SOCIAL NORMS IN TRANSITION: IDENTITY, ETHICS, AND SOCIAL CHANGE

WILLIAM RICHARD O'SHEA

A DISSERTATION SUBMITTED TO THE FACULTY OF GRADUATES STUDIES IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR IN PHILOSOPHY

GRADUATE PROGRAM IN PHILOSOPHY

YORK UNIVERSITY

TORONTO, ONTARIO

AUGUST 2025

ABSTRACT

Social change is an incredibly important but difficult contemporary problem. The world stands to benefit from a vegetarian diet and lowering our carbon footprint, for example, but changing such behaviors, habits, or norms, is notoriously difficult. Even if we were to find out how to effectively change such behavior, how should we? Philosophers offers indispensable guidance on these matters. My dissertation contributes here, to the practice of social change, by offering conceptual clarity on how change occurs and providing ethical guidance on how change should occur.

A central claim of my dissertation is that change is, and sometimes should be, emotional or non-rational. If we treat agents solely as rational deliberators, then we will fail to understand why people behave the way they do. But if we can appreciate people's desire to belong, fit in, and validate their identity, then we can design truly persuasive interventions that will respond robustly to how people decide how to behave in the real world.

The dissertation begins with a conceptual analysis of social norms. The goal is to offer a concept of social norm that is useful in guiding interventions. I argue an empirically grounded, psychologically accurate accounts of social norms must speak to social identity theory. The analysis I put forward says people adhere to norms because they identify with social groups. This offers a framework for making novel predictions about how norms change, for instance, that people will adhere to norms to validate their social identity when there is intergroup conflict or outgroup threats.

Some non-rational means to persuade people might raise moral concerns. But should we persuade people only on rational grounds to not eat factory farm meat or eat meat more generally? I argue no; when it comes to getting people to eat less meat this is wrong for two reasons. One reason is it is likely to cause serious backlash, but the other is that moral judgements about what is acceptable to consume are systematically non-rational. If we want to persuade people to not consume meat, then we ought to pursue non-rational means such as undoing or supporting propaganda.

But to what extent can we use non-rational means to facilitate social change? Martha Nussbaum famously argues that shame is impermissible as a means of social change. Shame can fail to respect others' worth and bypass their rationality and agency. I show that there is an important exception to this: if the aim of shame is to include someone in moral persuasion it actually becomes a duty to shame someone, as it expresses a respect for someone by attempting to include them into the social practice of morality.

DEDICATION

For Dad.

Let this be a sign that your love of learning, curiosity, and discipline live on.

ACKNOWLEDGEMENTS

A massive thank you to Kristin Andrews for her grace and tireless effort in guiding me through the dissertation. Perhaps more importantly thank you to Kristin for inspiring me with her rare ability to combine intellectual rigor and clarity with a commitment to community and care.

Thank you to Regina Rini for so much support particularly towards the end of my doctoral studies and as my interests in ethics grew. It was exhilarating to be able to work through my ideas with you and watch them grow with your feedback.

Thank you to Mike Giudice for such friendly, calm, and continuous support from my entrance to the program all the way until the end.

To my cohort and colleagues as well as the faculty at the Department of Philosophy at York University: thank you for all the support, discussions, and encouragement. I especially want to thank the staff at the Department of Philosophy: Mike, Christine, Manpreet, Nisha, and Debbie – thanks for being so warm and kind to me and for all the help over the years.

To Mum and Louse: without your never-ending support I would never have come this far in philosophy.

To my partner, Lia: your encouragement, love, and discussions were welcome and surprising contributions which were essential in finishing my dissertation.

TABLE OF CONTENTS

Chapter 1 – Introduction to the Dissertation

Social change is a topic worthy of research more than ever. How can we ensure public and global health outcomes are met amongst technological advancements, climate change, and pandemics? How can we change norms to promote healthy and sustainable diets, develop healthy habits with technology, and learn to respect cultural differences in multicultural societies? These are questions that should concern us all, especially researchers of social, behavioral, and health sciences. A perennial worry is that despite the real-world importance of such topics, philosophers' arguments and conceptual analyses are too far removed and abstract to be of use.

The aim of my dissertation is to dispel this worry and show that philosophy is a crucial tool for thinking about how to change norms effectively and also how to change norms ethically. As a preview, I develop an account of social norms grounded in social identity theory – a robust and psychologically realistic interpretation of norms that should help researchers understand why people follow norms they do and offer insights into how to change norms. I then delve into ethical discussions surrounding how and whether we ought to change norms. I argue that there are limits to rational persuasion when it comes to changing dietary norms because the choice to eat meat is systematically non-rational. I also argue that shame can actually be a permissible way to change norms, but, only when it aims at rational persuasion because such an intent offers respect and dignity to those whose norms or behavior we want to change. I hope my dissertation offers some surprising conclusions and important contributions to philosophical discourse on changing social norms.

To set the stage it will help to understand how the topic of norm change piqued my interest. It all started with the research of philosopher Cristina Bicchieri, especially her more

recent conceptual analyses of social norms which were used to train aid foundations on how to change social norms. I believe Bicchieri offers an example of how philosophy can help solve real-world problems; her work should extinguish doubts about whether philosophy can be concrete and useful. The cases which occupy Bicchieri's work range from maladaptive social norms that interfere with public health like open defecation, female genital mutilation, to not wearing masks, as well as norms of distrust that lead to corruption and honor killings. Bicchieri's drive is admirable and offers a unique example of how philosophy can be concrete, practical, and make real-world differences.

Norms in the Wild by Cristina Bicchieri "stands to change of social practices as Che Guevara's Guerrilla warfare stands to political revolutions" says philosopher Ulf Hlobil (2017). I remember thinking after I first read the book was exciting and fresh as a work in philosophy, but also that it rarely responded to other philosophers who worked on norms. *Norms in the Wild* is part philosophy part social change practitioner's guidebook. The promise of Bicchieri's research, written on the back of *Norms in the Wild*, is to be able to "improve lives everywhere". These words were presented as a challenge to me: could we really apply this theory to changing *any* social norm? The answer I eventually landed on was no, there was a category limitation of Bicchieri's conceptual analysis of norms, namely, those kinds of social norms which people follow because they want to belong to and self-identify with a social group. To see how I arrived at this view requires appreciation of the questions which motivate Bicchieri's research over the last 25 years, and how Bicchieri answers these questions.

