated, indoctrinated, brainwashed, or coerced, to do any number of things. We know of people who were willing to greatly endanger themselves and their families in order to protect complete strangers, as well as of people who turned their family in to the secret police out of ideological conviction. Some of these methods may be unstable, costly, unlikely to succeed, or indeed morally repugnant, but they nonetheless seem at least possible. In any case, given that historical examples of people acting in extremely impartial and altruistic ways exist, it is not clear how the claim of 'motivational impossibility' can be seriously sustained. The important question is not so much whether people's motivation is malleable—clearly, it is—but how is it malleable. A factually accurate understanding of human motivation is indispensable for attempting to answer this question: unless we understand the psychological mechanisms that influence people's behavior, we

⁸ As Ruth Grant (2011) argues, however, this assumption is often misleading: incentives and deliberation are also forms of power. Following Grant's typology, we may consider coercion, bargaining, and persuasion as different instantiations of power—using force, exchange, or speech, respectively—to control or influence others' behavior. Not only are the latter two forms of power, but they may be, in some circumstances, more dangerous than coercion, given their potential to be 'an exercise of power that conceals power' (Grant 2011, p. 134).



are unable to assess the legitimacy of the institutions and policies required for the implementation of any given normative proposal.

Facts about motivation are directly relevant, therefore, for assessing the legitimacy of mechanisms of social change. I propose, moreover, that they are indirectly relevant, via the question of legitimacy, to the justification of social and political ideals. The assessment of the legitimacy of the social mechanisms required for its implementation affects the justification of a political ideal in two ways: through considerations of accessibility and through considerations of stability.

First, accessibility. Suppose an otherwise attractive political ideal which could be feasibly implemented. For example, suppose that a conception of equality of opportunity is best realized in a system in which property is collectively owned and private property is abolished, instead of one in which individuals have exclusive ownership over property. There is no reason, after all, to think that our current system of private property is necessarily ideal from the perspective of equality of opportunity. Collective ownership and the abolition of private property might arguably fare better in light of major moral theories. Perhaps this arrangement would bring about the best state of affairs, or will be most respectful of individuals' autonomy, or will most successfully contribute to the cultivation of moral virtue. There is also little reason to think that human nature, in itself, somehow precludes this possibility: there are sufficient examples in history of collective and communal ownership of property, such as the medieval Icelandic commonwealth or the Israeli kibbutzim, to suggest that it is plausible at least under some social circumstances. It is also quite plausible, however, to suggest that the transition from our current system to this one, given human psychology, may require the use of other social mechanisms beyond deliberation and incentives. Whether these social mechanisms would be legitimate or not will be highly dependent on the background theory of legitimacy in play—but assuming that they will not be, the collective ownership of property and the abolition of private property will not be justified as a political ideal (cf. Hall 2012, pp. 176–178; Miller 2013, pp. 32–33).

I am happy to concede Gilabert and Lawford-Smith's point that one should not conflate believed and actual costs of transition, and that we should not completely surrender normative justification to accessibility constraints, as these often reflect unjustified biases. Indeed, Gilabert seems to me correct in arguing that limits on accessibility trigger what he calls 'dynamic' duties, leading political actors to act in ways that will make the policy more accessible in the future (Gilabert 2017). According to my interpretation here, however, making a theory more accessible does not simply refer to the probability of its implementation, but to the legitimacy of mechanisms required to bring it about; for example, making it accessible through education and not by imposition. In other words, the justification of 'dynamic' duties is similarly subjected to considerations of legitimacy.

One might object further that given the malleability of motivation, this concession leads to my view collapsing into the self-oriented view I criticized above. For example, even if the means to implement collective ownership of property are illegitimate at present, because they involve excessive coercion, we can imagine



a gradual process over many years that will, through legitimate means, transform people's psychology to be more hospitable to this ideal, thus alleviating the need to resort to illegitimate means. It is important to note, however, that while we may plausibly make predictions about the transformative effect of some social mechanisms in the short-term future, the uncertainty of future transformations makes these evaluations increasingly difficult in the long term. Assuming that achieving the ideal in question would require illegitimate means in the present, the more distant the possibility of achieving it legitimately in the future, the more the burden of proof on the defenders of the ideal will become heavier.

