Ethics, Politics & Society
A Journal in Moral and Political Philosophy

The invisible hand of partisan irrationality

ORCID: 0000-0001-9451-291X
Received: 17/07/2023
Accepted: 13/05/2024

Jimmy Alfonso Licon
Arizona State University
jimmy.licon@asu.edu

ABSTRACT Why do we vote, protest, and boycott? Economists explain partisan actions, despite their costs, by arguing political irrationality by a single partisan isn't costly to them as an individual - they can afford the political irrationality, despite the social costs. And some philosophers worry about the moral and epistemic costs of political irrationality. Here I argue that political irrationality has some benefits: it encourages partisans to engage in virtue signaling and rationalization in politics. And while virtue signaling and rationalization are often epistemically and morally bad, they can nonetheless confer benefits too, like facilitating societal and moral progress.
KEYWORDS Rational ignorance; political irrationality; rationalizations; invisible hand.

RESUMO Porque é que votamos, protestamos e boicotamos? Os economistas explicam as acções partidárias, apesar dos seus custos, argumentando que a irracionalidade política de um único partidário não lhe é onerosa enquanto indivíduo - pode permitir-se a irracionalidade política, apesar dos custos sociais. E alguns filósofos preocupam-se com os custos morais e epistémicos da irracionalidade política. Neste caso, defendo que a irracionalidade política tem alguns benefícios: encoraja os partidários a envolverem-se na sinalização da virtude e na racionalização da política. E embora a sinalização da virtude e a racionalização sejam muitas vezes epistémica e moralmente más, também podem conferir benefícios, como facilitar o progresso social e moral.
PALAVRAS-CHAVE Ignorância racional; irracionalidade política; racionalizações; mão invisível.

Introduction

Begin with a puzzle: partisans politically participate through voting, protesting, and other means even though individual actions are *highly unlikely* to change the election outcomes. Most people, even many academics, regularly overestimate the effectiveness of individual political actions. How likely is a single vote to decide an election outcome? Not very:

> [We] estimate the probability of a single vote being decisive as, at most, about 1 in 10 million in a few states near the national median. Averaging

> these probabilities over all the states and weighting by turnout yields a 1 in 60 million chance that a randomly-selected voter would be decisive (Gelman, Silver, and Edlin, 2012, p. 321).

The odds then that your *individual* vote will make a difference are slim. Here one may think that casting a vote actually requires little effort: one need only register, arrive at the polling place, and cast a vote. However, this underestimates the many hours of study required to be informed on a host of issues like economics, foreign policy, and education. It requires hundreds of hours of study to be an expert in *just one* of these areas – expertise enough to cast a vote across several such fields would require a lifetime. However, though elections are unlikely to be decided by a single vote, voting matters in the *aggregate*. As Jason Brennan explains,

> If *we*, the electorate, are bad at politics, if we indulge fantasies and delusions, or ignore evidence, then people die. We fight unnecessary wars. We implement bad policies that perpetuate poverty. We overregulate drugs or underregulate carbon pollution. But the problem is that we, the electorate as a whole, don't make choices about whether to be informed or rational about politics. Individuals decide for themselves in light of their individual incentives (2016, p. 24—original emphasis).

It thus takes a *long time* to learn enough to cast an informed vote—and yet an informed vote is comparable to an uninformed one where the outcome is concerned. We thus arrive at a puzzle: given that individual voters and partisans are highly unlikely to influence the outcome of an election—comparable to winning a state lottery twice—*what is the point of voting, boycotts and protesting*? A highly plausible explanation: people vote and politically engage to express tribal and social identities (Cohen, 2003). Here one may object: voter incentives and behavior admit of other plausible, competing explanations, e.g., someone who votes despite knowing their vote is highly unlikely to matter, may do so because they believe it is part of their civic duty (Jones and Dawson, 2008). Unfortunately, though, despite the fact that some voting behavior explained by factors like voters' sense of duty, the empirical evidence generally favors the social identity and tribal explanation for their behavior. As Barber and Pope explain:

> Using a research design that employs actual "conservative" and "liberal" policy statements from President Trump, we find that low-knowledge respondents, strong Republicans, Trump approving respondents, and self-described conservatives are the most likely to behave like party loyalists by accepting the Trump cue—in either a liberal or conservative direction. These results suggest that [...] *group loyalty is the stronger motivator of opinion than are any ideological principles* (2018, p. 38—emphasis mine).

The same applies to even the most informed voters and partisans:

> [People] are *often unable to escape the pull of their prior attitudes and beliefs*, which guide the processing of new information *in predictable and sometimes insidious ways*. But what does this mean for citizens in a democracy? From one perspective the average citizen would appear to be *both cognitively and motivationally incapable of fulfilling the requirements of rational behavior in a democracy*. Far from the rational calculator portrayed in enlightenment prose and spatial equations, *homo politicus* would seem to be a creature of simple likes and prejudices that are quite resistant to change (Taber and Lodge 2006, p. 768—emphasis mine).

