UNDERSTANDING NORMS AND CHANGING THEM

By Ryan Muldoon

Abstract: *It is crucial for policymakers to focus their attention on social norms if they want to improve policy outcomes, but doing so brings in new normative questions about the appropriate role of the state. Indeed, I argue that efforts to reduce coercion at the state level can create potentially pernicious and difficult to eliminate forms of coercion at the informal level. This creates a new normative challenge for thinking about the broader regulatory apparatus, and complicates our approach in utilizing social norms for democratic policy ends. I will distinguish between two forms of social norms orientations in policy: a diagnostic approach and a design approach. We will see that the diagnostic approach better models a Humean approach to supporting social norms, and a design approach has a more Millian character. While it is easier to justify a design approach in the abstract, as it has very little room for abuses of state authority, if Mill is right that social norms can be a source of coercive power that runs afoul of the harm principle, then a design approach will sometimes be necessary to counter this form of tyranny. However, this latter approach is complex, and as such we may want to take a recommendation from Mockus to focus on deliberative approaches to norm change.*

KEY WORDS: Social norm, policy, nudge, harm, coercion

I. Introduction

Let us assume that governments have a mandate to operate a regulatory apparatus in the interests of citizens. There remains a question of how to best carry out such a regulatory regime. Traditionally, this has been done through the main tools of policy: information, legal sanction, financial incentives, and government monitoring. The government can give its citizens better information to help them inform their choices, it can put into place laws and regulations that have some coercive power, it can make adjustments to tax law to encourage or discourage certain behaviors, and it can use monitoring to enforce its regulatory apparatus. All of this, in one way or another, relies on a model of human decision-making: one in which we are all *homo economicus*: ideally rational, able to incorporate new information and to optimally determine what we should do to best satisfy our preferences.

Thanks to a growing literature across the social sciences that has developed a much more nuanced understanding of human decision-making, policymakers are increasingly turning their attention to behaviorally informed policies. Books such as *Nudge* and *Predictably Irrational* have advanced the debate, primarily by focusing on detailing systematic cognitive limitations for individual decision makers, and potential policy

structures that can help to achieve better policy outcomes by taking these limitations into account. There is also an opportunity to use behavioral nudges to formulate policies that at least in theory *reduce* the amount of state coercion used in policy formation. So, policy interest in behavioral science stems both from an empirical interest (more reliably improving outcomes) and a normative interest (reducing the amount of state coercion used to do so). Cass Sunstein and Richard Thaler in particular have argued that a reliance on behavioral interventions creates the opportunity for what they call "libertarian paternalism"—government can help guide people's choices to better ones without foreclosing options.[1]

Core to this idea is that *homo economicus* is a poor, or at least, overly-idealized, model of humans. We are far more limited beings than is imagined in classical economic modeling. And so we do not simply change our behavior optimally in response to new information, or even a new regulatory environment. We find that people don't save enough for their retirement, that the poor often don't take advantage of beneficial programs like financial aid for college, and people are in general bad at managing their health. And so on. Improvements in choice architecture are meant to encourage better decision-making, while continuing to grant individuals the right to choose for themselves.

By focusing on behavioral patterns of individual decision makers, policy-by-nudge has an opportunity to grasp some low-hanging fruit: increasing retirement savings, the organ donation rate, and inducing healthier eating in cafeterias are real and significant policy victories. However, this fairly narrow focus may prove to be limiting. What's more, it opens itself up to the charge that so-called libertarian paternalism is mostly just paternalism. Since many nudges are subtle (changing sizes of plates, changing the location of the salad bar, changing the default choice, and so on), it is not clear that individuals consciously register the choice they are making. These changes may not be something that are the product of rational reflection and deliberation, but instead a form of (minor) manipulation. Thus, behavioral nudges are constrained in two ways: first, their domain of application may be limited; but second, and perhaps more importantly, their normative status is at least open to dispute. What makes nudges work—that they alter the choice environment to encourage a specific choice—may undermine the libertarian claim of keeping one's full option set available.

A complementary approach to individualistic behavioral policymaking leverages social norms to promote behavior change. A social norms approach treats people not as atomistic individuals, but instead as members of larger communities. Instead of thinking of behavior as a series of one-time decisions, a social norms approach focuses on shared rules of behavior.

---

[1] Richard Thaler and Cass Sunstein, *Nudge: Improving Decisions about Health, Wealth and Happiness* (New York: Penguin Books, 2009)

This approach aims to align a community's social expectations of its members with the policy goals that have been democratically determined. This approach promotes greater levels of community engagement, and has the potential to avoid a common critique of nudging policy: that nudges circumvent our rational decision-making process and instead substitute citizen choice with technocratic choice. This other realm of policy brings along its own share of interesting normative challenges that is often elided in both policy discussion and in political philosophy: How should we think about the role of the state, and in particular the state's coercive power, in relation to the social norms that can shape much of our private-sphere behavior. I consider two possibilities: first, that social norms can be used in a way that devolves power to these local informal institutions, consonant with a more Humean approach to social arrangements that defers to whatever conventions a given society has in place. In this approach, the state serves to facilitate local social norms as a means of regulating public morality. In the second approach, I consider a more Millian view, where social norms can be a source of tyranny, and argue that coercive state power can be used to increase individual freedoms.

In this essay, I will argue that it is crucial for policymakers to focus their attention on social norms if they want to improve policy outcomes, but doing so brings in new normative questions about the appropriate role of the state. Indeed, I argue that efforts to reduce coercion at the state level can create potentially pernicious and difficult to eliminate forms of coercion at the informal level. This creates a new normative challenge for thinking about the broader regulatory apparatus, and complicates our approach in utilizing social norms for democratic policy ends.