There are three questions that motivate Bicchieri's philosophical work. They are: What are social norms? How can we measure social norms? How can we change social norms? In 2010 Bicchieri had the chance to develop a project with UNICEF to train global aid workers

about her theory of social norms. The goal was to train aid practitioners to diagnose, measure, and target the above-mentioned norms for change, which was partially due to their previous resistance to past interventions. It is important to note that however one answers the question of what social norms are will influence the latter even if questions of change and measurement are not wholly dependent on one's conceptual analysis. If we can understand what the basis of Bicchieri's answer is to what norms are, then we can understand what its limitations are for understanding and changing norm-guided behavior.

Bicchieri's account is underpinned by rational choice theory – a theory which aims to explain human behavior with the assumption that people make decisions by weighing the costs and benefits of different options to maximize their self-interest. Bicchieri is especially influenced by David Lewis' use of rational choice theory in analyzing norms in *Convention*. It will help to start with Lewis' conceptual analysis of convention:

> *A regularity R in the behavior of members of a population P when they are agents in a recurrent situation S and S is a convention if and only if it is true that, and it is common knowledge in P that, in any instance of S among members of P:*
>
> *1. Everyone conforms to R*
>
> *2. Everyone expects everyone else to conform to R*

There are additional conditions in Lewis' account, but these two are the most important to understand what norms are and Lewis' influence on Bicchieri's account of norms. A central influence here is the rational choice framework. Recall that according to this framework individuals should behave in a way that maximizes their outcomes. This is important because it allows theorists like Bicchieri and Lewis to abstract from the empirical and psychological question of what motivates people to adhere to norms. For instance, the idea that someone would be motivated to behave in a way that does not benefit them such as donating blood or

volunteering is put to the side. Instead, it is assumed that people are motivated only by self-interest and what will lead to the best outcome for them. Bicchieri would argue that while this assumption abstracts or simplifies what actually motivates humans, it is an important assumption because it allows us to make testable models that predict the outcomes of individuals' interactions. This leads us to game theory – the study of mathematical models of strategic interaction among rational decision-makers. Game theory builds upon rational choice theory to analyze form models and make predictions about how people will behave given others' choices. These testable models are important for answering the kind of questions Bicchieri and I are interested in such as how to get people to cooperate with norms that promote public health. With the rational choice and game-theoretic frameworks in mind, we can now move to Bicchieri's first pass at an account of social norms.

Bicchieri's first pass at giving an account of social norms comes from *Rationality and Coordination* (1993):

> *Let R be a behavioral regularity in population P. Then more generally, R is a social norm iff R depends on the beliefs and preferences of the members of P in the following way:*
>
> 1. *Almost every member of P prefers to conform to R on the condition that almost everyone else conforms.*
> 2. *Almost every member of P believes that almost every other member of P conforms to R* (p. 252).

The similarity between the two accounts is striking, but there are important differences. One is that on Lewis' account for a norm to exist people are actually conforming, whereas on Bicchieri's view people *prefer* to conform given that people hold the above beliefs. The difference matters especially in cases where people are mistaken about what others actually do and yet still prefer to adhere to a norm despite being mistaken about how many people actually follow it. Notable examples of norms based off inaccurate perceptions of their prevalence

include alcohol consumption and energy usage – for instance, many college students might think that drinking alcohol is common and therefore a norm when it is in fact not very common. Bicchieri agrees with Lewis however that for a norm to exist it must be common knowledge that others conform to it and that one believes most others conform to it. This means the two accounts are nearly identical except for vocabulary such as common knowledge used in the case of Lewis and beliefs and preferences in the case of Bicchieri. Now, let us look at Bicchieri's most recent theory of social norms in *Norms in the Wild*:

> *A social norm is a rule of behavior such that individuals prefer to conform to it on the condition that they believe*:
>
> 1. *Most people in their reference network conform to it* (empirical expectation)
> 2. *Most people in the reference network believe they ought to conform to it* (normative expectation)

In this analysis Bicchieri now uses the language of beliefs and expectations while introducing two further elements. The first is that of a reference network, which is the collection of people whose expectations are relevant to adopting a norm. The second is that of normative language where people believe they "ought" to conform.

The reason that this evolution of the concept is interesting is because it shows the development of a theory that was concerned with one question and later another. Lewis' account was meant to answer one question, which influenced Bicchieri as a philosopher steeped in and intrigued by rational choice and game theory. The initial question that Bicchieri was interested in answering in 1993 was "how do we explain the fact that many people… engage in cooperation in situations in which it is a dominated strategy?" (Guala, 2017) Dominated strategy meaning there are other options that provide better results or better outcomes for an individual. In other words, why do people choose to cooperate even when there are other options that would lead to a better result for them? This is perplexing given the assumptions about the self-interested motivations of

humans that are central to rational choice and game theory. While some appeal to moral or altruistic motivations to answer this question, Bicchieri's insight is to suggest that in such conditions where cooperation is a dominated strategy it can still be rational to cooperate. Individuals should still prefer to cooperate on the condition that a social norm exists. This is because the presence of a social norm transforms the payoffs of a situation such that cooperation is no longer the dominated or suboptimal strategy. Take the examples of volunteering, tipping, or giving blood. These behaviors are individually costly and in terms of time, energy, or money, and are therefore suboptimal choices or dominated strategies. But if a social norm exists then the decision to cooperate with these norms becomes optimal given the presence of others' expectations. Instead, cooperation becomes the dominant strategy given the presence of others' expectations because failure to meet others' expectations will lead to being sanctioned or other sorts of costly consequences (Bicchieri, 2006).

We can now appreciate that while the framework of rational choice and game theory are features of Bicchieri's account, they are also bugs for the projects of figuring out how to change norms. This is because humans' motivations to adhere to norms are not only self-interested; people are also intrinsically motivated because they internalize norms as part of their self-image – an aspect of norm adherence I focus on in my dissertation. This shortcoming of Bicchieri's account of norms becomes clear when we ask: why do people follow what others do because they believe others expect them to? On this front Bicchieri remains vague. We could supplement Bicchieri's theory by suggesting that humans have an innate propensity to care about what people expect of us, but then this poses additional questions such as which people? Here Bicchieri's account of a reference network just delays again a satisfying answer: we know it is the sum of people who we care about, but why do we care about *them*? This vague motivational

basis of norm-abiding behavior and recognition that in the real world we are *selective* norm followers demands an answer that Bicchieri's theory does not possess. The limitations of rational-choice analyses of social norms are that not all norms are optimal choices or dominant strategies – for instance, driving on one side of the road or choosing to queue are something that one ought to do given what others do and that they expect they should. But some norms are quite clearly not about making optimal, strategic choices, but about being loyal to the group or fitting in with one's culture. For instance, Americans and Australians barbecue meat a lot whereas Hindus do not eat beef at all – these norms are culturally significant ways to bond with others and are representative of identities that people ascribe to. Thus, there are reasons for people to adhere to these norms of food consumption irrespective of whether they are optimal. This is something Bicchieri does not discuss because they are questions that exist outside of the choice theoretic framework – questions like whether people adhere to social norms because it is significant to them given how one categorizes oneself in terms of social groups.

