Title: The moral source of collective irrationality during COVID-19 vaccination campaigns

Authors:

Corresponding author: Cristina Voinea, Research Center in Applied Ethics, Faculty of Philosophy, University of Bucharest,

Splaiul Independentei, no. 204, Bucharest, Romania

Email: cristina.voinea@man.ase.ro

Lavinia Marin, Ethics and Philosophy of Technology Section, TU Delft,

Delft, The Netherlands

Email: L.Marin@tudelft.nl

Constantin Vică, Research Center in Applied Ethics, Faculty of Philosophy, University of Bucharest,

Splaiul Independentei, no. 204, Bucharest, Romania

Email: constantin.vica@filosofie.unibuc.ro

The moral source of collective irrationality during COVID-19 vaccination campaigns

Abstract: Many hypotheses have been advanced to explain the collective irrationality of COVID-19 vaccine hesitancy, such as partisanship and ideology, exposure to misinformation and conspiracy theories or the effectiveness of public messaging. This paper presents a complementary explanation to existing epistemic hypotheses for understanding collective irrationality, focusing on the moral reasons underlying people's decisions regarding vaccination. We argue that the moralization of COVID-19 risk mitigation measures contributed to the polarization of groups along moral values, which ultimately led to the emergence of collective irrational behaviors. Collective irrationality arises from groups explicitly or implicitly endorsing values that ultimately harm both themselves and those around. The role of social media platforms in amplifying this polarization and contributing to the emergence of collective irrationality is also examined. Finally, potential strategies for addressing the moral sources of collective irrationality are discussed.

Keywords: collective irrationality, social media, vaccine hesitancy, COVID-19, moral reasons.

Many hypotheses have been advanced to explain large-scale COVID-19 vaccine hesitancy, such as partisanship and ideology, exposure to misinformation and conspiracy theories, especially on social media platforms, the effectiveness of public messaging, etc. (Dror et al., 2020; Gerretsen et al., 2021; Machingaidze & Wiysonge, 2021; Murphy et al., 2021). Most of the work that unpacks the motivations for irrational attitudes focuses on the epistemic grounds for individuals' refusal, such as conspiracy theory endorsing and epistemic vices, and on the role social media

platforms played in fostering these attitudes (Meyer et al., 2021; Nguyen, 2020; Rini, 2017; Wilson & Wiysonge, 2020). Collective irrationality is usually understood as the aggregation of reasonable microbehaviors into a macrobehavior that nobody would endorse (Schelling, 2006). Irrationality was a common accusation thrown left and right in mass media and policy discourse since the pandemic's beginning. Despite the growing body of work on the epistemic grounds of collective irrationality, explaining the COVID-19 vaccine hesitancy as caused by individual epistemic failures alone does not suffice. This paper advances a complementary explanation to the epistemic account of collective irrationality and explores the role played by moral reasons in people's decisions regarding vaccination.

We start by showing that the decision to get vaccinated or not for COVID-19 cannot be solely explained by differences in knowledge or information: people who refused to get the shot were not necessarily worse informed than people who accepted. This shows that there are more than just epistemic reasons at play in people's decision regarding vaccination. In the following section, we delve into the process by which COVID-19 vaccination campaigns became highly moralized. The moralization of health behaviors can have both positive and negative effects on public health outcomes. On the one hand, it can mobilize individuals to hold outliers accountable and suppress deviant behaviors (Kraaijeveld & Jamrozik, 2022). On the other hand, it can lead to stigmatization and polarization among those who do not comply with certain health norms (Minson & Monin, 2012). In the context of COVID-19 vaccination campaigns, we observe how different moral values are cited as reasons for or against getting vaccinated, leading to polarization of groups along moral lines. In the next section we explore the moral sources of collective irrationality. We argue that the polarization along moral lines between those willing and unwilling to get vaccinated led to the emergence of collective irrational behaviors. More precisely, some groups prioritized personal freedom over concern for others and this became an identity-defining norm that led to the refusal of the COVID-19 vaccine. This refusal, however, brings on health risks for both the individuals with this value-belief package and everyone else. This results in collective irrationality as the values held by these groups have damaging consequences not only for others, but also for themselves. Additionally, we examine the role of social media platforms in amplifying this polarization through the importance of signaling group identity and positions in highly moralized debates. Finally, we suggest potential strategies for addressing the moral sources of collective irrationality.

All about knowledge?

During COVID-19 vaccination campaigns, social divides between those willing and not willing to vaccinate deepened as accusations of irrationality were directed by both sides towards the other. At an individual level, irrationality is usually defined as resulting from holding beliefs that are ill-grounded–not supported by evidence (Bortolotti 2020). Collective irrationality arises in social situations where individuals act in a self-interested manner, which leads to an outcome that is collectively less than optimal (Schelling, 2006) because individual interest does not sum up to collective interest. For example, in the case of a fire, each individual's interest is to get as fast as possible to the exit, but if every individual follows their self-interest, this will lead to a stampede, and very few will succeed in exiting. Countries with low vaccination rates and high vaccine hesitancy apparently fit perfectly the definition of collective irrationality. In these countries, people made the self-interested decision not to get the shot, which led to a collective sub-optimal outcome: the virus continued spreading, exposing people to severe consequences of contracting it. But a standard epistemic definition of collective irrationality does not exhaustively explain widespread vaccine hesitancy because in this situation, the individual interest aligns with the collective interest from an epistemic point of view. That is, it is in the interest of individuals to get vaccinated, as the risks are minimal and the benefit, the protection against the disease, is maximal (as contrasted with other passive risk mitigation measures such as social distancing). And the more individuals vaccinated, the better the collective outcome–herd immunity and protection even for those that for some reason could not get the vaccine.