Motivational facts are relevant to the question of justification in a second way, as they affect stability. Whether or not a normative proposal is accessible from where we are now, it cannot be fully justified unless it can be shown that, once implemented, it can be sustained over time. Motivational facts play an indispensable part in this, as the stability of any normative program for political institutions cannot be demonstrated without reference to the projected behavior of agents under these institutions. But again, this is not merely a technical question: if the stability of a political ideal cannot be achieved without recourse to illegitimate means, then it cannot be justified as a political end.

Consider, for example, Rawls's argument about the stability problem and the possible solutions to it (Rawls 1993; cf. Weithman 2010). For Rawls, the stability problem arises given certain psychological facts about human beings, specifically the conditional nature of their willingness to cooperate under conditions of uncertainty. Even if each participant in a social scheme would benefit from mutual cooperation, each will tend to defect if all others defect, rather than endure the costs of picking up the slack when others free-ride. Simply assuming cooperation will result in instability, unless all can be assured that others will co-operate. The stability problem is obviously central to questions of non-ideal theory, understood as dealing with problems of 'non-compliance' (Simmons 2010). Importantly, however, for Rawls it was relevant for the level of justification in ideal theory as well. Indeed, this is one of his main reasons for rejecting the utilitarian conception of justice: unless 'sympathy and benevolence can be widely and intensely cultivated', this conception is threatened with instability and will not be acceptable to the parties of the original position, and could only be maintained through unacceptable imposition (Rawls 1971, pp. 178/155).

We should not, of course, be wedded to either Rawls's particular depiction of the problem of stability or the specific solutions he gives to it. One may be skeptical with regard to the account of moral psychology with which Rawls is working, or question the particular conception of political legitimacy which informs his understanding of the stability problem. Rawls is committed to thinking that all solutions to the stability problem must be those that provide stability 'for the right reasons', or, in other words, show that a liberal, well-ordered society could provide its own internal motivation, without relying on either coercion or ideology. This is not an uncontroversial move, as some recent critics of Rawls argue (e.g., Sleat 2013). But

⁹ I would like to thank an anonymous reviewer for this objection.



for my purposes this debate could be safely side-stepped; whatever normative conception of legitimate power one holds, facts about human motivation would inform the conditions for political and institutional stability, and whether it could be legitimately attained.

My account is therefore fully compatible with the simple fact that motivational facts are both contingent and malleable, but it does not consider this to be a conversation-stopper about their role in normative justification. As I pointed out, the missing normative element from the debate over feasibility constraints is found in the question of the *means* of transforming behavior through social mechanisms. For considerations of accessibility, even if certain ways of behavior are contingent, it matters that they exist: otherwise we cannot know what will be required for the transformation of society from its current state toward the desired ideal. For considerations of stability, even if certain motivational facts are malleable, it matters *how* they are malleable; there is a difference between achieving stability through persuasion, incentives, coercion, or manipulation.

Moreover, by focusing on the legitimacy of mechanisms of social change rather than on what is required of individual agents, my account is not susceptible to the objection that one can get oneself off the moral hook by becoming less motivated. We may coherently think that individuals have stringent and demanding moral and ethical obligations that they are unwilling to perform, while still maintaining that imposing the fulfillment of these duties would not be justified; that is, that the agent is morally required to fulfill certain duties, but that there is no legitimate way to compel her to fulfill these duties. For example, given certain conceptions of legitimacy that rule out intrusive coercion as a social mechanism for behavioral change, we may accept that adultery is immoral, but that the state, or any other political actor, cannot legitimately act to abolish it. The abolition of adultery cannot be, therefore, part of a justified political ideal under this conception of legitimacy. This boundary between the moral and the political realms is of course a fuzzy one, highly dependent on context, and indeed, in itself a contested political question: for example, whether familial relations can be susceptible to political intervention is one of the key points of friction between feminists and their opponents. Yet even given these points, it would be a mistake to simply conflate the political domain and the moral domain, even if we wish to resist the realist view that political normativity is autonomous. 10

It should be clear by now that my account is not restricted to any particular substantive picture of legitimacy. As such, it is fully compatible with the objection to viewing motivational facts as normative constraints *simpliciter*, independently of normative considerations; as I showed above, motivational facts are relevant to normative justification indirectly through an assessment of the legitimacy of different mechanisms of social change. While I suggested some parameters for evaluating legitimacy, these should be taken as offering a framework rather determining the content of these parameters: my account is compatible with more and less permissive