We adhere to some norms because they part of who we are - our social self-understanding. The late philosopher Geoffrey Brennan, along with philosophers Lina Eriksson, Robert Goodwin, and Nicholas Southwood argue that their view of social norms, unlike Bicchieri's, can account for the fact that some norms are conformed to because they are meaningful or they "mean something to us", whether that meaning be an identity, role, or signal. Their view is that social norms are essentially clusters of normative attitudes plus knowledge of those attitudes (Brennan et al., 2013, p. 15). Notice that, like Bicchieri's latest definition of social norms in *Norms in the Wild*, normative commitments are given a central place in their account of norms:

A normative principle P is a norm within a group G if and only if:

> *(i) A significant proportion of the members of G have P-corresponding normative attitudes; and*
> *(ii)A significant proportion of the members of G know that a significant proportion of the members of G have P-corresponding normative attitudes (Brennan et al., 2013, p. 29).*

This account is quite close to Bicchieri's account except with the explicit focus on normative attitudes. Brennan and colleagues take normative attitudes to include the following: "(a) normative beliefs, judgements, and other cognitive states, (b) normative expectations, (c) reactive attitudes and dispositions to have such attitudes, and (d) any other attitudes that entail (a), (b), or (c)" (Brennan et al., 2013, p. 29). Bicchieri focuses instead on beliefs and preferences, which Brennan and colleagues take to be non-normative – however, the normative expectations of Bicchieri's account are clearly normative attitudes. What distinguishes the two views then is that Brennan and colleagues expand normative expectations to normative attitudes so as to include the normative force or content of these attitudes, e.g., not merely that one desires or believes something to be the case, but rather that norms involve explicitly normative attitudes such as clusters of approving, disapproving, obliging, sanctioning, and rewarding attitudes. But terminology aside, this does not seem a big threat to Bicchieri's account; rather a helpful amendment to recognize the varieties of normative attitudes which are not expectations. However, it is unclear how this account of norms sheds light on the connection between social identity and norm adherence, namely, this idea that norms are not contingent on what others do but how one understands oneself in terms of social groups. A possible answer from Brennan and colleagues is to point to their functionalist account of norms – that all norms serve the purpose of fostering accountability. This is also perhaps the biggest notable difference between Bicchieri and Brennan and colleagues. Let me return briefly to Bicchieri to see why she denies analyzing norms in terms of their functions, which will then let me come back to my question about identity and norms.

Bicchieri argues that functions are not essential to an analysis of norms and so abandons the idea altogether. One reason is because philosophers typically argue that social norms serve beneficial functions, e.g., they serve the purpose of maximizing welfare, facilitating coordination (Lewis, 1969), or making individuals accountable (Brennan et al., 2013). Bichierri argues that none of these functional accounts will suffice because some norms are not adaptive, serve no function, and are harmful – such as discriminatory norms against women or minorities (Bicchieri et al., 2023). So, Bichierri suggests we need to understand norms in terms of why individuals follow them and not norms in terms of some purpose they serve, and we also need to be weary that we do not believe all norms serve an adaptive purpose.

I think the point that some norms are harmful is clearly correct, but I do not think this fact and that some norms seem purposeless warrants a complete abandonment of the functionalist perspective. There is good reason to think that all of these can be true – individuals can take themselves to follow norms for certain reasons which serve important adaptive functions while the norm is harmful. The norms I have in mind are those which are personally meaningful for one's self-image, serve the purpose of fostering group identity, while also discriminating against others or being harmful in some other ways. Think about norms that prescribe eating factory farm meat to fit in for a certain group – these are harmful to one's health, the environment, and for factory farm animals, but they also serve an important bonding purpose for the social group which they are part of. Thus, Bicchieri focuses on individuals' cognitive representations of appropriate behavior, or in this case preferences, beliefs, and expectations. But this leaves out the question of how norms function at the group level and also how the group functions at the level of individuals' minds – both important aspects of norms for understanding the kinds of norms I just described. And it is entirely possible that these norms which are adaptive for some group-level purposes are

also the basis upon which people exhibit bad or dangerous behaviors – not wearing sunscreen, eating unhealthy food, not wearing a helmet, wearing a mask, and so on.

So, there are reasons to not discredit a functionalist understanding, which brings us back to Brennan and colleagues' view - that norms essentially serve the purpose of creating accountability between people. What norms facilitate is accountability, which creates expectations and demands for what kind of behaviors are appropriate and not. An account that deserves mention at this point is philosopher Margaret Gilbert's. Gilbert offers an account of convention, which she takes to be a species of norm, that focuses on the importance of joint commitments. Gilbert describes a convention as a collective commitment to jointly accept that a shared practice is to be conformed to. Since the convention is described as a collective commitment, the commitment puts a constraint on which actions are permissible for all parties involved. Gilbert offers the following analysis of convention:

> *A population P has a convention of conformity to some regularity in behavior R in situations of type S if and only if the members of P are jointly committed to accept as a body, with respect to themselves, the fiat: R is to be conformed to (Gilbert, 2008, p. 12).*

When a joint commitment is made, the parties involved owe it to each other to carry on adhering given that rescinding the commitment cannot happen unilaterally. This seems to mesh nicely with what Brennan and colleagues state which is that parties to a norm *collectively* form commitments and mutual accountability. They accept the fiat that 'this is what we ought to do around here' together.

Both Margaret Gilbert and Geoffrey Brennan and colleagues recognize the importance of the group for analyzing social action. According to the latter, norms are a vehicle for creating accountability to "galvanize meanings and identities, or to express shared values" (Brennan et al., 2013, p. 53). These accounts recognize that there are important group-level or social functions that

social norms help achieve. These social functions are important for understanding norm change; how norms persist, emerge, and fade. This is because norms can become constitutive of identities and roles such that what people believe they should do, what norms they think they should follow, are a function of these roles and identities. Trying to change a norm when it constitutes a person's identity or role will be difficult, just as packaging a norm as part of a new role or identity may help catalyze change. Notice that these accounts offer something which rational choice theoretic accounts cannot. When norms are constitutive of identities or roles, we can make certain predictions about what norms people will follow based on knowledge of these roles and identities. But Lewis and Bicchieri take these things for granted and analyze norm adherence based on preferences that bottom out at self-interest. When people adhere to norms which they've internalized, however, what is rational to do is not based on self-interested preferences but based on who you are, how you identify, and what roles you occupy. These are determinants of social action, but do not fit in well with the foundational assumptions of rational choice theory.