I will distinguish between two forms of social norms orientations in policy: a diagnostic approach and a design approach. We will see that the diagnostic approach better models a Humean approach to supporting social norms, and a design approach has a more Millian character. While it is easier to justify a design approach in the abstract, as it has very little room for abuses of state authority, if Mill is right that social norms can be a source of coercive power that runs afoul of the harm principle, then a design approach will sometimes be necessary to counter this form of tyranny. I follow Lessig in suggesting that the Civil Rights Act is one such example of using state power to dismantle tyrannical social norms.[2]

II. Defining Social Norms

Social norms are widely studied across several academic disciplines: sociology, social psychology, economics, and philosophy all promote different

---

[2] Lawrence Lessig, "The Regulation of Social Meaning," University of Chicago Law Review 62 (1995): 943.

accounts of what a social norm is.[3] While these competing definitions have been a boon to academic research, in that they have fostered many different kinds of research programs that have investigated different aspects of how we take normative cues from our social environments, it has made it more difficult for policymakers to productively engage with norms. Social norms can appear to be a very vague category, becoming more akin to a general idea of cultural values, rather than an operationalized concept that can be productively measured and used to make predictions about behavior. For this reason, I will focus on a single account of social norms that is both amenable to measurement and has some predictive power. This makes it more suitable as a framework for policy-driven norm change.

Cristina Bicchieri offers a formal definition of norms that can be deployed in a policy context.[4] On this account, a social norm is a behavioral rule R in a population P that applies in a context C. An individual within P has a conditional preference to follow R in C if and only if two kinds of expectations are met. First, her *empirical expectations*: she expects that a sufficient subset of P complies with R in C. Second, her *normative expectations*: she expects that a sufficient subset of P believes that she should follow R in C. Failure to comply with R may result in sanctions from others in P. Complying may bring with it positive rewards. Bicchieri defines a *descriptive norm* in the same way, except that there are no normative expectations.

There are a few important things to notice about this definition of norms. First, the definition is built out of individuals' beliefs—both types of norms involve expectations that are beliefs about others. Empirical expectations are beliefs about what others do (and will do in the future), whereas normative expectations are beliefs about what others believe. This suggests that there can be heterogeneity in understanding certain social situations. For example, we might imagine three college students drinking at a party. Alice is drinking because she sees everyone else drinking, and thinks that she is also expected to drink. For her, there is a social norm present, and she would rather conform to the norm, even if she would rather not drink. Bob is drinking because he sees everyone else drinking, but he doesn't think anyone cares what he does. Still, he would rather do what others do, so he drinks. Finally, Carol doesn't particularly care about what others do, but enjoys drinking, and so she drinks. Even though all three follow

the same behavior, Alice is following a social norm, for fear of social sanction, whereas Bob drinks because he'd rather follow the crowd, and Carol is just following her personal preferences. A public policy intervention that offered an informational campaign on the dangers of drinking might sway Carol, but it is unlikely to affect the behavior of Alice and Bob, since they aren't motivated by the pleasure of drinking as much as the desire to fit in, and (in Alice's case) the fear of sanction if they do not. In fact, several successful campaigns to reduce binge drinking on American college campuses have involved not just health information, but social information: how many beers the average student would like to drink in an evening. Prior to intervention, students thought others wanted to drink around five drinks in an evening, when in fact everyone wanted about two. Armed with this information, empirical and normative expectations can (and did) shift downward.[5]

As this example illustrates, motivations for action matter: if people are more like Alice than Carol, they need different interventions to respond to their reasons for action. In the real cases of binge drinking on many college campuses, not only did students drink because of their beliefs about what others expected of them, the students were all mistaken about how much everyone else wanted to drink. This is known as *pluralistic ignorance*— when everyone's public expressions are a belief counter to their private belief, because they mistakenly believe everyone else endorses the public sentiment. Fear of sanction prevents people from exposing themselves as someone who would rather do something else. Without intervention, this pluralistic ignorance is stable, as everyone's beliefs are in equilibrium: they are all motivated to go along with everyone else, even if privately everyone would rather do something else.[6] While uncommon, pluralistic ignorance can result in the continuation of practices that no one likes or would privately endorse, simply because it is too socially risky to announce that they do not like the norm. The binge drinking interventions proved successful because they could credibly serve as a clearinghouse for private beliefs: students were surveyed about what they thought others wanted to drink, and how much they would choose to drink if it were up to them. Once this widespread disconnect was revealed, everyone could reduce their drinking without fear of social sanction.

Pluralistic ignorance brings out a second aspect of the Bicchieri definition of norms. There can be a disconnect between an individual's personal preferences for her actions and what her actions are. Norms do not have to be endorsed by those who follow them. Pluralistic ignorance is the extreme version of this, where no one endorses the norm, but it persists anyway.

---

[5] See Michael P. Haines, "A Social Norms Approach to Binge Drinking at Colleges and Universities," *The Higher Education Center for Alcohol and Other Drug Prevention*; H. Wesley Perkins, "Social Norms and the Prevention of Alcohol Misuse in Collegiate Contexts," *Journal of Studies on Alcohol and Drugs* 14 (2002): 164.

[6] Hans Christian Anderson's *The Emperor's New Clothes* is an exemplary account of pluralistic ignorance in a fictional context.

Much more common is the scenario in which some do endorse the norm, and some don't, but most or all people follow the norm. This matters a great deal, as we tend to model people's choices as following their personal preferences. Under a standard approach, we might use information campaigns or individual exhortation to convince people that a particular activity, like binge drinking, is unhealthy. The presupposition of this approach would be that once an individual is convinced, she would change her behavior in light of the new evidence. However, social norms are an important constraint on that model. Even if every individual is successfully convinced, if they do not know that their peers have also changed their minds, everyone will continue to persist in following the unhealthy behavior. This also points to an important way in which social norms take on deep political relevance. While we tend to think of the primary source of (unjust) coercion as the state, as Mill argued in *On Liberty*, the rest of society can be even more coercive than any state apparatus. Social norms can leave us stuck doing what we don't want to do, fearing sanction if we fail to go along with the majority.