So, even if there are voters who are motivated more by their sense of duty than their social and tribal identities, they aren't *only* motivated by a sense of civic duty—they are biased and engage in motivated reasoning. Nor are civic minded citizens representative of voters in a democracy. Voters are often too tribal, and identity minded, where it suited their tribal biases and irrational beliefs, to be epistemic rational in their political activities. For example, at the time of writing, Democrats and Republicans are divided over public labor unions. Democrats favor teachers' unions, but not police unions, Republicans favor the reverse. This is strange: if one thinks police unions are corrupt because of the difficulty firing a corrupt or incompetent police officer, then similarly, they should deem teachers' unions corrupt too. And vice versa. If, however, support for one union or the other were an exercise in signaling partisan affiliation, this strange mix of policy positions makes sense (Brennan and Lomasky, 2008).

We should pause here to make a clarification essential to our thesis. Though in the earlier parts of this paper we will be focusing on voters, the discussion over political irrationality and how it feeds into practices, like virtue signaling and rationalization in politics, *applies*

equally to political partisans and to other political activities like protesting, marching, boycotts, and letter and email writing campaigns, and the like. Voting, boycotts, and similar, by an individual, have but a slim chance of shifting a political or a policy debate or deciding the outcome of an election. Why then do we focus to such a degree on voting earlier in the paper? The reason is incidental: academic work on the effectiveness of political involvement has an outsized focus on voting. However, other partisan practices, like protesting and email campaigns *by a single partisan*, are not likely to be any more effective at influencing outcomes than single voters.

Philosophers and economists argue since that partisans lack incentives to be informed (Downs, 1957; Somin, 2014), and instead use politics to signal tribal identity, we get worse political results *in aggregate* than if voters and partisans were objective and informed. There is clearly something to this point: it would be good to have an informed electorate for no other reason than it may produce a better functioning democracy[1]. Despite that, in this paper, I highlight some *upsides* of partisan irrationality and political ignorance: voters and partisans use politics, not to pick the best candidates or policies, but to signal tribal affiliation by engaging in virtue signaling and in rationalization of partisan actions. Although these practices have many costs, they have benefits too, like social and moral progress—call this thesis *the political invisible hand* (PIH):

> Despite the intentions of partisans, virtue signaling and rationalization, regulated by self-interest to avoid plausible charges of hypocrisy and reputational harm, can improve societal and moral norms and behavior. While there is the risk that practices like virtue signaling and rationalization will often, though not always, increase tribalism and political polarization, here we will focus on the benefits, not the cost, of such practices.

Two clarifications are in order before proceeding. First, the thesis doesn't claim that the benefits of virtue signaling and rationalization in politics, like social and moral progress, outweigh the obvious costs of such practices—the costs likely outweigh the benefits (Hill and Fanciullo, 2023). The point of highlighting such benefits is to balance out the discussion of the epistemic vices of the political domain

1 One problem here is that voters, even if informed, would engage in motivated reasoning. So, informing voters isn't likely to improve democracy (Dusso and Kennedy, 2015).

involving virtue signaling and rationalization—it is easy to stress the costs of such practices and ignore the benefits. Second, throughout the paper, terms like 'rationality' and 'irrationality' refer to *instrumental* rationality and *epistemic* irrationality respectively—we will unpack this further in section two. The aim in using such terms is to describe and to make sense of voter and partisan behavior in the political domain, given their incentives, rather than to normatively judge it—this paper is meant to be more empirical than normative in that sense, merely highlighting potential, but small, benefits to virtue signaling and to rationalization given the poor and weak incentives to become an informed political actor.

In the first section, we delve into the details of why voters and political partisans are practically rational to remain ignorant of political information, and instead use political activities to signal identity. Second, we explore how political inefficacy incentivizes practices like virtue signaling: using moral language to boost one's reputation which, despite the downsides, sometimes confer benefits like fostering improve moral and social norms. Finally, we argue that rationalization can produce reputational pressure to morally improve behavior.

1 The rational irrationality of voters and partisans

Voters and partisans are often uniformed about politics, not due to lack of intelligence or a lack of education: it is rational to stay uninformed about politics, as an individual, because the odds that a single vote or protest sign will determine the fate of an election or public policy debate is shockingly low. Here we should clarify that terms like 'rationality' and 'irrationality' as they are used throughout the paper refer to *instrumental* rationality and *epistemic* rationality respectively. As Jason Brennan explains again:

> *A person is rationally irrational when it is instrumentally rational for that person to be epistemically irrational. Instrumental rationality* is about taking courses of action that serve one's ends. *Epistemic rationality* is about forming beliefs with the goal of seeking truth and avoiding error, using a scientific evaluation of the best- available evidence. It can sometimes be useful— instrumentally rational— for

> us to form our beliefs in an epistemically irrational way… [Suppose] one lived in a fundamentalist theocratic monarchy or something close to it, such as most of Europe in the Middle Ages or Saudi Arabia right now (2016, p. 48—emphasis mine).