While Gilbert and Brennan and colleagues offer important lessons for group-based norms and norm change, ultimately, they lack a focus on this crucial connection between an individual's self-image and social identity. Both accounts focus on why one ought to adhere to a norm when they know others are committed or when they commit with others. But norm adherence is not solely about what others do or what normative attitudes others have; norms are also personal for individuals. They give individuals a sense of meaning such as fitting in, belonging, or constructing a valuable self-image. The shortcomings of the aforementioned accounts therefore are that they do not make conceptual space to make sense of this empirical commonality, which is that adherence is both personal and social - personal because it is one's self-image and social because they are group-level facts about the group or norm which one self-categorizes as.

There is a wealth of evidence in social psychology that attests to the idea that people adhere to norms because of their social identities. Henri Tajfel, social psychologist and creator of social identity theory argued that a social identity is "part of an individual's self-concept which derives from his knowledge of his membership of a social group (or groups) together with the value and emotional significance attached to that membership" (1979, p. 255). Tajfel's theory is significant for thinking about norms in at least two ways. The first is that it fills a gap in research on norms about the significance of affect and emotion. We know anecdotally that people feel strongly about some of the norms they follow - some motorcyclists will never wear helmets but wear their tattoos with a keen sense of pride. The norms we adhere to are infused with personal attachments and connecting norm adherence to social identity helps explain this. The second way is that it opens space to make sense of the way that social norms can be both personal and social, which was the shortcoming of previous accounts. People adhere to norms because they identify and self-categorize with a social group. This is important because it is not solely according to what others expect or know or do that motivates adherence - as Bicchieri, Brennan and colleagues, and their predecessors on convention would have it. Rather, it is self-knowledge or a self-image and its ties to social groups that often motivates people to adhere to norms. This is where the focus of my interventions lies: to understand how to change real-world norms where identities play such an important role what norms are, how they change, and how to change them.

What excited me about Bicchieri's research was that it showed how the tools of philosophical analysis could be of use to those 'on the ground' who are engaged in changing harmful social norms. Analyzing and refining the concepts we use to think about social norms or social behaviors more broadly is a particularly important task for a time where deliberately changing norms is important for so many reasons. Climate change creates a need for people to

adopt sustainable norms like using less energy, driving and flying less. Public health in the Global South can stand to benefit from more communities adopting sanitary norms while many communities in the Global North stand to benefit from creating norms around healthy lifestyle choices like diets, exercise, technology usage, and more. My dissertation takes inspiration from Bicchieri and builds on her discourse about what social norms are. An important aspect of norms is they are tied to social identities and navigating these identities are crucial to real-world projects of norm change. And while philosophers can help develop concepts to measure and diagnose when norms exist and how to change them, another important set of questions that philosophers are poised to answer beyond how to change norms effectively, is the ethical question of how and whether we *should* change norms. In what follows, I discuss in more detail my philosophical contributions to the psychology and ethics of norm change.

1.1 Social identity norms

In the first chapter of my dissertation, I offer a conceptual analysis of norms tied to social identities, or simply, social identity norms. I argue that there are several kinds of social norms, and in particular argue that *social identity norms* are an important kind if our concern is psychologically realistic interpretations of why people follow norms. As mentioned, social identity norms do not fit well into existing categories of norms because they are both personal and social. They are personally meaningful because people self-identify or categorize according to these social groups but are clearly social because it is the dynamics of groups at an interpersonal level which affect what is internalized. The central claim is that to explain why people follow social norms, and how to change them, we must attend not merely to external expectations or coordination dynamics, but to agents' desire to self-categorize in terms of the groups they belong to and the roles they play within these groups.

The concept of a social identity norm is distinct from both social and descriptive norms in prevailing accounts like Cristina Bicchieri's. Identity-based norm adherence involves a relation between individuals and their groups. People self-categorize in terms of social groups for several reasons. The most basic one is to satisfy the fundamental need to belong. More sophisticated desires include emotional attachments to the social group or role such as they feel proud or purposeful. In intergroup contexts where people feel their group identity is threatened, such as during immigration, governmental, or other external influences, they will adhere to norms as a way to avoid the group's dissolution.

These insights about why people adhere to norms are clearly important for thinking about norm change. Most interventions to change norms focus solely on the individual and what they think about others or how they come to decide what norms to follow. But my perspective shows that the group matters just as much. If desired change, say getting people to wear masks or eating less meat, is perceived as a threat to the existence or identity of a group, then change will be much more difficult. Conversely, simply labelling a norm as part of a social identity can encourage or dissuade people to adhere to norms as a sign of membership. Think of labels like a voter, a litterbug, or a vegetarian. People wear these labels with pride or disdain, and this can encourage norm change as well.

1.2 Propaganda and persuading people to eat less meat

In the second chapter of my dissertation, I take this insight that some norms are adhered to because of social identities and apply it to a case of norm change: getting people to eat less meat. I argue that in some cases people eat meat because that is what one ought to do as a member of a specific social group. Social identities can operate at different levels of social groups such as at the national level, or more locally in one's municipality, community, and of

course family. Eating meat is something one ought to do at each of these levels of social identities. Think of national dishes such as Kiwi's meat pies or chicharron in Latin America or a Brazilian grandma's Arroz con Pollo. These are not simply norms that people follow because others are doing it, but because they understand themselves in terms of these norms. This means persuading people to not eat meat will require careful navigation of these identities.

Once we have this social identity norm lens to understand why people eat meat, it should become apparent that persuading people this is wrong will likely cause backlash. This second chapter therefore accomplishes two things. The first: I make the empirical claim that rational persuasion is likely to cause backlash - and that rational persuasion is largely ineffective in the case of getting people to eat less meat. The second: a normative claim that we ought to not use rational persuasion but can instead look to its main alternative which is non-rational persuasion or propaganda to get people to eat less meat.

Persuading people that eating meat is morally wrong induces significant backlash because it comes as a threat to people's identity, including their self-image, and to various freedoms such as what one purchases, cooks, and eats. Here I draw on the first chapter of the dissertation to argue that eating meat is a kind of social identity norm. People eat meat because it is what one ought to do as part of the groups they belong to. Again, this might involve just fitting in with one's family or friends, or other identity-based attachments to eating meat such as due to one's gender or nationality.