Maybe the reason for the social divide between those willing and unwilling to vaccinate is knowledge: even if it would have been in the individuals' best interest to take the shot, given the relatively low risks compared to the high benefits, it is possible that people did not have access to accurate information regarding vaccine safety, which lowered their willingness to actually take the shot. But empirical data does not confirm this hypothesis: not all those who hesitated to get the vaccine lacked access to relevant and accurate information (Stoica & Umbreș 2021; Tsang 2022; Goldstein et. al., 2015). As (Pfattheicher et. al., 2022) show, information about the efficiency of vaccines has to be supplemented by the receptors' empathy in order for it to

actually influence vaccine intention. Accurate information might be important, but by itself it is not enough to change people's behaviors; empathy is also necessary to increase vaccination intention (Korn et al., 2020). So, it seems that knowledge alone or, rather, lack of, does not predict vaccine hesitancy. Even more interesting is the fact that vaccine skeptics do not necessarily distrust medicine as a whole: they are very selective about the things that they doubt (Funkhouser, 2022). This just shows that taking a position on the adoption or rejection of risk-mitigation measures, including vaccination, was to a great degree a matter of signaling one's adoption of a value-belief package. In other words, one's position on vaccines and their acceptability was not only a reflection of one's epistemic beliefs, but also a reflection of one's moral stance. But how can a medical situation become so highly moralized, turning into an identity-defining issue and what does each position express from a moral point of view, more precisely? In the following section we explain how moralization of vaccination campaigns happened and how it led to polarization of groups along moral lines.

Moralizing COVID

Vaccination campaigns often elicit heated public debates, but COVID-19 vaccination campaigns and the decision to take or not to take the shot were even more moralized than usual (Bor et al., 2020, 2022; Wakefield & Khauser, 2021). This is due to the fact that the COVID-19 pandemic was colored with moral tones as almost every decision, at every level, involved a moral trade-off (Reimer et al., 2022). The increased salience of moral trade-offs inherent in options regarding responses to the pandemic, especially vaccination, determined people to make moral values definitory for their individual and collective identity. This led to polarization along moral lines, as people chose to prioritize some values instead of others, which in turn motivated people to pursue actions that otherwise may appear irrational. But polarization is not a homogenous phenomenon, and it can be motivated by different types of moral trade-offs and values. In what follows we present two mutually supporting explanations, focusing on the individual and collective levels, for how polarization emerges and spreads within communities. While the processes behind polarization we describe below are different, the effects are the same:

imposition among in-groups of a behavior, constitutive of membership to that group, and, implicitly, strengthening the divide between in-groups and out-groups.

On an individual level, moderately young and healthy individuals faced a decision with moral undertones, which involved weighing the risks from getting a new vaccine against a disease that was not seen as a threat to them against their duty to protect the most vulnerable members of society, including the elderly and those with immune problems and other medical conditions that put them at higher risk. In other words, individuals had to weigh in immediate self-interest and personal freedom against longer-term collective interest and duties towards others (Machingaidze & Wiysonge, 2021). Taking the shot was framed by those willing to do it as a moral imperative, an obligatory action whose violation is not only harmful to others but also immoral, as it doesn't respect the minimal ethical requirement to try and improve other people's situation when one can do so with minimal costs. In other words, failure to take the shot was interpreted by those willing to do it as a failure to prevent harm and suffering costs (Rosenfeld & Tomiyama, 2022). Furthermore, those unwilling to vaccinate placed more value on personal freedom, believing that the decision to get the shot or not is a matter of personal choice (motivated either by religious reasons or other beliefs regarding the purity of the human body or by group conformity) that is non-negotiable, even in a situation of global urgency (Bor et al., 2020). This shows that what was once a usually benign action, getting a vaccine, became a moralized issue as people attached moral values to the apparently neutral act of getting vaccinated (Graso et. al., 2021).

The moralization of COVID-19 risk mitigation measures led to an acute polarization of groups, as each side considered that the other is not just different, but plainly 'wrong'. Vaccination was thus not a matter of factual disagreement that could be tackled by correcting people's biases and injecting more information in the public domain but also a matter of moral disagreement, which is known to become a very divisive and hostile form of disagreement (Haidt et. al., 2003). Values are often seen as integral parts of one's identity and may be perceived as non-negotiable. In the context of pandemic responses, certain values may be perceived as being threatened by certain proposals, leading to polarization and conflict. This is because moral values are constitutive for who we are, thus we feel personally attacked when someone attacks one or more of our moral values. Both those willing and unwilling to vaccinate saw themselves as having the higher moral

ground. Moreover, one of the first global studies on the emergence of prejudice against the vaccinated and the unvaccinated during the COVID-19 pandemic (Bor et. al., 2022) shows that the fracture between vaccinated and unvaccinated spills into day-to-day interactions through the formation of prejudice against the out-group. Thus, while, on the one hand, those vaccinated or willing to vaccinate condemned those who refused to take the jab, the latter felt excluded, shamed, and treated paternalistically by the former (Bor et al., 2020; Rosenfeld & Tomiyama, 2022). This divisions, although expressed on social media platforms, bled into day-to-day life, informing people's attitudes towards others.

On a collective level, ending a pandemic, just as tackling climate change, is a collective action problem that requires large-scale cooperation (Sunstein, 2021). But cooperation, even in the face of existential threats, such as a pandemic, is not a piece of cake, as it requires individuals to bear a cost towards others' benefit or to contribute to the common good (Bavel et al., 2020). This type of trade-off is especially evident in the case of vaccines. It has long been recognized that vaccination is not only a self-interested action, but also a pro-social one (Bauchet. al., 2003; Böhm et. al., 2016; Giubilini, 2019). Obviously, by getting a vaccine one protects themselves against a disease, while also reducing the chances of further infecting others. Ideally, when a critical mass of people gets vaccinated, herd immunity is accomplished. But this also creates the perverse incentive to not get the vaccine, as one can still benefit from herd immunity even without getting the shot (James et al., 2021). Individuals can be free-riders, but their motivation to contribute to collective action increases, it seems, as they break the bounds of and extend moral consideration beyond their immediate in-groups. Sociologists have long established that feelings of connection towards strangers can motivate people to help even distant others in times of crises and even in cases where they have to sacrifice something for the strangers' well-being (Oliner, 1992). Even more interesting, when people feel a connection and identify with humanity at large, a phenomenon termed as "all humanity is my in-group" (McFarland et al., 2012) they show no in-group favoritism but, on the contrary, empathically extend moral consideration beyond one's family or nation and towards humanity at large (Barragan et al., 2021; McFarland et al., 2012, 2013). Barragan et. al. (2021) found that identifying with all humanity predicts cooperative health behaviors during COVID-19. More precisely, the feeling of having moral responsibilities and duties towards those who are not part of our immediate in-group, is an essential predictor of compliance with risk-mitigation measures and the desire to get vaccinated.