Some critics argue that it is mistake to think individual votes do not matter with respect to the outcome of an election: the expected utility of voting, especially in a swing state, is high because there is a chance that one's individual vote will be decisive, and the resulting policies which are implemented by the politician one elected would be worth a lot (MacAskill, 2015, p. 86). The issue though with arguments like this is that it is opaque, even under the best of conditions, how much your vote would actually be worth. Such calculations are often made under the shadow of partisan motivated reasoning: where someone finds reasons and arguments for their favorite political views and candidates stronger than arguments and reasons for the opposing side as a product of their political loyalties. Partisans are thus in a poor epistemic position to objectively assess their own political activity (Freiman, 2021, pp. 58-62; Kunda, 1990).

Here one may object that in many democracies voting is often a private matter, and thus a poor avenue for signaling ones' social identity and tribal affiliations. There is some truth in that, but it isn't especially relevant for our purposes. Why? For the simple reason that while many voters opt to keep their political preferences and voting behavior a secret, sometimes even from family and close friends, many others blatantly advertise their political affiliations: they wear partisan clothing, attend political rallies, place partisan bumper sticks on their cars and political signs in their lawns. Voters and actors in the political domain do not operate homogeneous with regard to sharing or guarding their political allegiances. There need only be a major subset of political participants, who advertise their political activities, to encourage others to broadcast their social identity and tribal affiliation using politics too (Williams, 2020; Funkhouser, 2022).

Voters and partisans use political activity to signal social and tribal identity. It is worth pausing here to appreciate why human spend so much time and resources signaling social identity and tribal affiliation. As a species, we depend on cooperation from others to survive (Henrich, 2015) to the point that: people often prefer 'jail time, amputation of limbs, and death to' various forms of reputation damage' like acquiring a reputation as a Nazi or child molester (Vonasch et. al., 2018,

p. 604), kids as young as five prefer to maintain a good reputation refraining 'from cheating at the cost of losing a highly desirable prize' (Fu et. al., 2016, p. 277), and strong moral reputations are a major factor in mate selection, especially female sexual selection (Miller, 2007).

The strong tendency to signal social and tribal affiliation is no doubt the result, in part, of our individual need to belong to survival and flourish. These tendencies are manifest in the political domain because the cost of acquiring political information nearly always exceed the benefits of acquiring it for the individual, and because good (moral, social) reputations and tribal affiliation are adaptive. The instrumental rationality of voters and partisans preclude them wasting time acquiring political information that will not matter in their influencing politics or policy (Caplan, 2007; Downs, 1957).

Because casting a vote or attending a protest, for example, is ineffective on the individual level, this affords voters and partisans an opportunity to instead signal loyalty to the tribe, even and often at the expense of the facts, and at a very low personal cost to them as individuals—a bad vote is at most a miniscule cost to an individual voter. Due to this incentive structure, there are many signaling opportunities for reputation boosts and repair in politics since the outcomes of elections influences access to resources, power, and influence for the winning political coalition. As Dan Kahan writes,

> Where positions on some policy-relevant fact *have assumed widespread recognition as a badge of membership within identity-defining affinity groups*, individuals can be expected to selectively credit all manner of information in patterns consistent with their respective groups' positions. The beliefs generated by this form of reasoning excite behavior that expresses individuals' group identities. *Such behavior protects their connection to others with whom they share communal ties* (2016: 2—my emphasis).

In the political space where lack of political knowledge isn't costly *for the individual*, mechanisms like virtue signaling and rationalization can aid in affirming and protecting one's tribal identity and reputation. War is a great analogy: when groups fight over limited natural resources, often no single individual can make a difference if the armies are big enough. Because no individual can often do that much to determine the outcome of the war, individuals in each group must be loyal

and cooperate to win. Similarly with politics: group cohesion, identity, and reputation are both key to the survival and flourishing of the group and the individual members.

However, even within a group, there are often conflicts over things like resources, prestige, and power; the individual finds herself in a tension where she must belong to the group to survive, but also has a strong incentive to gain more from the group than she contributes. This is why group members engage in practices like virtue signaling and rationalization: individuals must convey that they belong to the group and are willing to contribute while angling to gain more than they contribute. And it isn't just that rational agents want more goods from the group than they contribute, but that they self-represent as morally good (Hardy and Carlo, 2011). There is substantial virtue signaling and rationalization in politics to look good to others and to preserve one's moral self-image, and even some moral and social benefits too. To elaborate, we first turn to the nature and benefits of virtue signaling.

2 The upsides of virtue signaling

Virtue signaling is the practice of using moral talk to improve or repair one's moral reputation and social status. Most people find the practice duplicitous and irritating. Here though we must inquire as to whether there are benefits to virtue signaling, especially since politics is rife with it. And given how many people view virtue signaling as fake and self-serving, it might seem odd to suggest there are benefits to it. Allow me to make the case.