The first contribution of this chapter is to make the empirical claim that we should not morally persuade people to eat less meat because it will be ineffective. Moral persuasion that predictably provokes backlash, as is the case with meat consumption, can entrench opposition by triggering defensive identity closure, a process in which individuals reject moral appeals not on

their merits, but because they perceive them as threats to their social identity. Research in social psychology shows that when people's values or group memberships are challenged, they often experience identity threat, leading to reactance and a doubling down on prior beliefs (Kahan, 2015; Haidt, 2012). Thus, beyond questions of short-term effectiveness, the use of moral persuasion in contexts of likely backlash can have long-term counterproductive effects, hardening divisions and impairing the conditions for future social progress.

The second contribution is to argue that we should engage in nonrational persuasion or use propaganda to convince people to eat less meat. I make this argument on utilitarian grounds given that so much good can be produced in the form of reduced suffering for humans and non-human animals if vegetarian diet becomes widespread or from drastically reducing the amount of meat we consume. I also make it on the grounds that given the systematically non-rational basis upon which humans make judgements about what animals are acceptable for consumption. Thus, propaganda is a live alternative once we recognize the difficulty in employing rational persuasion here.

I think it is doubtful that moral rational persuasion will work when it comes to eating meat. Most of the judgements people make about eating meat are not rational but due to their identity or learned cultural concepts of what animals are for consumption, which in turn are infused with advertising and propaganda. I think this calls into question not simply whether rational attempts to persuade people not to eat meat will work, but whether it even makes sense in this case. My view is that it does not and as such we should judge the consumption of meat to be a case where propaganda is permissible.

The importance of this chapter is to argue for an effective alternative to rational persuasion. Propaganda might seem a stretch to those who find it a disdainful means of social

change, but I think that the case of meat consumption is an important exclusion. Understanding that propaganda can be permissible is important in the context of encouraging people to eat less meat because it acknowledges the ethical and strategic boundaries involved in promoting behavioral change. When advocating for reduction in meat consumption, leveraging persuasive messaging or propaganda can be necessary to counteract cultural norms of eating meat, as well as to motivate healthier and more sustainable choices. Recognizing that such persuasion is permissible within ethical standards helps ensure that campaigns are designed responsibly, while still being effective in influencing public attitudes and behaviors toward more sustainable diets.

1.3 Shame and norm change

Given the previous chapter one might wonder: should we use non-rational persuasion such as emotional appeals to change people's behavior more generally? Even if we can persuade people through emotional appeals, does that mean we should? Many people would argue no: we should persuade people rationally and not pull on their heart strings or induce fear and especially we should not shame people to try and change their behavior. In this chapter, I zero in on shame and argue that there is an important exception to when shame is permissible, namely when it aims at moral persuasion.

Shame has long been used as a tool for social change and operates through a devaluation of self to change people's views or behavior. This worries some ethicists that using shame as a tool for social change is coercive or manipulative. Shame might successfully change how people behave, but it corrodes human dignity and respect for other people. The most famous philosophical opponent of using shame to change society is likely Martha Nussbaum. Nussbaum puts forward the view that shame is destructive, narcissistic, and expresses a contempt for inadequacy and neediness, which are ultimately defining factors of the human condition. When

we shame others, we try to devalue their self-image thereby coercing them into change in a way which is disrespectful to that person but also disrespectful of the neediness and incompleteness of humans more generally. Thus, in Nussbaum's view, shame ought to have no place in social change.

I argue that these are not necessary features of shame and that it can be constructive when it aims at moral persuasion. Shaming others can be justified if the intent behind it is to promote rational exchange, rather than to humiliate or oppress individuals. In fact, if someone is closed to moral persuasion then it is a duty to try and reintegrate them into this practice because it expresses a respect for them that they may not have for themselves given their failure to participate in rational exchange. This suggests using shame is even an obligation in such cases.

My argument contributes to the broader debate on the ethics of social change in global development but especially pertaining to norm change surrounding public health. I challenge the notion that shame is inherently harmful, suggesting that in certain contexts, shame can serve as a moral motivator, encouraging individuals to open up to rational persuasion. My argument emphasizes the importance of intent behind shame, as a way to include people in the practice of moral persuasion, rather than as a way to simply tell people what they should or should not do. By arguing that shame is acceptable if it aims at moral persuasion, I offer a principled approach to when shaming is permissible.

Understanding whether shame ought to be used for social change is important and timely. Increasingly it is this case that public health urgencies create a demand for tools that are effective in achieving population-level behavioral change, e.g., mask-wearing, handwashing, etc. But shaming can easily fail to respect human dignity, autonomy, and agency. How can we be sure that people aren’t shaming others and expressing a kind of contempt for them and looking down

at them as people who are less than because they aren't doing things the way we believe they ought to? Given this ethical risk involved with shame, it may seem that we should not use shame at all as a tool for social change. My argument provides an important exception to this: that if the intent behind shaming is to include people in the practice of moral and rational exchange, then it can still be respectful to shame someone.

My view provides guidelines for how to use shaming appropriately in development contexts: when someone is closed to moral persuasion. However, there is an important caveat. While everyone ought to be open to moral persuasion in some circumstances, it is not a perfect duty, and we cannot be open all the time, nor can we demand openness from others all the time. Beyond this complication is the additional wrinkle that one must be a competent judge of how people engage in moral persuasion and what constitutes a failure to be open. Nevertheless, if people are informed judges of when others are closed, then shaming can be used as a tool to open others up and invite them to partake in moral discourse. My argument is important because it offers clear guidelines for when shaming is permissible as a tool for social change and what are shame's limitations.

Chapter 2 - Social identity norms

2.1 Introduction

On October 31st 2018, climate activists known as Extinction Rebellion (XR) protested with a sit-in at the UK headquarters of Greenpeace. They claimed that the non-profit Greenpeace was not radical enough and that more government action was needed to prevent social and ecological collapse, climate disaster, and biodiversity loss. Today XR operates through a decentralized network of grassroots movements comprising 989 local groups. XR is fascinating because they have encouraged many people to change their lifestyles and to see themselves differently. When people start to take part in the activist culture of XR, such as the lighting of a flame, the funerary march, the sit ins, or even the more mundane events of handing out flyers, they start to see themselves differently and develop a new identity (Stacey, 2019). This new identity involves identification with the XR movement itself and helps to spur collective action (Furlong & Vignoles, 2023). For instance, XR members believe that to be part of the group one ought to engage in non-violent protests (Jansma et al., 2024).