¹⁰ See the recent discussion of this question in (Erman and Möller 2015a, 2021; Jubb and Rossi 2015; Leader Maynard and Worsnip 2018; Sleat 2021).



substantive understandings of how to evaluate the legitimacy of transforming motivation, and what means of social change should be considered absolutely illegitimate, not just comparatively less legitimate. Some, for example, will reject all use of physical violence as illegitimate; and others will maintain that means should only be evaluated instrumentally in light of the ends. Both are compatible with my suggested framework. All I am proposing is that by explicitly acknowledging the importance of a background theory of legitimacy, we can and should deny the implicit assumption in many existing accounts that our focus should be on whether the agent is able to motivate herself into action.

Even if my argument so far is persuasive with regard to consideration of legitimacy, one may object that it is begging the question against views such as Estlund's, that wish to exclude facts about motivation from the justification of political ideals. Perhaps, if what we are interested in is what justice requires, then considerations of legitimacy (and thus, of accessibility and stability), however valuable, are irrelevant (Estlund 2019). If the only way to achieve full justice is through illegitimate means, and so justice can only be brought about at the cost of legitimacy, that is regrettable and perhaps a reflection of lamentable moral tragedy (cf. Miller 2013, pp. 229–249). But this does not require a modification of the ideal itself.

I do not wish to deny that some forms of utopian political theory are valuable pursuits, even if I am skeptical about their importance given political theory's public, action-guiding role (Frazer 2016). Importantly, one implication of my view is that the justification of political ideals is contextual—a political ideal is not justified in the abstract, but justified to people like us, as we are now, or as we can expect to be. I am happy to concede that for some highly abstract, 'pure' forms of political theory, which do not aspire to be action-guiding, my argument would not be as persuasive (e.g., Cohen 2008). Insofar as most political theory aspires to be, in some way, action-guiding—if it aims to orient the actions of particular political actors rather than articulate the content of moral oughts in the abstract—there would still be a need to take motivational facts into account in the way I described.

A Conservative Bias?

Nevertheless, the account I defend above will undoubtedly strike some as being overly conservative. In some sense, this is unavoidable; after all, any account that justifies limitations on normative theorizing will lead to some otherwise attractive political ideals being ruled out. The interesting question here is not whether my account is conservative in this trivial sense, but whether it is *objectionably* conservative; that is, whether this supposedly methodological intervention is ideological in the pejorative sense (Geuss 1981). In this final section, I wish to suggest reasons to think that it is not. To do so, I will consider three potential objections which advance this line of criticism in different ways.

In her critique of political realism, Lorna Finlayson (2017) compares the idea of human nature as a normative constraint to 'victim blaming'. In the same way that a victim of rape is accused that her behavior is the cause of her abuse, realist theorists 'cite the deficiencies of human beings—the large majority of whom are being



(and more or less always have been) well and truly screwed by the political structures under which they live—as the root of the shortcomings of their political institutions, and hence of the suffering that those institutions cause' (Finlayson 2017, pp. 276–277). Even if certain political ideals were possible if people behaved differently, it is just as perverse to justify the status quo in this case as it is in the rape case.

The point is well taken, but I think that it does not apply to my proposed account. First, unlike other similar proposals, my account does not consider motivational facts as constraints independently of normative considerations regarding institutions, norms, and other loci of political power. Second, the legitimacy of mechanisms for social change is always considered vis-à-vis the people at which they are directed; and as such, restrictions of freedom could be legitimate to secure greater freedom. Finally, if we consider the problem of motivation as one which is related to questions of legitimacy in the use of political power, then the relevant questions—effectiveness, accessibility, and stability; force vs persuasion; collective action problems, etc.—arise not only for individuals as subjects of political power, but also (and perhaps primarily) for individuals in positions of political power. This insight, for example, has motivated thinking about the problem of political corruption and the robustness of institutions to abuse by self-interested actors, which could be progressive in nature (Buchanan and Tullock 1962; Philp and Dávid-Barrett 2015). 11

Criticizing motivational constraints from a different angle, Lisa Herzog (2015) warns that leveling down normative theories to fit 'human nature'—specifically, economic theories of humans as self-interested, utility maximizing beings—is not just misguided, but politically dangerous. As moral motivation depends, among others, on epistemic cues of others' cooperation, the spread of economic ideology, in which individuals understand themselves and others as self-interested, will generate feasibility gridlocks. Instead of promulgating restrictive theories of human nature and constraining normative theories to fit them, political theorists 'can help to shift the feasibility frontier for more egalitarian institutions, by making clear to individuals that others might also be willing to do their bit in the struggle for justice' (cf. Gilabert 2017; Herzog 2015, p. 968).