All of the studies quoted above uncover the importance of moral reasons in motivating people to act, even in times of crisis. If an individual identifies with a bounded social group, they can comply with the groups' norms and standards, even if this goes against their personal preferences. In this situation, individuals are less prone to extend moral consideration towards socially distant others.

Vaccination has been highly moralized, meaning that preferences regarding vaccination became universal and objectively true action-guiding principles that drove people to harshly judge and even punish those that disrespected their intuitions (Bor et al., 2020; Graso et. al., 2021). In the following section we explain how moral reason can play a role in the emergence of collective irrational behaviors in the case of COVID-19 vaccination campaigns. We show that collective irrationality arises when groups of people prioritize certain values over others, leading to a discounting of their duties and responsibilities towards others. This discounting can have negative consequences not only for others, but also for those who hold these values.

The moral sources of collective irrationality

Underlying moralization processes of COVID-19 vaccination campaigns is people's reliance on automatic moral intuitions. Automatic moral intuitions are the primary source of moral judgments, basically unconscious evaluative judgments about characters and actions, later followed by conscious deliberation (Graham et al., 2013). The Moral Foundation Theory categorizes six types of intuitions, called foundations, that underlie most of our moral judgments (Graham et al., 2011): Care/harm, Fairness/cheating, Loyalty/betrayal, Authority/subversion, Sanctity/degradation, Liberty/oppression. These moral intuitions have been shown to play a role in people's decision regarding vaccination: for example, (Amin et al., 2017) show that sanctity and liberty are strongly associated with vaccine hesitancy among parents and the same results were obtained by (Rossen et al., 2019). In the case of COVID-19 vaccination campaigns, hesitancy is closely associated with a high endorsement of liberty and purity and a low endorsement of authority (Schmidtke et al., 2022), while vaccine acceptance has been associated with fairness, that is the sense that one has a moral obligation to protect the most vulnerable

(Reimer et al., 2022). Thus, empathy and care for others predict a stronger willingness to vaccinate, while an over-valuation of personal freedom, alongside a belief in the purity of the human body (most probably informed by a distrust of medicine and strong religious beliefs) predicts lower intention for vaccination.

In the case of COVID-19 vaccination campaigns, collective irrationality can be explained in terms of people's favoring of some moral values instead of others. More precisely, those who refuse to take the vaccine discount the duties and responsibilities they have towards others. This is because they see some other values, such as personal freedom, resistance to authority or the purity of the human body, as more important, thus overriding the moral responsibilities for the well-being of other people. Holding these types of beliefs and discounting responsibility towards others has damaging consequences not only to other people, but also to those having these beliefs. This is a case where these values boomerang against those holding them: not getting the shot affects everybody, including the one that does not take the shot. And the more people hold these beliefs, the worse the outcome. Thus, collective irrationality arises from people failing to see that the values they prioritize have damaging consequences not only for other individuals, but also for themselves.

But this does not mean that every individual that finally took or refused to take the vaccine engaged in a process of moral reasoning that presupposed weighting in different types of values. Especially under conditions of informational overabundance and uncertainty, as was the case with COVID-19 vaccines, individuals are particularly prone to look at how others in their immediate social network are behaving, in order to both evaluate a certain behavior but also to gain new information about what the others in their social group approve or disapprove of (Horne, 2001; Horne & Johnson, 2021). In fact, it has been shown that health related behaviors, such as getting vaccinated, oftentimes depend on the perceived extent in which one's in-group also engages in them (Rabb et al., 2022; Schmelz & Bowles, 2021).

In groups, social norms serve the purpose of group-level self-regulation; norms prescribe which behaviors should be followed, regardless of individuals' preferences. Norms instill compliance even in cases where there is no possibility of punishment or withdrawal of esteem by group members. One of the reasons for social norms compliance, even in the case of a lack of social

control, is the perceived legitimacy of other people's normative expectations (Andrighetto et al., 2015; Bicchieri, 2017; Bicchieri & Xiao, 2009). Normative expectations are a subclass of beliefs concerning what individuals believe others should do (Bicchieri, 2017; Andrighetto et. al., 2015). According to Andrighetto, Grieco and Tummolini (2015), the perceived legitimacy of normative expectations provides a vital reason to motivate people to live up to the expectations of in-group members, irrespective of being watched. Thus, beliefs regarding how others ought to act have a significant motivational power that can explain why individuals in a certain group conform to a social norm although compliance might go against their immediate self-interest (Bicchieri et al., 2021). Even if prioritizing personal freedom over responsibility for others' well-being is not a conscious decision but is a result of giving in to group pressure, it is still irrational in virtue of the consequences that it brings about for everybody, including the person having those beliefs. On the contrary, when getting the shot is the result of caving in to or being influenced by one's group, the consequences do not backfire, as both the individual and the others are protected against developing serious forms of the disease. To put it simply, irrationality stems from explicitly or implicitly endorsing values that harm both the believer and others.

We live in a world where the size of our social networks, alongside high social mobility, greatly overcome our capacity of tracking those we can trust. Nowadays groups are constituted, especially online, around value commitments. This mean that "a specific group identity is the main relation among our social network rather than an intimate interpersonal relation" (Brady et al., 2020). This is why signaling mechanisms are more and more important in showing group membership (Marlowe et al., 2008). As Neil Levy argues, "costly and credibility enhancing signaling help fill the gap between reputation tracking and formal regulation" (Levy, 2021a). In other words, signaling one's adherence to a group becomes an important means of establishing trustworthiness. What is worrying is that when group identity is more important than individuals' identity, in-group members are prone to dehumanize socially distal others (Waytz & Epley, 2012). In the following section we show how this phenomenon manifested on social media platforms and how it contributed to strengthening the moralization of debates regarding COVID-19 vaccination campaigns.