I think such social movements can teach us something important about why people engage in social change, and how to intentionally design future social movements. Notice that an important mediator between collective action like protests is because someone identifies with a group and think that is what they ought to do as part of the group. Psychologists call this a social identity. I think the kind of social normativity involved in matters of social identity is unique. This is because the motivation to adhere to identity-backed norms is deep: it involves intrinsic sources of motivation like self-esteem, personal identity, and self-worth (Bar-On & Lamm, 2023). And because it is a social identity, i.e., one which involves complexities such as what the group stands for and who one is within the group, there are important, group-related reasons

people follow norms such as a way to solidify the group's existence or as a way to differentiate one's group from others (Marques et al., 1998). In this chapter, I will argue that we should view this kind of normativity as unique and hence should view such norms – social identity norms – as a distinct type of social norms.

Before I get there, however, let us take a step back. What are social norms more generally? A common answer is that social norms are informal rules that prescribe what one ought to do in specific situations based on what is commonly approved or disapproved within a group (Bicchieri, 2006). When enough people start to think enough others act and approve of certain conduct, they will match their behavior to fit in. It helps to know if a pattern of behavior is a social norm because it offers certain means to change the pattern, e.g., by changing who people care about or by changing perceptions about what most people do. But it is insufficient to know whether a behavioral pattern is a norm or not. It also helps to know what kind of norm one is dealing with. For instance, gender norms are often thought to be unique and distinct from other kinds of social norms because they involve power relations as well as being highly embedded in institutions (Cislaghi & Heise, 2019). If one was trying to change a gender norm it would help to know that power relations support them and to offer strategies which are sensitive to such peculiarities, e.g., by developing strategies which allow renegotiations of power, by empowering female voices, or by using role models that disrupt stereotypes.

By starting with the aim of changing harmful norms, we can see that the current accounts of social norms presented in the literature have deficiencies when it comes to norm change projects, namely they are not fine-tuned enough to offer suitable advice on how to change certain norms. And, at least at first glance, recognizing that not all social norms are the same, that there are distinct kinds, can help inform strategies about what interventions are likely to succeed.

Thus, the first step is to show that there are indeed significantly different predictions about what motivates people to follow social norms.

The main goal of this chapter however is to put forward the view that "social identity norms" are a distinct subclass of social norms not captured by existing theories of norms. In particular, social identity norms depend on identification with a group and are motivated by group-related dynamics such as group survival and distinction. These kinds of motivations are not captured by current accounts of social norms, like Cristina Bicchieri and Charlotte Witt's, whose accounts I will contrast with my own account after presenting it.

This chapter is divided into six sections. In the following section I show that philosophers aim to provide comprehensive accounts of social norms, but in fact only focus on certain reasons why people follow social norms. In §3, I offer my account of social identity norms with a particular focus on the social identity of XR membership. In §4 I show that Bicchieri cannot account for social identity norms because she categorizes norms as either personal or social but social identity norms are both rooted in self-understanding and socially determined making them an exception. In §5 I show Witt can handle such an ambiguity but that ultimately, she cannot account for identity norms since the latter involve self-legislation and internally motivated norm adherence which Witt explicitly denies. In §6 I explain how an account of social identity norms can support novel predictions about how to change norms.

2.2 Different kinds of social norms

In this section I begin by providing a general overview of norm psychology research. I then go on to show that philosophers aim to offer a general, overarching account of social norms. I will argue contrary to this that it is particularly important to identify subtypes if the goal is norm

change. I will then focus on identity norms, what seems to make them distinctive, and offer an example which I use throughout the chapter.

Norm psychology is a name for a suite of psychological mechanisms that are responsible for and explain how agents learn, comply, and enforce norms (Westra & Andrews 2024; Heyes 2024; Kelly, Westra, & Setman 2025). Many theorists take a cognitive-evolutionary approach because they believe these psychological mechanisms are adaptations to important selection pressures in human evolutionary history which favor the acquisition and enforcement of culture (Richerson & Boyd 2005; Chudek & Henrich 2011; Kelly & Davis 2018). Early accounts of norm psychology suggest that it is a functionally integrated suite of psychological mechanisms that were responsible for soaking up culture in humans in a Pac-Manesque manner (Sripada & Stitch 2006). This idea is corroborated by a wealth of empirical evidence regarding the ontogeny of norm psychology, which shows that from an early age humans follow, enforce, and infer norms (Rakoczy, Warneken, & Tomasello 2008; Schmidt & Rakoczy 2023). A focus on norms and their psychology has led to many interesting research contributions such as cross-cultural differences between norm strength, i.e., how norms in some cultures are strictly enforced with little leniency for deviation while in other cultures norms may even be routinely violated or only 'loosely' policed (Gelfand 2018).

A focus on norm psychology in this paper is due to the role it plays in norm change or norm-based interventions. By far the most prominent and widespread account of social norms and norm change comes from philosopher Cristina Bicchieri. In Bicchieri's research, especially *The Grammar of Society* (2006) and *Norms in the Wild* (2017) it is implicit that her aim is to explain social norms as a unified category which does not consist of subtypes. In other words, she aims to explain what all social norms share in common rather than pick out a specific type of

social norm. The goal of her account, Bicchieri writes, is to "single out what is fundamental to social norms", thus providing a general analysis as opposed to an account of one type of social norm (2006, p. 10). Bicchieri (2017) contrasts social norms with other kinds of constructs like conventions, descriptive norms, and moral norms, but she takes her account of social norms to be definitive. Similarly, philosopher Geoffrey Brennan (2013) and his colleagues distinguish between moral norms and social norms but take social norms to resemble each other enough such that there is no need to further subdivide the category.

What matters for changing social norms is to figure out what reasons people have for norm adherence to develop causally efficacious interventions around these factors. If there are substantially different reasons why people follow social norms, then a natural move is to further subdivide the category of social norms into its types. These types can then be used to reflect the different kinds of adherence to social norms. There are in fact several, distinct theories about why people follow norms. One view is people adhere because people perceive it is in their best interest to do so given what others do or believe (Bicchieri 2006; 2017). Another view is that norm adherence follows from role. People occupy social positions and they must adhere to the associated norms even if they care little or not at all about the norm itself (Witt 2024). Yet another view I advocate for here is that people follow norms because it becomes constitutive of their group membership and social identity. I will argue in the penultimate section that no unified account of social norms will be useful in speaking to these varieties of norm adherence. We should acknowledge different types of social norms because a unified account runs the risk of homogenizing a phenomenon that is multi-faceted and complex, or rather, runs the risk of lumping together several distinct phenomena which would benefit from separate treatment. In the following sections I will argue in detail that there are these three subtypes of norms. Each

subtype incorporates a distinct analysis of why people follow norms and how to change them. I contrast identity norms with social learning norms first and then move to contrasting identity norms with role norms after. In the following section, I flesh out what makes identity norms distinctive.