Does my proposed framework preclude such a progressive role for political theory? No, for three reasons. First, being sensitive to facts about motivation, as I suggest political theorists should be, includes being aware of feedback effects different policies have on shaping people's behavior and motivation (e.g., Frey 1997). Thus, when considering the legitimacy of different social mechanisms, the long-term and unintended effects are also being taken into account. Second, there is nothing in my account which suggests that *Homo Economicus* is the most accurate psychological model of human behavior; it may or may not be, but that is a question for empirical psychology. Third, even if we assumed that *Homo Economicus*, or any other

¹¹ It should be noted that Finlayson remains suspicious about the last point: 'When political failures are blamed on "human" ones, the implication—at least, the conversational implicature—is very often that it is the "human" failings of the people at large ("the masses") that are explanatorily salient, rather than those of ruling elites' (p. 277).



similarly restrictive model of human nature, is the right one, this should not suggest that human behavior is the same regardless of incentives, norms, or institutions. In other words, to borrow from the terminology of social choice theory, granting the methodological usefulness of motivational neutrality should not lead us to assume *behavioral* neutrality. So political theorists interested in expanding the feasibility frontier should be interested both in having an empirically accurate picture of human psychology, and in the different social mechanisms that influence behavior.

A final powerful objection is the revolutionary critique, vividly captured in Leon Trotsky's 1938 famous pamphlet *Their Morals and Ours* (Trotsky 1973). To paraphrase Trotsky's attack on 'moralist' critics of Bolshevik violence: it may be the case that the implementation of a radical normative theory in practice, given certain motivational facts about the relevant political agents, would require the use of morally objectionable means, while still being fully justified. As a clear example, the abolition of slavery in the United States required the use of violent force—indeed, an outright civil war—to impose a new institutional structure on former slaveholders; and the liberal legitimacy of these political institutions in the Reconstruction Era was at the very least disputed. Could these extreme measures mean that the abolition of slavery in the United States was somehow unjustified? Surely not. As Trotsky puts it, '[a] slave-owner who through cunning and violence shackles a slave in chains, and a slave who through cunning or violence breaks the chains—let not the contemptible eunuchs tell us that they are equals before a court of morality' (Trotsky 1973, p. 44).

Trotsky's argument could be understood in one of two ways. The first is to see it as an all-things-considered argument about moral costs. It may be that the benefits of the utopian end-state—or, more realistically, the radical injustices of the status quo—are so great that they outweigh or override the costs of achieving (or overcoming) it. It seems plausible that the injustice of institutional slavery was so great that these costs were, all-things-considered, necessary. This is only a critique of my framework insofar as we assume that facts about motivation serve to justify absolute deontic side-constraints. Now, under certain substantive theories of political legitimacy, that may be true; but this is in no way entailed by my methodological account. That said, unless one assumes that the value of utopia (or the disvalue of the status quo, for that matter) is infinite—that is, justifying any cost—it seems that at least conceptually some costs would not be worth paying. In other words, how can we know whether the ends justify the means, if we do not know what means are going to be required?

As I argued above, however, my account cannot be reduced to an argument about moral costs; evaluations of means-legitimacy are internal to the conception of legitimacy in play. It is possible, then, to interpret Trotsky's argument as claiming that political philosophers' beliefs about legitimacy are themselves a product of the unjust status quo. Thus, any supposed costs of transition—and the background theory of legitimacy which grounds them—would turn out

¹² On the principle of motivational neutrality, see (Buchanan and Tullock 1962, pp. 2–3). On the distinction between motivational and behavioral neutrality see (Brennan and Lomasky 1993, pp. 1–19).



to be deceptive, and should not be taken into consideration. The 'violation' of the slaveholder's property rights over his slave is no violation at all—it is not a cost that should be weighed against potential benefits. In this case, the choice between different social mechanisms of change would be solely on the basis of their expected success in abolishing the unjust status quo. As considerations of legitimacy would be indistinguishable from considerations of comparative efficiency, this would be a limiting case for my framework that may render it unnecessary. At best, it would mean that facts about motivation would be relevant for the justification of the use of certain means over others, for example preferring non-violence (Mantena 2012), but they would not play a role in the justification of political ends. Note that framed in this way, the case for bracketing away facts about motivation is much more radical than it originally appears: it is not merely a methodological argument about the role of facts in normative justification, but a substantive argument about the deep injustice of the status quo.