A third feature of this account of norms, already implicit in the discussion thus far, is that norms are fundamentally about our perceptions of others. Norms are not simply what everyone does. For instance, brushing one's teeth in the morning is not a social norm, because we tend to be personally motivated to brush our teeth, or just find it's a habit. If I found out that everyone else had stopped brushing their teeth (and there had not been a large oral hygiene technological breakthrough to explain this change), I would continue to brush my teeth regardless of that they've done. Likewise, wearing shoes outside to protect my feet, or wearing a coat in the winter to stay warm are all actions that I take, and assume others will take. In instances like these, there are common solutions to common problems, but we are not socially motivated when we follow these common patterns of behavior. Norms are those areas where we monitor and regulate each other's behavior, not just where we all behave or believe similarly. For there to be a norm present, we must have consistency in our social expectations: we all think that others are following a particular behavioral rule in a particular context, and we all think that others think that we should follow it too.

So what should we take away from this account of norms? This definition is fundamentally epistemic in nature: it is simply about mutually consistent social expectations in a population. It does not itself delve into the importance of social identity or the expressive meaning of a particular norm, or the relationship between social norms and more deeply embedded cultural values. This is not a rejection of the importance of those concepts, but rather a claim that much policy work can be done with the more directly measurable facets of norms. Since this offers a general treatment of all social norms, it is difficult to say much on these contextually richer features of norms beyond that people often deeply care about the expressive meaning of some more

"core" norms and that the importance of our social identity is a human universal. Policies that rely on people directly rejecting deeply held values or social identities are unlikely to be successful. What this account can offer in much more detail is an approach for uncovering existing social norms, and identifying levers for promoting norm creation or change.

III. Norms and Coercion

Antanas Mockus, a former mayor of Bogota, Colombia, argues that there are three main mechanisms for regulating human behavior: moral, social, and legal norms. Policymakers traditionally have focused on legal norms. However, passing laws does not guarantee that the desired behavior is obtained. This can be seen across many areas: banking regulation can fail if banks follow the letter but not the spirit of the law, and take advantage of more and more loopholes or regulatory arbitrage opportunities. Intellectual property laws fail to achieve their regulatory goals if young people see all their friends downloading music illegally. Tax law is subverted when everyone knows everyone cheats on their taxes, as is the case in countries like Italy and Greece. Traffic laws are ineffectual if everyone speeds. Laws themselves have no authority until people agree to grant them that authority. This is granted via our moral norms and our social norms acting in concert with the legal norms. In Western countries, we are fortunate that we all largely have agreed to grant laws such authority, but as I have indicated, there are many cases where particular laws are ignored. Problems like this can be even more severe in developing countries, which can lack a history of or capacity for effective governance. Mockus argued that an important role for government was to work on harmonizing moral, legal, and social norms, so they mutually reinforce each other. When they are in harmony, they are powerful regulators of behavior. When they are close enough together, they can serve as a ratchet—moral and legal norms can pull social norms closer to them when they are seen as feasible changes from the status quo. However, when they are far enough apart that they all point in different directions, legal norms will be ignored, and often, moral norms will be as well. Social norms are motivationally powerful, and while we can all understand the importance of moral and legal regulation, when such regulation is too distant from the realities we face, it has no particular sway over us. Just as it is dangerous to drive the speed limit when everyone around you is speeding, it can be risky to be the only person who doesn't accept or give bribes.

Let's consider these forms of regulation for a moment. Laws clearly coerce—the state backs laws with its monopoly on force. Failure to comply with the law can bring on the police, the court system, and ultimately prison. The state is authorized to restrict one's freedom in a variety of ways as a means of enforcing the law. Legal regulation is clearly an external constraint on action.

Morality, on the other hand, is broadly understood as an internal constraint on one's action. Following one's moral compass involves eliminating possible actions oneself. One's guilt is purely first-personal. One feels guilty about having committed some offense even if no one else knows. Moral norms, then, aren't coercive in any politically relevant sense. Morality causes us to limit ourselves, regardless of what we might see others do or what they might expect from us.

Social norms share some features of both moral norms and legal norms on this front. Social norms can be a form of external constraint—we may experience shame from the violation of a norm, rather than (or in addition to) guilt. That is, when we violate a social norm, we can worry about what others think of us, and can internalize a feeling of the approbation of others. Further, those who violate social norms are frequently punished by others in the community. So with social norms, we find that people can regulate their own behavior because of their concerns of what others will think of them, and when that is insufficient, communities can punish social norm violations.

Seeing the coercive structure of norms can help us appreciate why norms are an important social phenomenon to understand, since they can powerfully inhibit or enable legal norms if they are counter to or in alignment with them, respectively. We can also consider them as potential *replacements* for legal norms in a variety of contexts. For instance, while we may have very good reasons more generally for wanting to support friendships, we likely do not want the state to have a role in coercively ensuring that people treat their friends well. Social norms governing friendships can fill that role. If Anna treats Barbara poorly for no good reason, Carol might stop inviting Anna to dinner, and find other ways to punish Anna for her bad behavior. Social norms can coercively prop up good behaviors in a variety of areas where we have no interest in state control.

Because of this, social norms look like potentially good candidates for the broader policy toolbox, alongside nudges. Both aim to reduce explicit state regulations with all that this might entail—police authority, exposure to the court system, and so on—all while aiming to achieve democratically determined policy goals.