Why are people averse to virtue signaling? First, and perhaps most obvious, is that people are motivated to virtue signal to boost their social status and reputation, to look good without the intent to do good—a practice that feels cheap and insincere. Second, some philosophers argue that virtue signaling parasitizes moral discourse because the practice is motivated by status-seeking behavior, virtue signaling can result in exaggerated moral claims, expressions of moral outrage, pilling on, and shaming that weakens public confidence in moral discourse. These effects can undercut public moral discourse as an avenue for moral and political progress (Tosi and Warmke, 2020). Since the practice of virtue signaling is often selfishly motivated, many are

tempted to conclude that little moral and social good could result from it.

There is, though, a subtle mistake in this line of thinking. It parallels the mistake the economist and philosopher Adam Smith identified in the quote at the start of the article: we benefit from the self-interested actions of others, even if that wasn't intended, or perhaps was the opposite of what was the intended. As Smith writes in *The Wealth of Nations*, it isn't 'from the benevolence of the butcher, the brewer, or the baker that we expect our dinner, but from their regard to their own interest. We address ourselves, not to their humanity, but to their self-love, and never talk to them of our own necessities, but of their advantages' (Ch. II, p. 19—original emphasis). Here it is worth noting that, just as with virtue signaling and rationalization, the marketplace processes are not always beneficial—there are environmental costs to unbridled consumerism—and that many benefits of the market result from greed and the profit motive. There is a disconnect here between aim and outcome just like with virtue signaling and rationalization, motivated by self-interest, but with some (unwitting) benefits for third-parties in both cases.

It may seem that there is a major disanalogy between the marketplace, and virtue signaling and rationalizations since the former doesn't need capitalists to base their actions on moral reasons to benefits others (unwittingly), while the latter practices of virtue signaling and rationalization, to influence social and moral norms for the (moral) better, presumably must comply with moral reasons for the right reasons. But that is false: virtue signaling and rationalizations may improve the social and moral norms in a group by incidentally complying with salient moral reasons, not because such practices were motivated by such reasons. Sometimes self-interested individuals virtue signal and rationalize in accordance with the proper moral reasons *simply because doing so improves their moral reputations*. Just like with markets, virtue signaling and rationalization need not be motivated by salient moral reasons to improve social and moral norms. Even though virtue signaling and rationalization must align with moral reasons for changing norms to count as moral progress, this could be incidental to the motives of the political actors just like market forces may incidentally hit upon innovations that align with moral reasons, e.g., consequential reasons whereby an innovation makes people's live better overall.

However, before we delve into how virtue signalers can unwittingly confer societal and moral benefits, despite their selfish motives, we

must understand how social norms work. To do that, we use work by Christina Bicchieri who defines social norms as,

> Rules of behavior to which individuals prefer to conform, but only on condition that they believe most people in their reference network conform to it, and that those same people believe that they ought to conform to it (2006).

People are more likely to conform to a social norm if they believe most people in their group conform to the norm (*the empirical condition*) and if most of them believe that everyone *ought* to conform to that norm (*the normative condition*). Robert is more likely to avoid tossing empty beer cans onto the sidewalk, and recycle them instead, if he believes that people within his reference network—those whose opinions he weighs when he makes decisions—do not toss their cans on the street, but recycle them, and expect others in the reference network to do likewise. If Robert believed that most people in his reference network believed everyone should recycle instead of littering, but the people in his reference network didn't act on that belief, then such knowledge wouldn't like be sufficient to change Robert's behavior due to moral mediocrity: people usually aim to be about as good and about as bad as their peers (Schwitzgebel, 2019).

In any case, Robert isn't likely to litter if he knows that his neighbors didn't and expected others not to either simply because doing so may hurt his reputation. Social norms are powerful tools to influence the behavior of others in one's reference network like family members, high-status individuals like celebrities. There are plenty of examples of social norms influencing behavior from wearing masks in public to paying income taxes—when the empirical and normative line up, social norms can effectively shape and influence behavior.

This brings us to the (limited) benefits of virtue signaling: there is a three-part explanation of how virtue signaling can potentially improve social and moral norms (Westra, 2021, p. 165). First, a group of sincere, but naïve, advocates disseminate a positive normative standard into the public discourse. Second, virtue signalers, spotting an opportunity to boost their reputations and social status, reinforce the new norm by praising those who follow it, and criticizing those who do not. And finally, support from virtue signalers conveys to a broader audience evidence of a social norm that they should follow to avoid sanction and

reputational damage since 'whatever else it does for the virtue signalers themselves, virtue signaling provides evidence about social norms' (Westra, 2021, p. 164—original emphasis).