Social media platforms as a battlefield for moral conflicts

Social media platforms increased the heat of public debates around the question "should one take the vaccine or not?" by charging them with morally loaded language (as it often happens on social media, see Brady et al., 2021) that turned this particular choice into a way of manifesting one's identity as allegiance to a group (Pascual-Ferrá et. al. 2021). The groups were not necessarily the large camp of anti-vaxxers or anti-lockdown protesters, any random group could have sufficed. For example, a group for manifesting allegiance to can be the group of co-workers, fans of an online game, or any political group. It does not follow from being left- or right-wing politically inclined that one would need to be for or against the COVID-19 vaccine, yet once the members of one's group converged toward a position around the vaccine, one needed to signal the membership by agreeing publicly with the others. In this section, we show how this specific kind of irrationality played out and was amplified on social media platforms.

We construct our argument on several interrelated claims. First, social media platforms make group identity the primary factor that one considers–Brady et al. (2020) call it *hyper-salience*– when posting or interacting with others on such platforms: "for some users social media serves as a constant reminder of our political group identities" (Brady et al., 2020, p. 984). Identity is incessantly performed on social media (Tolentino, 2020) as everything we do online communicates something about who we are. In addition, group membership becomes our primary source of identity when confronted with polarized debates. Social media platforms tend to restrict the level of visibility for users at the group or community level and polarize these groups by drawing hard lines around group identity. While we are aware that other groups are out there, we perceive them as polarized and exaggerated (e.g., we perceive their opinions to be more radical than ours).

Secondly, social media engagement promotes highly emotional debates, employing moralized language around polarizing issues that propagate messages through emotional contagion, achieving higher visibility than a neutral tone could do (Brady et al., 2020). Hence social media platforms are highly effective for in-group communication that solidifies the group allegiance and solidarity around shared moral norms. They also implicitly strengthen divides between in-groups and out-groups. In fact, it has been shown that out-group animosity is very sucessful in

driving engagement on social media (Rathje et. al., 2021). Social media does not have deterministic powers for causing people to rally around some shared moral norms, the situation at hand needed to be complemented by a certain lack of coherence in the public discourse. Social media platforms had such a strong effect because they acted in an informational ecology where mass-media channels were incoherent and fragmented. To argue for this, we compared two countries with highly different vaccination rates, Portugal and Romania. Just like everyone else in the EU, the Portuguese had access to social media platforms, yet the vaccination rate in Portugal reached a record high in Europe and worldwide (Mathieu et al., 2021). To explain the difference in vaccine hesitancy, Portugal led the vaccination campaign with a coherent public discourse, entirely depoliticized and it has been shown that by depoliticizing the COVID-19 discourse, citizens were more likely to trust their government's recommendations (Falkenbach & Willison, 2022). By contrast, in the case of Romania, in the absence of a coherent public discourse around the vaccine in the main-stream news, social media debates were the ones that steered the discussions and ultimately determined many Romanians' decision to (not) take the vaccine (Stoica & Umbreș, 2021).

There are empirical studies linking the intention to take the vaccine with the belief in the gravity of the pandemic: the stronger belief that COVID-19 was not a severe threat, the less inclined were people to take the vaccine (Hamel et al., 2021); other studies correlate the belief in conspiracy theories with the anti-vaccination stance (Yang et. al, 2021). Finally, studies link the propensity for analytic thinking with the willingness to take the vaccine (Anderson, 2016). The studies were done on different online populations and presented various competing explanations about the epistemic grounds for refusing to take the vaccine. Suppose we were to focus mainly on the epistemic agency of those refusing the vaccine. In that case, we would conclude that competing beliefs (gravity of the pandemic), competing worldviews (conspiracy-prone), or reliance on specific cognitive modes (such as system 1 or intuitive thinking) are at fault for deciding not to take the jab. However, we do not think the entire explanation lies in the epistemic realm. False claims about vaccines and the fake news around vaccination side-effects or the 5G chip conspiracy were very visible on social media and in the academic debates around the pandemic and the vaccines. Yet debunking a false claim is not enough to change the beliefs of those holding it, since often people will hold on to their beliefs even stronger after attempts to debunk it (Levy, 2017). This shows that, when the debate at hand is not oriented towards

epistemic values such as understanding or knowledge, epistemic arguments about the truth or falsehood of a proposition are mere ammunition in a morally-charged debate. Truth and falsehood claims come only as an afterthought after the actors have already decided the best course of action. This means that we need to make room for a normative interpretation of the debate at hand, side-stepping epistemic issues. We do acknowledge that truth matters and that throwing around false claims on social media or in the news did have devastating effects for the undecided, however we maintain that the epistemic-inclined analysis is only half the picture. In addition to the normative arguments and the morally-loaded language, we need to understand fully why people cling to their vaccine hesitant stance in the case of COVID-19 debates.

Previously, we examined how the collective irrationality of widespread vaccine hesitancy can be traced back to the polarization of groups along moral lines. Specifically, we observed that collective irrational behaviors emerge when group identity takes precedence over public moral commitments, such as concern and responsibilities towards those outside of the immediate in-group. But people often have a basic level of public moral commitments to their country, fellow citizens, and humanity in general. However, it can be difficult to effectively communicate and showcase these commitments to others, especially when discussions about group identity dominate social media. To address this issue, we need both individual qualities or virtues that allow us to prioritize public commitments and a communications infrastructure that helps make these commitments visible to others. While social media can serve as this infrastructure, it does not have to be the only option. The following section offers suggestions for how to foster public moral commitments that focus on our shared membership in humanity.

From virtuous citizens to vicious netizens: where to look for solutions

Online, where communication becomes impersonal and we are lost in and exposed to millions of other people, our identity is mainly defined by the groups we belong to (Brady et al., 2021). In this respect, one might argue that social media platforms are, in essence, not too different from modern nation-states, which also have to deal with individuals' different interests and conflicting moral commitments. How should we deal with this preference for in-group allegiance when the stakes are high and demand a public commitment? In this section, we examine three possible

approaches when tackling the problem of public moral commitments. Given that public commitments are part of political life, we look toward political models that might help us. We have identified the model of the multiethnic state, the communitarian model of civic virtues, and the liberal model of the free-market. Below, we explore what these models can teach us about how we should steer people toward extending moral consideration towards out-groups as well.