There is a danger to the coercive structure of social norms as well. As we will see in the later discussion of Mill, social norms can sometimes coerce in ways that are more harmful than state coercion. The state can imprison you, but social norms can reshape who you are. And while we can repeal overly coercive laws, it can be far harder to get rid of a harmful social norm. Once set in motion, norms can persist across generations, even after the animating reason for their existence has gone away. For example, Alberto Alesina, Paola Giuliano, and Nathan Nunn[7] have argued that plow-based

---

[7] Alberto Alesina, Paola Giuliano, and Nathan Nunn, "On the Origins of Gender Roles: Women and the Plough," *Quarterly Journal of Economics* 128, no. 2 (2013): 469–530.

agriculture generations ago dramatically shaped the gender norms that persist to this day, despite the fact that people are no longer relying on human-powered plows. If the state chooses to wield this kind of regulatory tool, there should be heightened scrutiny about whether this is the kind of coercion that we want for ourselves.

IV. Diagnosis

There are two distinct cases in which a social norms approach might be applied: a diagnostic scenario, and a design scenario. In the diagnostic approach, the government agent seeks to understand the current state of affairs. In the design case, the government agent looks to use social norms as part of an intervention to improve the current state of affairs. Though these are both parts of the larger policymaking process, it is important to treat them as distinct entities. This is for two reasons. First, there are distinct normative issues at play between being sensitive to social norms in diagnosis and trying to create or change them in policy design, and as a result, we should consider each case separately. As mentioned before, a policy-induced change of norms or a policy-initiated creation of new norms carries with it the risk of policymakers engaged in social engineering.

Diagnostic tools do not run that risk. Second, the academic literature on social norms is much more definitive on how existing norms affect people's behavior, as compared to how to intentionally create a particular norm to achieve a particular policy goal. While we can identify some successes, and identify what structural features might need to be present, there is still a great deal of room for policy innovation and experimentation.

Let's first consider the diagnostic stage. To better situate the discussion, I offer a very brief sketch of a diagnostic process tailored to investigating social expectations.[8] While the details of this diagnostic approach can be found in Bicchieri,[9] our goal here is merely to get clearer on the main concepts at play.

After all, we frequently think about community or cultural norms. A challenge with this language is that it can be overly broad and imprecise, making it difficult to translate into diagnostic tools or policy interventions. So our first task is to get clear on the consequences of the definition that we are working with. If we recall the definition of a social norm, norms are situated in particular populations. Deploying social diagnostics makes sense when one finds a population of people exhibiting the same kinds of behavior, or publicly expressing similar attitudes. For instance, a city full of people who all follow traffic rules very carefully is much more likely to be explained by the presence of a social norm than lots of

---

[8] This process follows analytically from Bicchieri's definition of norms in *The Grammar of Society*.
[9] Bicchieri, *Norms in the Wild*.

individual preferences that just so happen to be aligned.[10] Our next task is to determine if there is such a group of people who follow the same rules of behavior.

Establishing that there is a common behavior does not yet establish that there is a norm. An equally plausible explanation is simply that the behavior is just a good idea, given the choices everyone faces. For instance, brushing one's teeth is in everyone's interest, and we tend to have preferences to do it regardless of what anyone else does.

Our next step is thinking about whether members of the community under consideration believe that most other members of the community follow the behavioral pattern—that is, do they share a consistent set of empirical expectations. If we find that the population under consideration has a consistent set of empirical expectations, we need to know whether they think others attach normative significance to the behavior in question. So far, we just have people imitating each other, which is similar to a fashion or a fad. It can be cognitively simpler to just follow what others do in a given situation rather than to try to work out what is best to do. However, if people believe that others expect them to behave in the same way, then this suggests that there is a social norm.

The diagnostic approach offers us an opportunity to combine our more general knowledge that humans are reliably social creatures with details of how this social nature manifests itself (or doesn't) in a particular context. While we may appreciate that people are in general conditional cooperators, have a sense of fairness, and like to punish those whom they take to be violating social rules, these tendencies alone can't inform us of how communities use them. For instance, the general human preference for reciprocity can manifest itself in positive ways, such as neighbors inviting each other over for dinner, or friends helping each other move, or it can manifest itself in negative ways, like revenge killings and long-term familial or tribal feuds, sustained by tit-for-tat attacks. We can even find that what might appear to be positive is in fact negative: gift-giving can induce obligations to the recipient to reciprocate, to the point where people will avoid or refuse gifts to avoid the instantiation of the obligation. Gift exchange economies often involve ever-larger gifts to enhance social status. Just knowing that humans tend to like reciprocity does not let us know how reciprocity will manifest itself.

---

[10] Importantly, social norms can also cause people to fail to exhibit certain behaviors: norms may dictate that individuals should not trust people outside of the group, for instance. Norms can likewise be so onerous that people avoid putting themselves in certain situations, to avoid triggering the obligations that a norm might impose: earning more money than one immediately needs may obligate the earner to loan or give her excess money to extended family or friends who have less. Accepting help or a gift may trigger an obligation to reciprocate even more generously. These norms can be powerful, even if we never see their prescribed behavior. Norms can also proscribe behavior: anti-corruption norms will mean that we systematically would not see public officials accept bribes.

Instead, we have to delve more deeply into the particular set of social expectations that hold a population's norms in place.

For example, Uri Gneezy and Aldo Rustichini argued that social incentives stemming from reciprocity norms are a more effective deterrent to late pick-ups by parents at daycare facilities than monetary fines—indeed, monetary fines induced Israeli parents to show up *later* than under a no-fine regime because they treated the fines as a *price* for tardiness—and the price was low enough to make staying at work longer worth it. What's worse, after the fines went away, tardiness remained higher than before, as the social incentives to be on time were broken by the fine.[11]

By uncovering this information about the population, it becomes much more apparent how deeply social we are. It helps reveal how communities can constrain and support each other, by helping to enable certain behaviors while making others more difficult. Understanding which behaviors are constrained, and which are enabled, and for what reason, can inform any form of subsequent policymaking, whether or not it is norms-based.