Why do changes to social norms influence moral progress? The simple answer is that changing people's moral *beliefs* is insufficient to produce moral progress (Green, 2016). For example, even if everyone believes that laws against hiring women are morally dated, if they fail to convey such beliefs to their reference networks, then moral progress will stall. It is insufficient for moral progress simply to change the moral beliefs of even a sizeable subset of group members to enact changing social norms if they fail to act on or communicate those beliefs. Even if everyone in the group believes that it is morally wrong to exclude women from certain kinds of employment, they may still refrain from speaking up for fear of sanctions or costly criticisms since they do not know what everybody else believes. There is more to moral progress than a change in moral beliefs simply because such beliefs, even if true, may be too costly or socially unacceptable to act on or to express too openly. An example is offered by Westra:

> [Consider a world where] people privately believe that air travel has bad moral consequences. As we have seen, it might be very difficult for these people to act on those moral beliefs. They might feel social pressure to go on that vacation to Rome, or to attend a prestigious conference at Oxford. They might want to refuse to go on these trips, but they know that doing so would make them seem excessively moralistic and prudish. They might also miss out on important social and professional opportunities or damage their personal relationships. [...] Now let us return to that other imaginary world where the flight-shaming movement has been successful. [...] In this new normative environment, our reluctant flyers now have the social freedom to act on their moral convictions. *Where previously social pressures might have posed a barrier to moral action, now they facilitate it* (2021, p. 176—my emphasis).

Once we remove the social costs to folks taking moral reasons seriously, they are better placed to make moral progress stick. Even if the best moral reasons and arguments will rationally win out over time (FitzPatrick, 2015; Huemer, 2016), it doesn't follow this process will result in actual societal and moral progress. And *ceteris paribus*, if we

lower the cost of acting on moral reasons, this lowers the cost of acting morally. This oversimplifies matters a bit, but there should be no doubt that there are many robust moral reasons for people to behave better that they sometimes rationally ignore because the (social, practical) cost of acting on them is too high.

Here critics may worry that the motives underlying virtue signaling just encourages people to attach themselves to efforts at societal and moral progress to gain a reputation boost, thereby incentivizing social free-riders. This may then result in the collapse of the entire system if too many people virtue signal without acting on the virtues signaled. Notice that we find a similar issue in the marketplace too: people can sometimes be motivated by profit, but not so that it motivates them to provide the goods and services they claim. They may be motivated by profit to commit fraud, say. And yet this is not a widespread problem in the marketplace—it happens occasionally, of course, but more often than not defrauding people will eventually backfire. Just as people may attempt to free-ride on the efforts of sincere advocates to secure reputational and social status boosts without acting on the virtues signaled, some folks will attempt fraud in the market too. This is an issue any cooperative system, to the extent it is stable, must police.

One of the approaches that we have evolved to mitigate this mismatch between virtue signaling and behavior is the robust ability to spot liars and hypocrites. People like hypocrites even less than liars and will sometimes even absorb a loss simply to punish them, even in the absence of a social return on investment. Likely, this is because cooperation is key to human survival, and free riders, left unchecked, can undercut it. As moral psychologists have discovered:

> [People] dislike hypocrites more than direct liars because hypocrites falsely signal. One straightforward explanation for why hypocrites' false signals inspire moral outrage is that misleading other people is generally regarded as wrong [...]. A hypocrite's false signals may rouse further disapproval, moreover, because they lead to negative outcomes, such as unfairly boosting the hypocrite's reputation or shaming other people into changing their behavior while the hypocrite carries on (Jordan et. al., 2017, pp. 366-67).

Social tools like reputations and gossip help disincentivize folks from engaging hypocritical virtue signaling, since hypocrisy, once discovered, is often a death knell for one's reputation—to engage in phony virtue signaling, to boost one's reputation, will often backfire and result in the loss of that reputation when the hypocrisy is discovered. People are good at cheater detection (Lier et. al. 2013) and often refrain from doing things like hypocritically virtue signaling in order to preserve a good reputation (Vonasch et. al., 2018; Altay, Hacquin, and Mercier, 2020). There are thus some social assurances against hypocritical virtue signaling.

There is another worry lurking though. The account we sketched as to how it is virtue signaling can change norms for the better may make us worry that sometimes norms, even if they were for the better, would change *too quickly* so as to destabilize society. There is no doubt something to this—we don't want norms regularly changing overnight if only because people need time to adapt to change, even if it is positive—but we should keep in mind that there are checks against norms changing *too quickly*. Often it takes many people in a wide range of reference networks virtue signaling to change a norm across society. Virtue signaling to change social norms is a lot like voting and protesting: when enough people virtue signal, this can shift norms by conveying information to others about what members of their reference networks expect.