Modern multi-ethnic and multicultural states, like the USA, Canada, Belgium, former Yugoslavia, and Czechoslovakia, faced the same problem as social media today: a single social and political platform for people, that is, a single state, and many conflicting moral values. However, there is a significant difference between nation-states and social media platforms: a modern state has the duty to create the conditions of possibility for citizens to exercise their rights while at the same time creating the conditions for cooperation and collective identity. While states assumed neutrality regarding personal moral values, they nonetheless focused on fostering individual development and civic virtues, such as tolerance and patriotism or equal rights. But lines were also drawn in the form of imposed political limits to moral particularism, such as punishment of hate speech. Undoubtedly, there have been and still are nationalist tendencies and discrimination against minorities in some of these (former) states. However, social peace and harmony were seen as the fundamental public good without which the state and its society could not stand the test of time. Instead, social media platforms have no explicit duty to inspire cooperation between individuals and no need to articulate the framework of a fulfilled common life. On social media, we are not entitled to genuine civic rights because we are at best their (digital) immigrants, not citizens.

Another closely related solution to the multi-ethic state, can be inspired by ideas from communitarian, deliberative democracy and even classic liberal perspectives, following the classical tradition of Plato and Aristotle. Within this paradigm, we can redirect individual epistemic vices toward a public good by catalyzing moral virtues at a state-community level as public or civic virtues. Democracy is a rational political order based on collective intentions and thus normatively binding. In other words, democracy is the exercise of "distributed collective intelligence of the people" (Landemore 2011, 6), in which institutions must be able to gather and process information (Landemore 2011, 10; Vică 2015, 175). When epistemic vices impair judgment, moral commitments should be backed by public virtues. For Walzer, civic virtues are

"the moral and political qualities that make a good citizen" (1974, 593). Examples are moral respect for everybody (Edenberg, 2021, 273), civility (Walzer, 1974), tolerance, commitment to reciprocity, and civic-mindedness (McCabe, 1998, 558). It is debatable if public virtues are collective virtues or a superior, universalistic expression of individual moral virtues (Cordell, 2017).

Nonetheless, one thing social media can learn from the history of multicultural/multiethnic societies or the communitarian and deliberative models of democracy is the power of public or civic virtues to counter private vices. But in order to do this, social media platforms must promote social peace as a public good. Yet platforms are not incentivized to promote social harmony due to their business models, based on harvesting personal data and targeted ads. First and foremost, platforms are on the side of those who pay, whether we are talking about users interested in promoting their ideas, products, or services or marketers interested in user profiles and data. The more you pay, the louder your voice and the larger your audience. Moreover, on social media platforms individuals' behavior "is dominated by additional social motives, especially the desire to maintain or enhance their social status in relation to a specific group identity" (Brady et al., 2021). This means that group-identity motivations are especially salient on social media platforms, while people's identity is less important than group identity. This only reinforces the differences between in-group and out-group, thus satisfying people's need to belong and maintain a positive group image (Van Bavel & Pereira, 2018). Moreover, platforms encourage the expression of personal opinions and experience, leading to a cacophony of voices that only further hampers individuals' capacity for rational deliberation. When confronted with never-ending information streams, people "will tend to take shortcuts to relieve the burden of complexity" (Voinea et al., 2020). The most common shortcut is to rely on other people's reasoning and judgment (Horne, 2001; Horne & Johnson, 2021). In other words, platforms do too little to foster bridging instead of bonding.

From a liberal perspective, private moral vices can produce public benefits if there is a mechanism of aggregating them to work in favor of everybody. One solution to direct moral vices, such as egoism, toward the public good is put forth by classical thinkers, such as Bernard Mandeville or Adam Smith. The free market was the medium for this sublimation of vices into public benefits because the market is a cooperative game that has a signaling system based on

prices and a trust system based on contracts and reputation. However, online social platforms are two or multi-sided markets, but not free markets; they lack genuine transparent signals. In these digital ecosystems, algorithms govern both the transmission and visibility of information (Gillespie 2014). Algorithms are tweaked for engagement or, to put it more simply, they rank and make more visible precisely that information that "grabs more eyeballs" (Williams, 2018), even if that information might actually be disinformation, used for purposes of manipulation and radicalization (Benkler et. al., 2018). In other words, algorithms are not neutral and they do not push the interests of individuals users. On the contrary, algorithms are mainly designed with the purpose of maximizing engagement, even if that is contrary to the immediate or long-term interests of both individuals and society. In this sense, the algorithms governing social media platforms distort the epistemic and moral environment users engage with by pushing, promoting and making more visible the content that will most likely keep users online, which happens to be content that scandalizes and engages people in emotional responses. This removes the possibility of an emerging 'invisible hand' able to nudge 'cognitively vicious' users into producing public benefits, as algorithms actively participate in the co-construction of online sociality and, implicitly, of cognitive vices and less frequently, virtues. Thus, the problem of 'bad beliefs' on social media platforms, such as conspiracy theories regarding the origins of the virus or the efficiency of vaccines which contributed to strengthening vaccine hesitancy, is not exclusively an individual problem, but a 'structural' one, heavily influenced by the epistemic environment users engage with online. As Levy so pertinently shows (2021b), social media platforms create a medium where unreliable higher-order evidence is abundant; and the solution to this situation that leads people to have bad beliefs lies in "managing the epistemic environment"

Platforms are not neutral, they embed and promote, through their affordances, specific values. If so, then it is easy, in principle at least, for them to nudge, at least by design, the expression of public, and civic virtues. However, implementing any of these political models for social media is problematic given that the platforms, by themselves, have no incentive to change. Hence some amount of political intervention is needed to make them change. Social media platforms operate in a free-market logic of maximizing their profit by enticing users to spend as much time and attention in their realm. What needs to change in order for social media platforms to foster public and civic virtues is precisely the business model that underlies these platforms and that incentivizes them to use users' personal data to stir and polarize them.

Conclusions

Our identities are formed by the moral commitments we have towards the groups we identify with, and our daily interactions are governed by the normative expectations we have of one another. The pandemic has disrupted our ability to fulfill these moral commitments and meet the normative expectations of others, such as being present for loved ones who are sick or dying, attending funerals, supporting the most vulnerable members of our communities, and welcoming new babies into the world. The pandemic and related risk-mitigation measures, as well as the accompanying atmosphere of fear and anxiety, have significantly altered people's moral lives (Usher et al., 2020). In this uncertain and fearful context, people have rallied around values that have been cited as justification for or against certain pandemic responses. In this context, moral values have served as a primary driver of polarization and the emergence of collective irrational behaviors.