At the same time, a diagnostic approach can serve to highlight areas where pre-existing social norms can be activated to better support regulatory goals. For example, the British "Nudge Unit" successfully increased the rate of tax compliance with what could be described as a social nudge: in a letter to tax delinquents, they alerted them the percentage of their neighbors who had paid their taxes on time. This increased tax revenues by more than a hundred million pounds. People knew what the right thing to do was, and framing it as a pre-existing social norm provided them with the extra motivation they needed to do it. Smaller-scale versions of this same idea have proven to be successful in other areas of policy as well. The best-known case of this is with OPower, which collaborates with energy companies to send consumers bills that alert them to whether they are using more or less power than the average of their neighbors. This is also accompanied with a smile if they are below average, or (in some cases) a frown if they are above average, along with suggestions on how to conserve power. This resulted in net power savings of 0.3-6 percent, for virtually no cost.[12]

Let's reconsider Mockus's work in Colombia. He exemplified the thought that it is not just the idea that laws and the regulatory state may be undermined when social norms are out of line with them, but that we should think of law and regulation in relation to existing norms and values.

[11] Uri Gneezy and Aldo Rustichini, "A Fine is a Price," *Journal of Legal Studies* 29, no. 1 (2000).
[12] Hunt Allcott, "Social Norms and Energy Conservation," *Journal of Public Economics* 95, no. 9 (2011): 1082–1095. Interestingly, some of the variation in power savings appears to be explained by ideological affiliation. Liberals were more likely to conserve, and conservatives were more likely to be bothered by the attempt at invoking a norm. Dora L. Costa and Matthew E. Kahn, "Energy Conservation "Nudges" and Environmentalist Ideology: Evidence from a Randomized Residential Electricity Field Experiment," *Journal of the European Economic Association* 11, no. 3 (2013): 680–702.

A diagnostic approach to social norms helps to reveal where there may be opportunities to draw support for policy from within existing social norms, and where there may be barriers. This encourages an incremental approach. Mockus himself modeled this insight through legislative initiatives. In a city with high rates of gun-related violence, rather than imposing a comprehensive gun ban, he banned guns on the weekend, arguing that for the sake of preserving life, people could leave their guns at home two days a week. Rather than being a radical change, Mockus argued that it was a small change from current practice, making it more likely that citizens would go along with it. Social norms help frame the space of possibilities for more formal institutional change, as their authority is independent of the state's authority.

A diagnostic approach helps us uncover where there are conflicts between our formal and informal institutions. While the diagnostics themselves do not tell us how to reconcile these conflicts, the larger approach helps us think through where opportunities for policy improvement might be. Before policymakers seek to make changes, it is crucial to have a clearer understanding of the social landscape. Without an understanding of existing social norms, it is difficult to know what constraints the policymaker faces, and what resources she can draw from.

This diagnostic approach has some technocratic virtues, but it also raises some interesting normative questions about the nature of the state's authority and its role in the broader system of social regulation that we saw in Mockus. Nudges, for instance, are meant to leave open our full option sets and not impose additional material constraints: nudges don't change prices or eliminate options via regulation. Instead, nudges impose differential cognitive costs on various choices. The socially preferred option is meant to be easy, whereas other options take more effort. So, the policymaker chooses good defaults, and while determined agents can choose something else, most agents will just passively accept the choices made for them.

This passivity is often the source of worry: nudges may obscure the fact that a choice is being made, and in doing so, reduce the citizenry's capacity for choice by depriving it of small-scale practice in choice making. While I don't wish to rehearse the arguments for and against nudging here, I do want to highlight that the arguments for and against tend view individuals as *individuals*—that is, independent choosers. Whether Bob eats salad or a burger for lunch is cast as *his choice*, independent of what Alan chooses. But as Christakis and Fowler demonstrated, our food choices are caught up in what our friends and colleagues choose.[13] In the extreme case, for Bob to move away from eating burgers for lunch and eating salads instead, it might mean not having lunch with Alan anymore, and having lunch with Colin instead. Nudges presuppose a wide range of

[13] Nicholas Christakis and James Fowler, "The Spread of Obesity in a Large Social Network over 32 Years," *The New England Journal of Medicine* 357 (2007): 370–79.

independent individual behaviors that may or may not be there, because they broadly ignore the social constraints that individuals are under. Interestingly, nudges occupy a somewhat peculiar spot in the normative landscape—they presuppose that we make individual choices, but that our choices can be effectively overridden by the choices of others by altering the cognitive cost of choice. Nudge-oriented policy focuses on the citizen-state relationship, and has the state hold traditional forms of coercion at bay by replacing them with differential cognitive loads. This could be justified because the citizen *authorizes the state* to act on his behalf, either as a product of democratic deliberation about the ends that the nudge aims to implement, or by a more general contract between the citizen and the state. Particularly on this latter view, the state may take itself to have an obligation to serve the ends of the citizens with as little coercion as possible. So rather than alter legal restrictions, fines, or subsidies, the state adjusts what is cognitively easy to do. But nudge-oriented policy ignores other people, and fails to treat individuals as caught up in a system of social relations.

By focusing on social norms, we more clearly see the social nature of a very broad set of our behaviors, precisely because the diagnostic approach aims to uncover the structure of our social expectations that holds those behaviors in place. This diagnostic approach to social norms, which simply aims to shine a light on the informal institutions that prop up or inhibit our behaviors, is consonant with what I am calling a more Humean approach to the state. Hume argued that social rules are conventional, in that they provide a solution to a coordination problem—a solution which enabled individuals to get much more of what they want than they could on their own.