Could this approach to progress be improved? Perhaps. Jesse Hill and James Fanciullo criticize Evan Westra (2021) and Neil Levy (2021): they argue that virtue signaling is a poor approach to improving social and moral norms, and that norm signaling—where one signals a norm but without the motive to boosts one's reputations—is better at conveying evidence of social norms than virtue signaling (Hill and Fanciullo 2023). Norm signaling is better than virtue signaling in some respects like doing less damage to public confidence in moral discourse. However, these critics also ignore a key feature of virtue signaling that norm signaling lacks: robust motivation. The reputation motive is often needed to motivate individuals to convey information about social and moral norms, just like the profit motive is often needed to motivate folks to produce affordable quality goods and services. Just like virtue signaling, rationalizations can sometimes improve social and moral norms too. We turn to that next.

3 The benefits of political rationalization

There are many opportunities to rationalize our behavior in politics for the same reason we find a lot of virtue signaling in politics: the costs of epistemic irrationality to the individual voter or partisan is extremely low. Voters and partisans indulge beliefs untied to the facts or evidence since the political cost of it is extremely low. The same applies to rationalization.

Before getting into the benefits of rationalizations though, we should begin with an account of what rationalizations are. Rationalizations are false, but sincere explanations of behavior. They are false in the sense that they are not the reasons motivating the agent offering them. When Sheri explains that she didn't donate to the food drive because she believes that donations only encourage people to stay unemployed, she is offering a rationalization for her behavior that has several components: (a) a coherent, plausible explanation of her actions that (b) is unrelated to her action reasons for not giving, but that (c) she sincerely believes (Summers 2017: 98). The explanation is a coherent, plausible explanation if it could prima facie explain the behavior of the agent in question. And it is a rationalization, not a lie, if the explanation is sincerely held by the agent, but it fails to convey her actual reasons for action.

An example should clarify: Sheri claims, in explaining her actions to her friend Beth, the reason that she doesn't want to give money to Robert is because he is homeless, and she worries he will spend whatever she gives him on alcohol and illicit drugs. She could donate the money instead to an effective charity and help more people. Suppose, later, Sheri is approached by Beth with solid research showing that a local charity is effective at both reducing poverty and helping the poor find employment. Given the rationalization Sheri offered her friend, though sincere, she is under pressure to remain consistent and not to look flakey. Sheri may, of course, claim that she changed her mind and doesn't think that donating, even to an effective charity, is such a good idea. However, if she changes her rationale enough she won't look good to others, including her friend—she will be seen as indecisive and unprincipled (Everett et. al., 2016).

Rationalizations need not pass severe scrutiny of motivated critics to count as rationalizations, but they do need to rest on reasons that provide a coherent and passingly plausible explanation for the actions

of the agent. There are good reasons that rationalizations are especially common in the political domain. The first reason has been explored extensively throughout this paper: a sincere, but false explanation of an agents' behavior is unlikely to be *politically costly* for the individual. The second is that people are better placed, in the political domains, to pursue their self-interest at the expense of everyone else, since doing so isn't that costly to them personally. If Beth can get a boost to her reputation and social status by giving a clever rationalization for her actions, even if epistemically irrational, she is more likely to do so if the costs are low.

The temptation to rationalize one's behavior is particularly strong in politics because it tends to be zero-sum and tribal. A downside of a tribal dynamic is that we sometimes must ignore good reasons and evidence from those in opposing tribes to signal tribal loyalty (Marks et. al., 2019). This creates an incentive to rationalize actions in a fashion that makes one look good to others in their tribe, while also engaging in self-interested actions. There is a strong motivation to look good to others within our political tribe, but there is also a strong incentive to benefit as much as possible at little cost. The political domain is especially fertile ground for rationalization.

There's another aspect of rationalization salient to it serving as a means to producing social and moral progress: rationalizations are often designed to make use look good to others. It doesn't make sense to offer rationalizations that would make one look selfish: doing so would undercut one's moral reputation, and harm ties with group members. One rationalizes their behavior to look better to others than they would otherwise look. Circle back to Sheri: she want to excuse not giving to Robert by relying on a makeshift paternalistic rationalization for her refusal. If she didn't care how she looked to others, especially her friend Beth, she would have said she didn't want to help Robert because she would rather spend the money on herself. However, she didn't do that because, like most people, she cares how she looks to others (Jordan et. al., 2017). And since we are strongly motivated to be consistent in word and deed, we are under social pressure to act morally better than we would have otherwise if we offer to a rationalization that makes us look better to others (Campbell and Kumar, 2012).

Rationalization can morally improve behavior, though the costs of this practice also likely dwarf the benefits. Since we have strong incentives to both offer rationalizations that make us look good, and to

preserve our reputations, we will likely offer morally good rationalizations for our actions that, in the future, can be used by others to pressure us to be consistent and act on those reasons or to take a hit to our reputation. As Jesse Summers writes,

> Offering sincere justifications of one's actions can contribute to moral progress by creating pressure to become consistent with those justifications, and hence to become better over time, even when the rationalized justifications misrepresent one's motivation. This benefit explains a significant way in which rational moral progress is possible even when actors rationalize their own actions (2017, p. 101).