This paper explored the moral sources of collective irrational behaviors in the case of COVID-19 vaccination campaigns. The main claim of the paper is that epistemic reasons alone cannot exhaust people's reasons for refusing to take the shot. We showed that COVID-19 risk mitigation measures, including vaccination campaigns, were highly moralized. While moralization of health-related behaviors can be a force for positive change, it can also lead to stigmatization of deviant behaviors. This moralization of decision regarding vaccination led to an acute polarization of groups along moral lines, which ultimately gave rise to collective irrational behaviors. We explained collective irrationality as the prioritization by groups of people of personal freedom over concern for others, leading to the refusal of the COVID-19 vaccine despite the resulting health risks for both themselves and others. Additionally, we discuss the role of social media platforms in strengthening this polarization through the importance of signaling group identity and positions in highly moralized debates. Finally, we considered potential solutions for addressing the moral sources of collective irrationality which point towards the fact that collective irrational behaviors cannot be tackled only through epistemic measures. We showed that when people divide along moral lines and feel they have the higher

moral ground, the only possible solution is to find ways to harness the power of civic virtues to counter private vices.

References

Amin, A. B., R. A. Bednarczyk, C. E. Ray, K. J. Melchiori, J. Graham, J. R. Huntsinger, and S. B. Omer. (2017). Association of Moral Values with Vaccine Hesitancy. *Nature Human Behaviour* 1 (12): 873–80. https://doi.org/10.1038/s41562-017-0256-5.

Anderson, D. (2016). Analytic Thinking Predicts Vaccine Endorsement: Linking Cognitive Style and Affective Orientation Toward Childhood Vaccination. *University Honors Theses. Paper 215*. https://doi.org/10.15760/honors.220

Andrighetto, G., Grieco, D., & Tummolini, L. (2015). Perceived legitimacy of normative expectations motivates compliance with social norms when nobody is watching. *Frontiers in Psychology*, *6*. https://www.frontiersin.org/article/10.3389/fpsyg.2015.01413

Barragan, R. C., Oliveira, N., Khalvati, K., Brooks, R., Reinecke, K., Rao, R. P. N., & Meltzoff, A. N. (2021). Identifying with all humanity predicts cooperative health behaviors and helpful responding during COVID-19. *PLOS ONE*, *16*(3), e0248234. https://doi.org/10.1371/journal.pone.0248234

Bauch, C. T., Galvani, A. P., & Earn, D. J. (2003). Group interest versus self-interest in smallpox vaccination policy. *Proceedings of the National Academy of Sciences*, *100*(18), 10564–10567.

Bavel, J. J. Van, K. Baicker, P. S. Boggio, V. Capraro, A. Cichocka, M. Cikara, M. J. Crockett, et al. (2020). Using Social and Behavioural Science to Support COVID-19 Pandemic Response. *Nature Human Behaviour* 4 (5): 460–71. https://doi.org/10.1038/s41562-020-0884-z.

Benkler Y., F. Robert, and R. Hal. (2018). *Network Propaganda. Manipulation, Disinformation, and Radicalization in American Politics*. Oxford: Oxford University Press.

Bicchieri, C. (2017). *Norms in the Wild: How to Diagnose, Measure, and Change Social Norms*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780190622046.001.0001

Bicchieri, C., & Xiao, E. (2009). Do the right thing: But only if others do so. *Journal of Behavioral Decision Making*, *22*(2), 191–208. https://doi.org/10.1002/bdm.621

Bicchieri, C., Fatas, E., Aldama, A., Casas, A., Deshpande, I., Lauro, M., Parilli, C., Spohn, M., Pereira, P., & Wen, R. (2021). In science we (should) trust: Expectations and compliance across nine countries during the COVID-19 pandemic. *PLOS ONE*, *16*(6), e0252892. https://doi.org/10.1371/journal.pone.0252892

Böhm, R., Betsch, C., & Korn, L. (2016). Selfish-rational non-vaccination: Experimental evidence from an interactive vaccination game. *Journal of Economic Behavior & Organization*, *131*, 183–195. https://doi.org/10.1016/j.jebo.2015.11.008

Bor, A., Jørgensen, F. J., & Petersen, M. B. (2022). Prejudice Against the Vaccinated and the Unvaccinated During the COVID-19 Pandemic: A Global Conjoint Experiment. *PsyArXiv*. https://doi.org/10.31234/osf.io/t2g45

Bor, A., Jørgensen, F. J., Lindholt, M. F., & Petersen, M. B. (2020). Moralizing the COVID-19 Pandemic: Self-Interest Predicts Moral Condemnation of Other's Compliance, Distancing and Vaccination. *PsyArXiv*. https://doi.org/10.31234/osf.io/3rczg

Brady, W. J., Crockett, M. J., & Van Bavel, J. J. (2020). The MAD Model of Moral Contagion: The Role of Motivation, Attention, and Design in the Spread of Moralized Content Online. *Perspectives on Psychological Science*, *15*(4), 978–1010. https://doi.org/10.1177/1745691620917336

Cordell, S. (2017). Group Virtues: No Great Leap Forward with Collectivism. *Res Publica* 23 (1): 43–59. https://doi.org/10.1007/s11158-015-9317-7.

Dror, A. A., Eisenbach, N., Taiber, S., Morozov, N. G., Mizrachi, M., Zigron, A., Srouji, S., & Sela, E. (2020). Vaccine hesitancy: The next challenge in the fight against COVID-19.

European Journal of Epidemiology, *35*(8), 775–779. https://doi.org/10.1007/s10654-020-00671-y

Edenberg, E. (2021). The Problem with Disagreement on Social Media: Moral not Epistemic. In E. Edenberg & M. Hannon (Eds.), *Political epistemology* (1st ed., pp. 259–279). Oxford University Press.

Falkenbach, M., & Willison, C. (2022). Resources or Trust: What matters more in the vaccination strategies of high-income liberal democracies? *Health Policy and Technology*, 100618. https://doi.org/10.1016/j.hlpt.2022.100618

Funkhouser, Eric. (2022). Dangerous Beliefs, Effective Signals. *Philosophical Psychology*, July. https://www.tandfonline.com/doi/full/10.1080/09515089.2022.2101444.