Concluding Remarks

Politics involves the use of power. My contention throughout this article has been that, at least with regard to the methodological debates on the role of motivational facts in political theory, this basic truth has been obscured. With the focus on moral motivation, and the capacity (or lack thereof) of the individual to motivate herself, the central question of *how* to motivate political action was neglected. Different social mechanisms to motivate political action are forms of power, differing in effectiveness and legitimacy. Motivational facts matter for the justification of political ideals, not because they are fixed points to which normative theory should be anchored, but because without them claims about comparative efficiency and legitimacy—and more specifically, about accessibility, stability, and transition costs—are meaningless.

As I hope to have shown in the final section, this line of argument is not as complacent or concessive as it may appear. As it is open to both narrow and broad interpretations of legitimacy, my framework is compatible with many substantive positions, from conservative incrementalism to radical transformation. The upshot of my account is a plea for honesty and consistency: that is, that in the justification of any political ideal, the costs and trade-offs of implementation should be justified as well.

Acknowledgements I would like to thank my PhD supervisors, Richard Bellamy and Cécile Laborde, as well as my examiners, David Miller and Avia Pasternak, for their comments and guidance on the doctoral research from which this article originates. A previous iteration of this paper was presented at the 'Facts and Norms' conference, University of Copenhagen, August 2013, where I benefited from the comments of David Estlund, Pablo Gilabert, and Anabelle Lever, as well as at the ECPR General Conference in Prague, September 2016. A special thanks also to Gabriele Badano, Brian Carey, Shai Gertler, Roni Hirsch, Robert Jubb, Steven Klein, Attila Mraz, Andrei Poama, Enzo Rossi, and Kai Spiekerman for their written comments at different stages. Funding for some of the research period in which the article was written was provided by the UCL Overseas Trust and the Anglo-Israel Association Kenneth Lindsay Scholarship.



Declarations

Conflict of interest None.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

- Brennan, G., and L. Lomasky. 1993. *Democracy and decision: The pure theory of electoral preference*. Cambridge: Cambridge University Press.
- Buchanan, A. 2003. *Justice, legitimacy, and self-determination: Moral foundations for international law.* Oxford: Oxford University Press.
- Buchanan, J. M., and G. Tullock. 1962. The calculus of consent: Logical foundations of constitutional democracy. Carmel, CA: Liberty Fund.
- Burelli, C. 2022. Political normativity and the functional autonomy of politics. *European Journal of Political Theory* 21 (4): 627–649. https://doi.org/10.1177/1474885120918500.
- Carey, B. 2016. Justice for jerks: Human nature, selfishness and non-compliance. *Social Theory and Practice* 42 (4): 748–766. https://doi.org/10.5840/soctheorpract201642426.
- Chahboun, N. 2017. Three feasibility constraints on the concept of justice. *Res Publica* 23 (4): 431–452. https://doi.org/10.1007/s11158-016-9336-z.
- Cohen, G. A. 2008. Rescuing justice and equality, 1st edn. Cambridge, MA: Harvard University Press. Cross, B. 2021. Normativity in realist legitimacy. Political Studies Review 19 (3): 450–463. https://doi.org/10.1177/1478929920917834.
- Dewey, J. 1916. Force and coercion. *International Journal of Ethics* 26 (3): 359–367. https://doi.org/10.2307/2377050.
- Elford, G. 2015. Pains of perseverance: Agent-centred prerogatives, burdens and the limits of human motivation. *Ethical Theory and Moral Practice* 18 (3): 501–514. https://doi.org/10.1007/s10677-014-9534-8.
- Enoch, D. 2018. Against utopianism: Noncompliance and multiple agents. *Philosopher's Imprint* 18(16). Available at: http://hdl.handle.net/2027/spo.3521354.0018.016 (accessed 3 January 2022).
- Erman, E., and N. Möller. 2015a. Political legitimacy in the real normative world: The priority of morality and the autonomy of the political. *British Journal of Political Science* 45 (1): 215–233. https://doi.org/10.1017/S0007123413000148.
- Erman, E., and N. Möller. 2015b. Practices and principles: On the methodological turn in political theory. *Philosophy Compass* 10 (8): 533–546. https://doi.org/10.1111/phc3.12245.
- Erman, E., and N. Möller. 2021. Distinctively political normativity in political realism: Unattractive or redundant. *Ethical Theory and Moral Practice*. https://doi.org/10.1007/s10677-021-10182-8.
- Estlund, D. 2008. Democratic authority: A philosophical framework. Princeton, NJ: Princeton University
- Estlund, D. 2011. Human nature and the limits (if any) of political philosophy. *Philosophy & Public Affairs* 39 (3): 207–237. https://doi.org/10.1111/j.1088-4963.2011.01207.x.
- Estlund, D. 2019. Utopophobia: On the limits (if any) of political philosophy. *Princeton University Press*. https://doi.org/10.1515/9780691197500.
- Finlayson, L. 2017. With radicals like these, who needs conservatives? Doom, gloom, and realism in political theory. *European Journal of Political Theory* 16 (3): 264–282. https://doi.org/10.1177/1474885114568815.