2.3 Social identity norms

Social identity theory is a distinct theory that aims to explain why people categorize themselves and others into social groups (Tajfel & Turner, 1979; Turner et al., 1987). The idea is that groups play a large part in how people understand themselves and what norms they will follow. Cross cultural psychology and anthropology research attests to this idea that individuals conform to norms to validate their social identity (Cancian, 1975; Molinsky, 2007; Rathbone et al., 2023). What these researchers found is that as people entered different groups, they quickly adopted the norms according to the commonly held ideas about what people should do in the group, or what people should do according to the social status or position they hold. Some of the hypotheses social identity theorists have developed include the prediction that people treat ingroup members more favorably, express distrust about outgroup members, and that people attach great motivational and emotional significance to their membership of groups they identify with. This research shows that when people voluntarily categorize themselves according to groups, they understand themselves differently and readily adhere to the group's norms.

A key feature of norms according to social identity theory is that the motivation to follow norms is intrinsic, meaning it is internal to the agent, here specifically tied to self-understanding and self-esteem. Adherence to norms symbolizes one's group membership, social status, or shared values. It is insofar as people value the group they belong to, or who they are in the group, that they care about following certain norms. In this sense, the motivation to adhere to norms is

intrinsic because it is tied to how people perceive themselves, or who they want to be. And, for social identity norms in particular, people are motivated because the existence of a legitimate group offers a way to bolster one's own subjective reality and self-worth (Marques et al., 1998). I offer the following account of identity-based, social norms:

Social identity norm: an expectation about how one ought to act which stems from identification or self-categorization with a social group.

A clarification about the definition is that I intend for this account of social identity norms to be fairly general and impartial on what counts as identification with a social group. On a narrow interpretation, social identity causing norm adherence might refer to some agent having an explicit belief about how a norm relates to their social identity (how they self-stereotype or categorize themselves), which motivates adherence to the norm (Turner et al., 1987). But social identity might play a broader causal role with no such belief such as when people have internalized a social identity or feel the need to secure or validate their social identity (Bar-On & Lamm, 2023). What the account posits is that humans are motivated to identify with social groups, and identification with social groups is a distinct form of adherence to norms or normativity.

Humans identify with social groups for a variety of interconnected reasons. At a basic level, group identification fulfills fundamental human needs for social belonging and connection. It provides social bonds, fostering a sense of community that contributes to emotional well-being (Baumeister & Leary, 1995). Beyond these foundational needs, another significant motivation for social identification is the way it confers a sense of legitimacy and purpose to one's self-worth. When individuals see themselves as members of a group, they often perceive their actions and aspirations as part of something meaningful and worthwhile. This group-based self-concept

can elevate personal self-esteem, as success or positive regard for the group reflects back on the individual, reinforcing a sense of value and importance.

However, this raises important questions about why individuals choose to associate with particular groups in the first place. One reason is that groups often serve as frameworks for differentiating oneself from others. Such differentiation can strengthen a person's identity by emphasizing unique group attributes or values that set them apart from outsiders (Marques et al., 1998). Alternatively, attachment to group members—the interpersonal bonds and shared experiences within the group—can motivate identification, as close relationships and shared goals create a sense of loyalty and emotional investment.

Once someone strongly identifies with a group and cares about maintaining that belonging, the group's coherence and unity become especially important. In contexts of intergroup conflict or competition, this importance becomes even more pronounced. Individuals are often motivated to adhere to group norms that reinforce the group's distinctiveness and legitimacy, thereby strengthening group cohesion and defending its social position (Tajfel, 1978). Norm adherence in such settings functions as a demonstration of group solidarity and unity. For example, minority ethnic groups frequently emphasize their cultural norms more strongly when facing pressures to assimilate, viewing such norms as expressions of their cultural identity and sources of collective strength (Barth, 1969; Tajfel, 1981; Phinney et al., 2001).

In summary, human identification with social groups is driven by a blend of fundamental psychological needs, strategic differentiation, emotional bonds, and the desire for social recognition and legitimacy. These motivations work together to promote cohesion and resilience within groups, especially in challenging social or intergroup circumstances.

An example of a social identity norm which I will revisit throughout the paper is the rise of 'non-complacency norms' in the Extinction Rebellion (XR) group during their rise as a grassroots, social movement. The local norms that XR changes is that of creating a "culture of participation" or basically getting people to care and play a role in advocating for sustainable development. When people join the culture of participation and abandon complacency it is very likely they do so because they identify with other XR members, and they sympathise with the purpose that the group exists for. XR embodies nonviolent civil disobedience, and they promote rebellion as a way to spur action against climate change, including to stop support of fossil fuels. Many people refer to XR as a family, as a cause which many identify with, and which has many roles and ranks within it such as leaflet designers, stewards, public outreach, demonstrators, and more. Their shared identity follows from the idea that everyone is affected and that if enough people leave their "comfort zones" to act, then hopefully with enough momentum they can bring about change (Furlong & Vignoles, 2023). The change that people who identify with XR want is bigger than changing local norms; they want governments and financiers to divest in fossil fuels and for countries to focus on reducing carbon emissions to sustainable levels, but the more local norm change they inspire is that of extinguishing complacency and replacing it with taking action such as occupying streets, leading demonstrations, and spreading information about climate change. XR advocates want to show that individuals can band together to stand up for climate change even if their governments will not. This has led to many local chapters or 'affinity groups' of XR such as Lawyers and Doctors for XR, and XR Youth. Those that identify with XR commonly state that the movement gives them a purpose and that they are proud of the community that they are creating.

2.4 The limits of Bicchieri's view of social norms

In this section I will show that while philosopher Cristina Bicchieri gives an account of social norms she takes to be general, there are some things we take to be social norms she cannot handle well. For example, Bicchieri cannot fully account for norms rooted in identity and intrinsic motivation. Furthermore, Bicchieri's classification of norms separates norms driven by personal choices from those driven by social, external influences. This separation does not mesh well with social identity norms since they are driven by processes of self-legislation with social groups and are therefore both personal and social. Finally, I consider whether Bicchieri's work on scripts and schemata change might be able to account for processes of identity-backed norm change such as those we find in XR members adopting norms of non-complacency. I argue that identity norm change is different because it involves a substantial affective component tied to self-understanding which schema change does not. A reasonable resolution in light of these shortcomings is to conceive of identity norms as a distinct subtype of social norms which are not amenable to Bicchieri's categories of norms.

Bicchieri (2017, p. 35) offers the following account of social norms:

A social norm is a behavioral rule which individuals prefer to conform to on the condition that they believe (a) most people in their reference network conform to it (empirical expectation), and (b) that most people in their reference network believe they ought to conform to it (normative expectation.