Gerretsen, P., Kim, J., Caravaggio, F., Quilty, L., Sanches, M., Wells, S., Brown, E. E., Agic, B., Pollock, B. G., & Graff-Guerrero, A. (2021). Individual determinants of COVID-19 vaccine hesitancy. *PLOS ONE*, *16*(11), e0258462. https://doi.org/10.1371/journal.pone.0258462

Gillespie, T. (2014). The Relevance of Algorithms. In *Media Technologies*, edited by Tarleton Gillespie, Pablo J. Boczkowski, and Kirsten A. Foot, pp. 167–94. The MIT Press.

Giubilini, A. (2019). Vaccination and Herd Immunity: Individual, Collective, and Institutional Responsibilities. In A. Giubilini (Ed.), *The Ethics of Vaccination* (pp. 29–58). Springer International Publishing. https://doi.org/10.1007/978-3-030-02068-2_2

Goldstein S., N. E. MacDonald, and S. Guirguis. (2015). Health Communication and Vaccine Hesitancy. *Vaccine* 33 (34): 4212–14.

Graham J., B. A. Nosek, J. Haidt, R. Iyer, S. Koleva, and P. H. Ditto. (2011). Mapping the Moral Domain. *Journal of Personality and Social Psychology* 101 (2): 366.

Graham J., J. Haidt, S. Koleva, M. Motyl, R. Iyer, S. P. Wojcik, and P. H. Ditto. (2013). "Moral Foundations Theory: The Pragmatic Validity of Moral Pluralism." In *Advances in Experimental Social Psychology*, edited by Patricia Devine and Ashby Plant, 47:55–130. Academic Press. https://doi.org/10.1016/B978-0-12-407236-7.00002-4.

Graso M., F. X. Chen, and T. Reynolds. (2021). Moralization of Covid-19 Health Response: Asymmetry in Tolerance for Human Costs. *Journal of Experimental Social Psychology* 93 (March): 104084. https://doi.org/10.1016/j.jesp.2020.104084.

Haidt J., E. Rosenberg, and H. Hom. (2003). Differentiating Diversities: Moral Diversity Is Not Like Other Kinds. *Journal of Applied Social Psychology* 33 (1): 1–36. https://doi.org/10.1111/j.1559-1816.2003.tb02071.x.

Hamel, L., Lopes, L., Kirzinger, A., Sparks, G., Stokes, M., & Brodie, M. (2021). KFF COVID-19 Vaccine Monitor: Media and Misinformation. https://www.kff.org/coronavirus-covid-19/poll-finding/kff-covid-19-vaccine-monitor-media-and-misinformation/

James, Erin K., Scott E. Bokemper, Alan S. Gerber, Saad B. Omer, and Gregory A. Huber. (2021). Persuasive Messaging to Increase COVID-19 Vaccine Uptake Intentions. *Vaccine* 39 (49): 7158–65. https://doi.org/10.1016/j.vaccine.2021.10.039.

Kraaijeveld, S.R., Jamrozik, E. (2022). Moralization and Mismoralization in Public Health. *Medicine, Health Care and* Philosophy, 25: 655–669. https://doi.org/10.1007/s11019-022-10103-1

Korn, L., Böhm, R., Meier, N. W., & Betsch, C. (2020). Vaccination as a social contract. *Proceedings of the National Academy of Sciences*, *117*(26), 14890–14899. https://doi.org/10.1073/pnas.1919666117

Landemore, H. E. (2011). Democratic Reason: The Mechanisms of Collective Intelligence in Politics. SSRN Scholarly Paper. Rochester, NY: Social Science Research Network, April 1, 2011. http://papers.ssrn.com/abstract=1845709.

Levy, N. (2017). The bad news about fake news. *Social epistemology review and reply collective*, *6*(8), 20-36.

Levy, N. (2021a). Virtue signalling is virtuous. *Synthese*, *198*(10), 9545–9562.

Levy, N. (2021b). Echoes of covid misinformation. *Philosophical Psychology*, DOI: 10.1080/09515089.2021.2009452

Machingaidze, S., & Wiysonge, C. S. (2021). Understanding COVID-19 vaccine hesitancy. *Nature Medicine*, *27*(8), 1338–1339. https://doi.org/10.1038/s41591-021-01459-7

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., & Henrich, J. (2008). More 'altruistic'punishment in larger societies. *Proceedings of the Royal Society B: Biological Sciences*, *275*(1634), 587–592.

Mathieu, E., Ritchie, H., Ortiz-Ospina, E., Roser, M., Hasell, J., Appel, C., ... & Rodés-Guirao, L. (2021). A global database of COVID-19 vaccinations. *Nature human behaviour*, *5*(7), 947-953.

McFarland, S., Brown, D., & Webb, M. (2013). Identification with all humanity as a moral concept and psychological construct. *Current Directions in Psychological Science*, *22*(3), 194–198.

McFarland, S., Webb, M., & Brown, D. (2012). All humanity is my ingroup: A measure and studies of identification with all humanity. *Journal of Personality and Social Psychology*, *103*(5), 830.

Meyer, M., Alfano, M., & De Bruin, B. (2021). Epistemic vice predicts acceptance of Covid-19 misinformation. *Episteme*, 1–22. doi:10.1017/epi.2021.18.

Minson, J. A and Monin, B. (2012). Do-Gooder Derogation: Disparaging Morally Motivated Minorities to Defuse Anticipated Reproach. *Social Psychological and Personality Science,* 3(2): *200–207.* doi:10.1177/1948550611415695

Murphy, J., Vallières, F., Bentall, R. P., Shevlin, M., McBride, O., Hartman, T. K., McKay, R., Bennett, K., Mason, L., & Gibson-Miller, J. (2021). Psychological characteristics associated with COVID-19 vaccine hesitancy and resistance in Ireland and the United Kingdom. *Nature Communications*, *12*(1), 1–15.

Nguyen, C. T. (2020). Echo chambers and epistemic bubbles. *Episteme*, *17*(2), 141–161. https://doi.org/10.1017/epi.2018.32

Oliner, S. P. (1992). *Altruistic personality: Rescuers of Jews in Nazi Europe*. Simon and Schuster.