In the *Treatise*, Hume says, "Two men, who pull the oars of a boat, do it by an agreement or convention, tho' they have never given promises to each other. Nor is the rule concerning the stability of possession the less deriv'd from human conventions, that it arises gradually, and acquires force by a slow progression, and by our repeated experience of the inconveniences of transgressing it. On the contrary, this experience assures us still more, that the sense of interest has become common to all our fellows, and gives us a confidence of the future regularity of their conduct: And 'tis only on the expectation of this, that our moderation and abstinence are founded. In like manner are languages gradually establish'd by human conventions without any promise."[14]

The rules that we follow, even our sense of justice, are purely conventional on Hume's account. They are slowly established by each of us adjusting to the others until we end up in a stable equilibrium of behaviors.

---

[14] David Hume, *A Treatise of Human Nature*, 2nd edition, ed. L. A. Selby-Bigge, (Oxford: Clarendon Press, 1978 [1896]), 60.

What drives this convergence to an equilibrium is that we are each seeking a set of rules that serve to align our individual interests with a common interest. Once this equilibrium is achieved, it's in our interest to follow the rules to better support the regularity of everyone's behaviors. Hume took this approach to explain the rules of public morality in large part—these methods of regulation are simply what conventions we have settled into. Once a convention is established, in the long run it is in one's interest to follow it, even if we can identify individual instances where we might benefit from going against convention.

Note here that Hume is explicitly rejecting any separate foundational claims about justice beyond a convention—hence the "without any promise" mention in the quote above. Hume thinks core questions of justice and the main tools of social regulation stem from informal institutions, not formal ones. Formal institutions merely reflect what is conventional. Hume doesn't offer an independent source of justice or authority beyond that which has been established by these processes.

We can readily find the connection between Hume's conception of justice and the diagnostic approach. This approach is by its nature somewhat conservative. It merely turns our focus to the sets of informal institutions that are often somewhat invisible to the state, and seeks to ensure that policy is done in concert with those informal institutions. Since on Hume's account the state doesn't have an independent claim to establish the rules beyond what is conventional, we can conceive of the state instead as a formalization of informal institutions. The diagnostic approach simply amplifies this idea by explicitly calling attention to informal institutions.

The most radical shift in behavior discussed in this approach was a reduction of binge drinking on college campuses. Note that this shift was possible because individuals *didn't want to drink as much as they were*. In cases of pluralistic ignorance, states can play the role of information-provider, but the information provided is *social* information, not merely factual information about health. Here, we find that the equilibrium settled upon was not responsive to the underlying individual interests of the members of the community.

Hume does not directly engage the possibility of pluralistic ignorance, but he does hold as a core principle that conventions serve the public interest. On those grounds we might suspect that this provides a principle for altering informal institutions: change is called for when compliance with them leads everyone to be made worse-off. Hume says, "[f]or since any considerable alteration of temper and circumstances destroys equally justice and injustice; and since such an alteration has an effect only by changing our own and the publick interest; it follows, that the first establishment of the rules of justice depends on these different interests."[15]

---

[15] Hume, *Treatise*, 70–71.

Likewise, Hume notes "'Twas therefore a concern for our own, and the publick interest, which made us establish the laws of justice."[16] Informal institutions serve us when they align our long-term interests with the public interest. While Hume doesn't seem to entertain the possibility that these informal institutions could fail to do so, the violation of this criterion would seem to serve as a good reason for society to throw those institutions out and try to develop new ones.

In other diagnostic approach cases, behavioral shifts were triggered by making social incentives more salient. That is, rather than deploying any formal state powers of coercion, the state worked to increase the recognized authority of communities to regulate themselves. In this role, the state functions more like a broker of social information, better enabling communities to monitor themselves. The OPower case is illustrative in this regard. While liberals took the information provided on their bills to work to lower their energy usage, conservatives used that information to increase their energy usage. A plausible theory for why that would be is that they are following different sets of social values, and using this increased monitoring capability to better express those values through their behavior.

While this is not a logical necessity, we can straightforwardly see the diagnostic approach to social norms as one consonant with a Humean conservative impulse. On this approach, the government provides tools to communities to regulate themselves, and in doing so it strengthens informal institutions and weakens, or at least does not strengthen, the formal power of the state. This would at least be consistent with an increase in local political control, and strengthening civil society as opposed to formal political authority. Particularly where we expect there to be regional variation in moral attitudes, such an approach would allow for more people to live under the kinds of rules that that align their underlying interests with the community interest. This comports well with a view that what Hume called conventions, and Bicchieri calls norms, are the relevant source of normative justification for conforming to social rules. Political authority is found in norms because for them to be stable it means that they align our private interests with the public interest. The state can then merely shine a light on these informal institutions and allow them to do the work of regulating our behavior.

V. Design

The design approach to policymaking is a mechanism for the state to use its understanding of social norms to attempt to change or dismantle existing norms, or create a new norm. As mentioned before, this practice can be fraught with ethical challenges. Social norms richly inform our lives,

---

[16] Hume, *Treatise*, 70.

and the thought of the state directly intervening to shape those norms to engineer certain outcomes can easily be seen as disturbing. The idea of the state somehow changing the nature of how we relate to each other in the private sphere is understandably disconcerting, as such change could go well beyond the legitimate scope of government.

However, if we instead turn to John Stuart Mill rather than David Hume, we can see a reason for why such an approach might be desirable, at least under certain circumstances: " . . . reflecting persons perceived that when society itself is the tyrant—society collectively over the separate individuals who compose it—its means of tyrannizing are not restricted to the acts which it may do by the hands of its political functionaries. Society can and does execute its own mandates . . . it practices a social tyranny more formidable than many kinds of political oppression, since . . . it leaves fewer means of escape, penetrating much more deeply into the details of life, and enslaving the soul itself."[17] Mill was relatively unique in identifying not just the state as a source of potentially unjust authority, but also the rest of society. So, if we are to maximize individual liberty, we may have reason to deploy the state to upset the coercive power of social norms. Mill contrasts the "tyranny of the magistrate" with the "tyranny of the prevailing opinion" and argues that one needs protection against both, as there is a "tendency of society to impose, by means other than civil penalties, its own ideas and practices as rules of conduct on those who dissent from them; to fetter the development, and if possible, prevent the formation, of any individuality not in harmony with its ways . . . ."[18] Mill saw the informal coercion of others as simply another form of political oppression, even if the exercise of this power did not go through the state.