- Flanagan, O. 1994. Varieties of moral personality: Ethics and psychological realism. Cambridge, MA: Harvard University Press.
- Fossen, T. 2021. Political legitimacy as a problem of judgment: What distinguishes moralist, realist, and pragmatist approaches? *Social Theory and Practice*. https://doi.org/10.5840/soctheorpract2021121 4148.
- Frazer, M. L. 2016. Utopophobia as a vocation: The professional ethics of ideal and nonideal political theory. *Social Philosophy and Policy* 33 (1–2): 175–192. https://doi.org/10.1017/S02650525160003 27.
- Frey, B. S. 1997. A constitution for knaves crowds out civic virtues. *The Economic Journal* 107 (443): 1043–1053. https://doi.org/10.1111/j.1468-0297.1997.tb00006.x.
- Geuss, R. 1981. The idea of a critical theory: Habermas and the Frankfurt School. Cambridge: Cambridge University Press.
- Gilabert, P. 2012. Comparative assessments of justice, political feasibility, and ideal theory. *Ethical Theory and Moral Practice* 15 (1): 39–56. https://doi.org/10.1007/s10677-011-9279-6.
- Gilabert, P. 2017. Justice and feasibility: A dynamic approach. In *Political utopias: Contemporary debates*, ed. K. Vallier and M. Weber, 95–127. Oxford: Oxford University Press.
- Gilabert, P., and H. Lawford-Smith. 2012. Political feasibility: A conceptual exploration. *Political Studies* 60 (4): 809–825. https://doi.org/10.1111/j.1467-9248.2011.00936.x.
- Grant, R. W. 2011. Strings attached: Untangling the ethics of incentives. Princeton, NJ: Princeton University Press.
- Griffin, J. 1990. Review of 'The limits of morality'. Mind. https://doi.org/10.1093/mind/XCIX.393.128.
- Griffin, J. 1998. Value judgement: Improving our ethical beliefs. Oxford: Oxford University Press.
- Griffin, J. 2015. What can philosophy contribute to ethics? Oxford: Oxford University Press.
- Hall, E. 2012. Political realism and fact-sensitivity. *Res Publica* 19 (2): 173–181. https://doi.org/10.1007/s11158-012-9199-x.
- Hamlin, A. 2017. Feasibility four ways. Social Philosophy and Policy 34 (1): 209–231. https://doi.org/10. 1017/S0265052517000103.
- Hamlin, A., and Z. Stemplowska. 2012. Theory, ideal theory and the theory of ideals. *Political Studies Review* 10 (1): 48–62. https://doi.org/10.1111/j.1478-9302.2011.00244.x.
- Herzog, L. 2015. Distributive justice, feasibility gridlocks, and the harmfulness of economic ideology. *Ethical Theory and Moral Practice* 18 (5): 957–969. https://doi.org/10.1007/s10677-015-9565-9.
- Horton, J. 2012. Political legitimacy, justice and consent. *Critical Review of International Social and Political Philosophy* 15 (2): 129–148. https://doi.org/10.1080/13698230.2012.651015.
- Jubb, R., and E. Rossi. 2015. Political norms and moral values. Journal of Philosophical Research 40: 455–458. https://doi.org/10.5840/jpr201511539.
- Krause, S. R. 2015. Freedom beyond sovereignty: Reconstructing liberal individualism. Chicago, IL: University of Chicago Press.
- Lawford-Smith, H. 2012. The motivation question: Arguments from justice and from humanity. *British Journal of Political Science* 42 (3): 661–678. https://doi.org/10.1017/S0007123412000063.
- Lawford-Smith, H. 2013. Understanding political feasibility. *Journal of Political Philosophy* 21 (3): 243–259. https://doi.org/10.1111/j.1467-9760.2012.00422.x.
- Leader Maynard, J., and A. Worsnip. 2018. Is there a distinctively political normativity? *Ethics* 128 (4): 756–787. https://doi.org/10.1086/697449.
- Mantena, K. 2012. Another realism: The politics of Gandhian nonviolence. *American Political Science Review* 106 (2): 455–470. https://doi.org/10.1017/S000305541200010X.
- McTernan, E. 2014. How to make citizens behave: Social psychology, liberal virtues, and social norms. *Journal of Political Philosophy* 22 (1): 84–104. https://doi.org/10.1111/jopp.12015.
- Miller, D. 2013. *Justice for earthlings: Essays in political philosophy*. Cambridge: Cambridge University Press.
- Nozick, R. 1974. Anarchy, state, and utopia. New York: Basic Books.
- Philp, M., and E. Dávid-Barrett. 2015. Realism about political corruption. *Annual Review of Political Science* 18 (1): 387–402. https://doi.org/10.1146/annurev-polisci-092012-134421.
- Raikka, J. 1998. The feasibility condition in political theory. *Journal of Political Philosophy* 6 (1): 27–40. https://doi.org/10.1111/1467-9760.00044.
- Rawls, J. 1971. A theory of justice. Cambridge, MA: Harvard University Press.
- Rawls, J. 1993. Political liberalism. Columbia University Press.
- Rousseau, J.-J. 1962. Rousseau: The social contract and other later political writings. Cambridge: Cambridge University Press.