And:

> Rationalization can bring about moral progress, then, because I thereby sincerely endorse a justification with the claim that the justification explains the action. This prompts me to defend and be consistent with that justification, even if that justification is not in fact the best explanation of my action. And this consistency leads me to change over time to be more in line with the justifications I offer (2017, p. 103).

This is another example of the political invisible hand at work: people offer rationalizations for their selfish behavior to look better to others, especially when the costs are fairly low. However, despite their intention for offering the rationalization, consistency and reputational pressures can be brought to bear to force them to act better than they would otherwise. And since actions over time can influence how our self-image, and how we instinctively act, this process could lead, despite self-interested intent, to people acting morally better than they would have in the absence of their self-interested rationalizations. And when someone acts better than they would otherwise due to consistency and reputational pressures given the rationalizations they offer, they are implicitly engaged in the practice of signaling to their future selves that they are the kind of people who are morally good. Such actions do not merely signal to others that someone is altruistic or trustworthy, they also *self-signal to the one who performs the action*, thereby shaping their moral identity,

> [People] may lack perfect information as to their moral type, such that prosocial behavior may lead them to update their view of themselves: If I behaved prosocially, I must be a prosocial kind of person—someone for whom prosocial behavior provides greater utility—and therefore I will behave more prosocially in the future … [We] suggest that a person engaging in costly moral behavior is likely to infer that she is a moral person: 'Why else would I incur a cost to be moral unless I am that kind of person?' (Gneezy et. al., 2012, p. 180).

These processes, especially over time, can not only improve the surrounding social and moral norms, but they can also shape the moral identity of the one engage in offering rationalizations (and virtue signaling) originally motivated by self-interest (and selfishness). There is a potential objection here though. Suppose that Sheri will continuously double-down whenever pressed on the quality of her rationalization and devise *yet another rationalization* not to help others, even when her friend has shown her conclusive evidence that her rationalizations simply do not hold up to scrutiny. Presumably, it would be cheaper for her to simply to explain why even the local charity, despite the solid evidence showing that it is effective at helping people find work and housing, still wouldn't be a good enough cause. Here consistency may not be enough to force Sheri to morally improve her behavior such that is lines up with the rationalizations she offered, but instead to double-double on yet another rationalization for not giving. Call this *the double-down objection.*

There are a couple problems with the objection though. The first is that our thesis is not simply that the benefits of political rationalizations outweigh the costs—the costs, in many cases, dwarf the benefits—but simply to highlight the benefits. And second, this objection rests on the false assumption that the only pressure to act better is from factors like rational consistency, but it is from reputational pressure too. Suppose Beth keeps trying to explain away her inconsistency by giving further, distinct rationalizations for not donating to effective charities. Eventually people will realize she isn't interested in donating, and that she will continue making excuses to avoid looking bad despite her selfness. People are good at detecting cheaters and liars, and they do not appreciate it when others cheat to improve their moral reputations on the cheap (Jordan et. al., 2017, pp. 366-67). Sheri can keep making excuses for her selfish behavior, but beyond a certain point, she will likely end up damaging her reputation.

4 Conclusion

Political irrationality, while often negative, can confer benefits too, such as improving social and moral norms as a by-product of virtue signaling and rationalization. Those practices cannot be divorced from the benefits that result from those versions of epistemic irrationality, just as we cannot divorce the profit motive from the societal benefits arising from market society. The claim here is not that the benefits outweigh the costs—it is likely the opposite—but to establish that just like with the profit motive, there are benefits to epistemic irrationality too. Our purpose has been to highlight and defend those benefits. There is no denying that, despite appearances, epistemic irrationality sometimes has (indirect) social and moral benefits.

References

Altay, S., Hacquin, A.S., and Mercier, H. (2020). Why Do So Few People Share Fake News? It Hurts Their Reputation. *New Media & Society 24*(6), 1303-1324.

Barber, M., and Pope, J. C., (2019). Does Party Trump Ideology? Disentangling Party and Ideology in America. *American Political Science Review 113*(1), 38-54.

Bicchieri, C., (2006). *The Grammar of Society: The Nature and Dynamics of Social Norms*. Cambridge University Press.

Brennan, J., (2016). *Against Democracy*. Princeton University Press.

Brennan, G., & Lomasky, L., (2008). *Democracy and Decision: The Pure Theory of Electoral Preference*. Cambridge University Press.

Campbell, R., & d V. Kumar (2012). Moral Reasoning on the Ground. *Ethics 122*(2), 273–312.

Caplan, B., (2007). *The Myth of the Rational Voter: Why Democracies Choose Bad Policies*. Princeton University Press.

Cohen, G. L. (2003). Party Over Policy: The Dominating Impact of Group Influence on Political Beliefs. *Journal of Personality and Social Psychology 85*(5), 808–22

Downs, A., (1957). *An Economic Theory of Democracy*. Harper and Row.