This offers us two possibilities worth considering. First, we might wonder whether the tyranny of the majority simply has two different manifestations—that of the magistrate and that of prevailing opinion—and they will always be in concert with each other. The second is that sometimes the power of the magistrate can serve to counter the power of prevailing opinion.

In the first instance, we have quite strong reasons to resist any effort on the part of the state to create social norms that more fully entrench the tyranny of the majority. This quite obviously extends the power of the state into those areas that can, ultimately "enslave the soul." The pervasive security state of East Germany may be a useful example of this phenomenon—not only did the state have repressive laws, but it encouraged citizens to monitor each other and report on those who were somehow suspicious. Here we find the magistrate amplify its power by means of capturing the power of prevailing opinion.

---

[17] John Stuart Mill, *On Liberty* (New York: Hackett, 1978), 4.
[18] Ibid.

Jim Crow laws in the American South appeared to go the other way: laws were created to better align with prevailing social norms about the moral status of different races. So prevailing opinion used the power of the magistrate to further entrench their despotism. In both of these cases, the two kinds of tyranny of the majority are aligned, and clearly this alignment is rather bleak for individual liberty.

The second possibility might prove to be more hopeful. Lawrence Lessig offers a useful example of how the power of the magistrate might be used to counter the power of prevailing opinion.[19] He argues that the Civil Rights Act worked to reduce racial discrimination by reducing the power of social norms around discrimination, rather than through courts and government enforcement. Essentially, the government never had enough agents to make it possible coercively to enforce desegregation of restaurants and businesses; and so if large majorities had wanted to continue the practice, it would have been quite challenging to do much about it legally. Instead, the law gave a number of individual small business owners a way to do what they wanted to do already—hire freely and serve a broader range of customers—without fear of social punishments. Discrimination was socially enforced. If a restaurant let minorities eat there, at best whites would boycott, and at worst, the business would be attacked, as the owner failed to uphold the norm. But in a regime where discrimination was illegal, the owners could have said that, while they would prefer to discriminate, they also did not want to break the law and endanger their business, and so they were forced to stop discriminating. Thus, the law gave someone who wanted to integrate an excuse for doing so that wouldn't run afoul of the social norm. State power, while coercive, was used to (at least partially) defang another source of coercion that importantly limited individual liberty.

We may, of course, ask whether the Lessig story is complete. The Civil Rights Act surely didn't eliminate all discrimination, nor did it end a variety of forms of segregation. Norms may well be far stickier than we might like. Lessig's point, of course, was that a legal change freed people who *didn't like the norm* to do something other than follow it without worrying as much about being punished. But plenty of people undoubtedly liked the norm well past the passage of the Civil Rights Act. Indeed, passing a law that is counter to an established social norm can either cause the law to simply be ignored (as it is with jaywalking, for instance) or increase the salience of the illegal behavior. But, if Lessig is right, the law can serve to chip away at the support of some norms if not everyone endorses them, as in those cases where antipathy for the social norm does not rise to the level of pluralistic ignorance, but does indicate less than full support.

By definition, social norms constrain one's option set. Sometimes, as in the case of norms that foster generalized trust, reciprocity, and other

---

[19] Lawrence Lessig, "The Regulation of Social Meaning," *University of Chicago Law Review* 62 (1995): 943–1045.

behaviors that enable us to live together cooperatively, those limitations actually free us up in a variety of other areas. But perhaps even more often, social norms serve to enforce homogeneity and can limit people's ability to endorse their own plan of life. They can informally codify a reduction of rights for a subgroup, or sanction and normalize systematic violence. While this coercive power can sometimes be rather mild, as in social norms that broadly govern what is appropriate to wear in different circumstances, they can be much more constraining, as with norms that govern the split of household obligations between men and women, or the use of domestic violence as a means of regulating gender norms. Outside the household, social norms can work to require violence at perceived slights, or limit one's academic achievement, or constrain one's economic ambitions. These are all areas where social norms not only coerce, but they can do so in ways that are potentially more objectionable than state coercion. While you can limit your contact with the state, you can't get away from your neighbors, friends and family. This is why Mill saw this kind of coercion as something that can "enslave your soul."

So, we may wish to follow Mill in deploying the harm principle not just to state coercion, but also to social coercion. In doing so, we may need to deploy the state to alter that social coercion. One possible approach is similar to what Lessig outlined: namely, we create laws that prohibit behaviors that are supported by social norms so as to undercut the empirical and normative expectations of the population that follows the social norm. However, there we are substituting one coercive authority for another. There are a few reasons to be worried about an overuse of such an approach. First, the government might not be the best situated to determine the presence or degree of coercion. While the Civil Rights Act addressed clear harms, there may not be that many cases of such obvious social coercion, such that the government will be an appropriate source of remedy. Second, we may worry that the government will overuse such an authority to eliminate merely disfavored social norms, rather than social norms that coercively harm individuals. Third, for Millian reasons it is far from clear that even if the government can appropriately identify bad social norms that ought to be changed or eliminated, it may not be well situated to determine what if anything should take their place. Even holding up the Civil Rights Act as an exemplar for the government using its coercive power to reduce informal coercive power, the state also wielded coercive power against African Americans before and after the law's passage. Finally, we can reasonably expect that social changes are more sustainable when they come from within, rather than from the outside. *Roe v. Wade* was decided quite some time ago, and yet there has been very little social resolution of the acceptability of abortion or the conditions under which it is allowable. The legal "resolution" of the issue did nothing to shift underlying social norms, and if anything further inflamed the "pro-life" movement.