Simmons, A. J. 2010. Ideal and nonideal theory. Philosophy & Public Affairs 38 (1): 5–36. https://doi.org/10.1111/j.1088-4963.2009.01172.x.

- Sleat, M. 2013. Coercing non-liberal persons: Considerations on a more realistic liberalism. *European Journal of Political Theory*. https://doi.org/10.1177/1474885112471296.
- Sleat, M. 2016. Realism, liberalism and non-ideal theory or, are there two ways to do realistic political theory? *Political Studies* 64 (1): 27–41. https://doi.org/10.1111/1467-9248.12152.
- Sleat, M. 2021. Realism and political normativity. Ethical Theory and Moral Practice. https://doi.org/10. 1007/s10677-021-10239-8.
- Southwood, N. 2016. Does 'ought' imply 'feasible'? *Philosophy & Public Affairs* 44 (1): 7–45. https://doi.org/10.1111/papa.12067.
- Southwood, N. 2019. Feasibility as a constraint on 'ought all-things-considered', but not on 'ought as a matter of justice'? *The Philosophical Quarterly* 69 (276): 598–616. https://doi.org/10.1093/pq/pqz012.
- Stemplowska, Z. 2021. The incentives account of feasibility. *Philosophical Studies* 178 (7): 2385–2401. https://doi.org/10.1007/s11098-020-01530-y.
- Trotsky, L. 1973. Their morals and ours, 5th edn. New York: Pathfinder Press.
- Valentini, L. 2012. Ideal vs. non-ideal theory: A conceptual map. *Philosophy Compass* 7 (9): 654–664. https://doi.org/10.1111/j.1747-9991.2012.00500.x.
- Weithman, P. 2010. Why political liberalism?: On John Rawls's political turn. Oxford: Oxford University Press.
- Wiens, D. 2015. Political ideals and the feasibility frontier. Economics & Philosophy 31 (3): 447–477. https://doi.org/10.1017/S0266267115000164.
- Wiens, D. 2016. Motivational limitations on the demands of justice. *European Journal of Political Theory* 15 (3): 333–352. https://doi.org/10.1177/1474885115578446.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