Dusso, A. & Kennedy, S. S., (2015). Does Ignorance Matter? The Relative Importance of Civic Knowledge and the Human Tendency to Engage in Motivated Reasoning. *Journal of Public and Nonprofit Affairs 1*(1), 59-72

FitzPatrick, W. J., (2015). Debunking Evolutionary Debunking of Ethical Realism. *Philosophical Studies 172*(4), 883-904.

Freiman, C., (2021). *Why It's OK to Ignore Politics*. Routledge.

Fu, G., Heyman, G. D., Qian, M., Guo, T., and Lee, K. (2016). Young Children With a Positive Reputation to Maintain Are Less Likely to Cheat. *Developmental Science 19*(2), 275-83.

Funkhouser, E., (2022). A Tribal Mind: Beliefs that Signal Group Identity or Commitment. *Mind and Language 37*(3), 444-64.

Gelman, A., Silver, N., and Edlin, A., (2012). What Is the Probability Your Vote Will Make a Difference? *Economic Inquiry 50*(2), 321-6.

Gneezy, A., Imas, A., Brown, A., Nelson, L. D., Norton, M. I. (2012). Paying to Be Nice: Consistency and Costly Prosocial Behavior. Management Science 58 (1), 179-87

Geoffrey, M. F. (2007). Sexual Selection for Moral Virtues. *The Quarterly Review of Biology 82*(2), 97-125.

Green, D. (2016). *How Change Happens*. Oxford University Press.

Henrich, J., (2015). *The Secret of Our Success: How Culture Is Driving Human Evolution, Domesticating Our Species, and Making Us Smarter*. Princeton University Press.

Hill, J. & Fanciullo, J. (2023). What's Wrong with Virtue Signaling? *Synthese 201*(117).

Huemer, M., (2016). A Liberal Realist Answer to Debunking Skeptics: the Empirical Case for Realism. *Philosophical Studies 173*(7), 1983-2010.

Jones, P., and Dawson, P., (2008). How Much Do Voters Know? An Analysis of Motivation and Political Awareness. *Scottish Journal of Political Economy 55*(2), 123-42.

Jordan, J. J., Sommers, R., Bloom, P., and Rand, D. G., (2017). Why Do We Hate Hypocrites? Evidence for a Theory of False Signaling. *Psychological Science 28*(3), 356-68.

Kahan, D., (2016). The Politically Motivated Reasoning Paradigm, Part 1: What Politically Motivated Reasoning Is and How to Measure It. Robert Scott and Stephen Kosslyn (eds.) *Emerging Trends in the Social and Behavioral Sciences*. doi:10.1002/9781118900772.etrds0417.

Kunda, Z., (1990). The Case for Motivated Reasoning. *Psychological Bulletin 10*(3), 480-98.

Levy, N., (2021). Virtue Signaling is Virtuous. *Synthese 198*, 9545-62.

Lier, J. V., Revilin, R., Neys, W. D., (2013). Detecting Cheaters without Thinking: Testing the Automaticity of the Cheater Detection Module. *PLoS ONE 8*(1), 1-8.

MacAskill, W., (2015). *Doing Good Better: How Effective Altruism Can Help You Make a Difference*. Gotham Books.

Marks, J., Copland, E., Loh, E., Sunstein, C. R.,, and Sharot, T., (2019). Epistemic Spillovers: Learning Others' Political Views Reduces the Ability to Assess and Use Their Expertise in Nonpolitical Domains. *Cognition* (July, 188), 74-84.

Schwitzgebel, E., (2019). Aiming for Moral Mediocrity. *Res Philosophica 96*(3), 347-68.

Simler, K. & Hanson, R. (2018). *The Elephant in the Brain: Hidden Motives in Everyday Life*. Oxford University Press.

Smith, A., (1776/2003). *The Wealth of Nations*. Bantam Classics.

Somin, I., (2014). Why Political Ignorance Undermines the Wisdom of the Many. *Critical Review: A Journal of Politics and Society 26*(1-2), 151-69.

Summers, J. S., (2017). Rationalizing our Way into Moral Progress. *Ethical Theory and Moral Practice* (20), 93-104.

Tabor, C. S., & Lodge, M. (2006). Motivated Skepticism in the Evaluation of Political Beliefs. *American Journal of Political Science 50*(3), 755-69.

Tosi, J. & Warmke, B., (2020). *Grandstanding: The Use and Abuse of Moral Talk*. Oxford University Press.

Vonasch, A. J., Reynolds, T., Winegard, B. M., and Baumeister, R. F., (2018). Death Before Dishonor: Incurring Costs to Protect Moral Reputation. *Social Psychology and Personality Science 9*(5), 604-18.

Westra, E. (2021). Virtue Signaling and Moral Progress. *Philosophy and Public Affairs 49*(2), 156-78.

Williams, D. (2020). *Socially Adaptive Belief. Mind and Language 36*(1), 333-54.