An alternative approach that would avoid such concerns would focus on community determination of norm change. On this model, the state serves as a facilitator and educator. Policymakers essentially help elevate issues that are brought to their attention, and create formal space for education and advocacy. This is a more deliberation-oriented approach to norm change, where citizens are intimately involved with choosing the norms that they want to be held to. On this sort of approach, the state functions as a trusted third party that can facilitate community discussion and commitment, rather than a source of imposed rules. This is a model that was widely used during Mockus's two mayoral terms in Bogata, and it also has formed the basis for Community-Led Total Sanitation efforts to eliminate open defecation in the developing world. To a much larger degree, NGOs such as Tostan have committed themselves to multi-year programs that are more about education and facilitation of norm change, allowing communities to make choices for themselves, rather than deliberately engineering particular outcomes without. Tostan has used this approach to encourage the elimination of Female Genital Cutting, primarily in Senegal, but because of their more comprehensive approach, it also sees spillover effects in gender relations, public health, and basic literacy and problem solving.

This approach is more long term than many policy initiatives, but it has the promise of being more sustainable. As it focuses on community deliberation around what values the community holds, and how they wish to express those values, there is an opportunity for greater stakeholder acceptance. In virtue of the greater levels of community input and engagement, it may also be easier to develop policies that get implementation details right. It also has more appealing normative features: it aims to alter coercive relations by means of a deliberative, participatory approach, and relies on the state not as an enforcer, but as a broker. The state can work to reduce social violations of the harm principle without itself relying on coercion.

This version of a design approach to social norms is promising, but more limited in scope than the diagnostic approach, both in virtue of the time costs of a deliberation-oriented intervention, and in terms of the appropriate areas of application. While this may work well for reducing gender violence or improving childhood nutrition, it is less likely to work against areas of policy where there are extreme power differentials. That is where the Lessig approach may make more sense.

VI. Conclusions

States have good reasons to look beyond the traditional tools of the state—legal restrictions, taxes and subsidies, and information—to better achieve democratically determined policy goals. There is ample empirical evidence that outcomes can be improved if we account for a more nuanced

understanding of humans. There is also the hope that the state can accomplish this all while reducing state coercion. This effort began with "nudge" policy. But we have seen how this ignores important aspects of our sociality.

Social norms, as Bicchieri and others have argued, present an opportunity for improving social outcomes without relying on the formal power of the state.[20] However, there are at least two basic ways of conceiving of the interplay of social norms and state authority. The diagnostic approach has the state merely uncover and highlight social norms, and perhaps point out where norms are out of step with individual interests. I've argued that this approach is consonant with a more Humean take on the source of legitimate authority, where the state is more of a formalization of the informal institutions that underlie it. The design approach places the state in a position of creating new coercive powers. In the example of the Civil Rights Act, formal coercive powers were introduced to counter informal coercion. This relied on Mill's distinction between the tyranny of the magistrate and the tyranny of prevailing opinion. I argued that when the state can put the two of them in conflict with each other, they may serve to weaken each other. An alternative formulation of the design approach is one in which a new norm is introduced, rather than one according to which the formal authority of the state attempts to weaken or eliminate an existing norm. Here, following Mockus and Tostan, I've argued that for norms to be legitimate, they must come out of a deliberative democratic process that arrives at agreement.

In this essay, I've sought to engage in a hybrid task—exploring the emerging norms policy landscape from the perspective of a policymaker, while evaluating the normative status of those policy approaches from the perspective of a political philosopher. There is a fundamental tension in this. Bare interest in efficacy ought to push the lawmaker toward the use of social norms as policy tools. But from a normative perspective, there are serious reasons for considering restraint. Social norms can have a far longer life than laws, and once social norms come into existence, the state doesn't have control over a new source of coercion over individual citizens. What I've called the diagnostic approach to social norms offers a relatively safe approach to using social norms in policy. Since the state is merely calling attention to existing norms, and perhaps showing where those norms no longer match underlying interests, the state is not working to increase coercion, but rather simply making existing informal institutions better aligned with underlying individual interests. The design approach is far riskier. It is fundamentally about introducing new sources of coercion. I've considered two sorts of circumstances—one where the state is aiming to eliminate informal coercion by means of introducing new formal coercion, and one where the state is aiming to create new informal coercion but no

---

[20] Bicchieri, *Norms in the Wild*.

new formal coercion. We see that the first case, which is opposed to the picture that Mockus offered on harmonizing sources of social regulation, is aimed at creating *disharmony* for the purpose of weakening or eliminating coercive regulation of behavior. This disharmony may work to dampen both countervailing forces, but it also may increase the authority of a harmful norm. This should give us reason to worry about whether and how such a policy effort is made. The alternative that Mockus proposes is to find a way to deliberatively engage citizens in developing new norms that people have reason to endorse.

This might point to a limitation of what I'm calling the design approach. The design approach works if there is a general background of individual liberty. Hume's arguments about the origin and authority of convention stems from this sort of liberty assumption. Mill at least recognized that there can be informal sources of coercion, and he rightly warned that they can be far more severe than formal sources of coercion. When we lack this background of liberty, it appears that we have two options: using the state to counter informal sources of coercion, or more deliberatively attempting to find a new set of norms that protect individual liberty. Both options can potentially be effective for freeing up more space for individual liberty, but setting state and informal coercion against each other has a variety of potential challenges. Mockus may well be right that a more deliberative, harmonious approach is the better one, even if it is the most difficult to pursue.

Philosophy, University at Buffalo