Pascual-Ferrá, P., Alperstein, N., Barnett, D. J., & Rimal, R. N. (2021). Toxicity and verbal aggression on social media: Polarized discourse on wearing face masks during the COVID-19 pandemic. *Big Data & Society*, *8*(1). https://doi.org/10.1177/20539517211023533

Pfattheicher S., M. B. Petersen, and R. Böhm. (2022). Information about Herd Immunity through Vaccination and Empathy Promote COVID-19 Vaccination Intentions. *Health Psychology* 41: 85–93. https://doi.org/10.1037/hea0001096.

Rabb N., J. Bowers, D. Glick, K. H. Wilson, and D. Yokum. (2022). The Influence of Social Norms Varies with 'Others' Groups: Evidence from COVID-19 Vaccination Intentions.

Proceedings of the National Academy of Sciences 119 (29): e2118770119. https://doi.org/10.1073/pnas.2118770119.

Rathje, S., Van Bavel, J. J., & van der Linden, S. (2021). Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences of the United States of America*, *118*(26), e2024292118. https://doi.org/10.1073/pnas.2024292118

Reimer N.K., M. Atari, F. Karimi-Malekabadi, J. Trager, B. Kennedy, J. Graham, and M. Dehghani. (2022). Moral Values Predict County-Level COVID-19 Vaccination Rates in the United States. *The American Psychologist* 77 (6): 743–59. https://doi.org/10.1037/amp0001020.

Rini, R. (2017). Fake News and Partisan Epistemology. *Kennedy Institute of Ethics Journal*, *27*(S2), 43–64.

Rosenfeld, D. L., & Tomiyama, A. J. (2022). Jab my arm, not my morality: Perceived moral reproach as a barrier to COVID-19 vaccine uptake. *Social Science & Medicine*, *294*, 114699. https://doi.org/10.1016/j.socscimed.2022.114699

Rossen, I., M. J. Hurlstone, P. D. Dunlop, and C. Lawrence. (2019). Accepters, Fence Sitters, or Rejecters: Moral Profiles of Vaccination Attitudes. *Social Science & Medicine* 224 (March): 23–27. https://doi.org/10.1016/j.socscimed.2019.01.038.

Savulescu J., A. Giubilini, and M. Danchin. (2021). Global Ethical Considerations Regarding Mandatory Vaccination in Children. *The Journal of Pediatrics* 231 (April): 10–16. https://doi.org/10.1016/j.jpeds.2021.01.021.

Schelling, T. C. (2006). *Micromotives and Macrobehavior*. W. W. Norton & Company.

Schmelz, K., and S. Bowles. (2021). Overcoming COVID-19 Vaccination Resistance When Alternative Policies Affect the Dynamics of Conformism, Social Norms, and Crowding Out. *Proceedings of the National Academy of Sciences* 118 (25): e2104912118. https://doi.org/10.1073/pnas.2104912118.

Schmidtke K.A., L. Kudrna, A. Noufaily, N. Stallard, M. Skrybant, S. Russell, and A. Clarke. (2022). Evaluating the Relationship between Moral Values and Vaccine Hesitancy in Great Britain during the COVID-19 Pandemic: A Cross-Sectional Survey. *Social Science & Medicine* 308 (September): 115218. https://doi.org/10.1016/j.socscimed.2022.115218.

Shpall, S. (2014). Moral and Rational Commitment. *Philosophy and Phenomenological Research*, *88*(1), 146–172.

Stoica, C. A, and R. Umbreș. (2021). Suspicious Minds in Times of Crisis: Determinants of Romanians' Beliefs in COVID-19 Conspiracy Theories. *European Societies* 23 (sup1): S246–61. https://doi.org/10.1080/14616696.2020.1823450.

Sunstein, Cass R. 2021. *Averting Catastrophe: Decision Theory for COVID-19, Climate Change, and Potential Disasters of All Kinds*. NYU Press.

Tolentino, J. (2020). *Trick mirror: Reflections on self-delusion*. Random House Trade Paperbacks.

Tsang, S.J. (2022). Predicting COVID-19 Vaccine Hesitancy in Hong Kong: Vaccine Knowledge, Risks from Coronavirus, and Risks and Benefits of Vaccination. *Vaccine: X* 11: 100164.

Van Bavel, J. J., Pereira, A. (2018). The partisan brain: An identity-based model of political belief the role of identity in political belief. Trends in Cognitive Sciences, 22, 213–224.

Vică, C. (2015). Open Data, Ideology of Liberation, and Open Government. *Re-Thinking the Political in Contemporary Society: Globalization, Consumerism, Economic Efficiency*, 163.

Voinea, C., Vică, C., Mihailov, E., & Savulescu, J. (2020). The Internet as Cognitive Enhancement. *Science and Engineering Ethics*, *26*(4), 2345–2362. https://doi.org/10.1007/s11948-020-00210-8

Wakefield, J. R. H., & Khauser, A. (2021). Doing it for us: Community identification predicts willingness to receive a COVID-19 vaccination via perceived sense of duty to the community. *Journal of Community & Applied Social Psychology*, *31*(5), 603–614. https://doi.org/10.1002/casp.2542

Walzer, M. (1974). Civility and Civic Virtue in Contemporary America. *Social Research* 41(4): 593–611. https://www.jstor.org/stable/pdf/40970199.pdf.

Waytz, A., & Epley, N. (2012). Social connection enables dehumanization. *Journal of Experimental Social Psychology*, *48*(1), 70–76. https://doi.org/10.1016/j.jesp.2011.07.012

Williams, J. (2018). *Stand out of Our Light: Freedom and Resistance in the Attention Economy*. Cambridge: Cambridge University Press. https://doi.org/10.1017/9781108453004.

Wilson, S. L., & Wiysonge, C. (2020). Social media and vaccine hesitancy. *BMJ Global Health*, *5*(10), e004206. https://doi.org/10.1136/bmjgh-2020-004206

Yang, Z., Luo, X., & Jia, H. (2021). Is it all a conspiracy? Conspiracy theories and people's attitude to COVID-19 vaccination. *Vaccines*, *9*(10), 1051.