The first condition is an expectation about how others behave. Its basis is a judgement about what one believes others commonly do such as holding the door open, speaking quietly in libraries, or wearing a mask during a pandemic. The second condition is an expectation about what others believe to be appropriate such as others' belief that you should speak quietly in libraries or that one ought to wear a mask during a pandemic. A reference network is the sum of

people whose opinion you care about when deciding whether or not to conform to a norm. So, when individuals wonder whether they should recycle they likely look to their neighbors, friends, or maybe climatologists.

This account captures much of what seems right for many norms, especially cases where norms are supported by misperceptions of what people really believe. For example, revenge norms and binge drinking norms have both been shown to involve cases where people think they know how many people actually follow this norm or what others believe but are in fact mistaken (Prentice & Miller, 1996; Smerdon et al., 2019). When these misperceptions are corrected and individuals form accurate expectations about what others do and approve of, then their desire to engage in those norms changes accordingly. For instance, previous research shows that a central factor in excessive consumption of alcohol on college campuses is the perception that students frequently consume alcohol, that they consume a lot, and that others approve of such behavior (Ellickson & Bell, 1993; Suls & Green, 2003). Once students find out that their perceptions about others' consumption of alcohol are incorrect their desire to drink reduces significantly (Haines & Spear, 1996). These findings have led to interventions that have been effective, like notes in hotels about the percentage of guests who reuse towels, information about neighbors' energy consumption, and more.

I will now switch from a description of what's right about Bicchieri's account to its limitations. I will use the counterexample of XR activism and its norms of non-complacency to show that Bicchieri cannot account for intrinsically valued social norms that are tied to identity. I will also show that Bicchieri cannot account for identity norms as a kind of moral norm because identity norms are not solely personal. On the contrary, identity norms are personally meaningful and socially dictated which cuts across Bicchieri's conceptual categories.

Let's return to the example of XR activism. The reason that people give for joining in on the XR movement is that they feel a sense of connection to the other members, that they feel in-group solidarity, and that often this leads to an integration of the group and its norms into one's own identity. A slogan used by XR activists (and other activist groups) is 'being the change', which "differentiates [them] from those who don't act in this way, building a thicker group identity" (Stacey, 2019). It seems plausible at first glance that this kind of non-complacency norm founded upon identity is not commensurable with Bicchieri's account of social norms. That is because the motivation to adhere is not wholly or even primarily contingent on what we expect others to do or believe, rather the motivation comes from self-perception. The person that XR members want to be is someone who is not complacent and may even risk being face to face with police during protests or demonstrations. The idea that I should follow this norm if I want to be such and such a person is different from what I should do based on what others expect of me, or so I will now argue against several objections Bicchieri might raise.

Bicchieri would likely object that her concept of a reference network captures the type of norm following I suggest here. A reference network is the sum of people whose actions and beliefs we care about when deciding whether or not to follow social norms. So, when people abandon their norms of complacency and endorse activism of the kind XR members do, then they are simply switching their reference network from their local friends and families to the members of XR. Instead of doing what their old friends and family expect them to do or think they ought to do, they are now following the XR lifestyle and what this community expects from them. This seems like a plausible objection, but the problem is that it doesn't actually speak to why people are motivated to follow these XR norms of non-complacency. When members go out and protest together, recycle, or buy organic food, they are embracing a new understanding of

who they are. Research by conservation social scientists Laura Thomas-Walters and Anandita Saberwhal (2023) explored what determines social identification with XR and found that the most significant factors is that they perceived other XR members positively, that they perceived XR's actions to be disruptive for good reason, and that XR meshed with their environmental self-identity. Environmental self-identity was measured by asking people whether they considered themselves a steward of natural resources and whether living a sustainable life was important for who they are. So, it seems some of the motivators behind adopting these new norms is that it gives XR members a sense of belonging to a group whose actions and members they view favourably, and that the XR pro-environmental identity aligns with their own. The social identity of XR members is someone who contributes to acts on their concerns for the climates by engaging in non-violent civil disobedience and pro-environmental action. This is not the same as doing what others expect you to. The latter is a form of extrinsic motivation contingent on others' behavior and beliefs; the former is a form of intrinsic motivation tied to one's sense of self and knowledge of group identity. A change in one's reference network alters who expects what of you, but it does not imply that there is a modification of one's self-understanding. However, the embrace of XR norms is motivated by exactly that: an internally driven desire to identify with XR. Following norms because you identify with them is not the same as changing your reference network.

Bicchieri has other conceptual resources to potentially deal with identity-backed norms such as her category of moral norms. Bicchieri argues that a moral norm is a norm which people prefer to follow because they believe it is the right thing to do, irrespective of what others do or believe. Moral norms are therefore independent of external or social influence in the sense that someone might follow them regardless of others' behavior or beliefs. Social norms are

interdependent because they depend on what others do or approve of. So, perhaps identity norms are really just moral norms, which would mean that there is no need to conceive of identity norms as a subtype of social norm. The problem with this response is that although identity norms are motivated internally, they are not followed irrespective of others because they are tied to knowledge about what social groups are including what other group members do. For instance, a predictor of social identification with XR is that other members adopt a sustainable lifestyle, participate in public protests, that it is important to take climate action, and that other members think everyone has a responsibility to take climate action. It would not be possible for a person to understand themselves in these terms without the existence of the group, but the choice to join the group and categorize oneself as a member is voluntary and internal. This makes identity norms a sort of quasi personal, quasi social norm. They are personal because people understand themselves with such norms and their motivation relies on self-understanding. However, the content of this self-understanding, i.e., knowledge of the categories by which people use to understand themselves, comes from a social provenance – social groups. It doesn't seem like identity norms fit well into either category of moral norms or social norms in Bicchieri's view.

Another way of understanding this point is that it exposes the limitation of Bicchieri's rational choice-theoretic framework. Bicchieri categorizes norms according to the explicit beliefs and preferences one has with respect to following them. So, the framework is not meant to capture differences in norms based on other causal, psychological factors such as the motivation to belong to a group. For Bicchieri's purposes, XR beliefs just are moral norms. But since Bicchieri applies her framework to the real-world dynamics of norm change and clearly seeks to

offer an account of why people really follow norms and how to change them on such a basis, it is fair to hold her account to such a standard (Bicchieri, 2017).

Another way Bicchieri might be able to handle this objection is through her work with Peter McNally on schema change. This work is substantially different from Bicchieri's rational choice-theoretic work on norms, as it seeks to offer an account of norms grounded in psychological mechanisms (scripts and schemata). Some philosophers suggest that this means Bicchieri really has a dual account: one account for what happens when norms rise to the level of explicit awareness and are considered as such, and another account for how norms are commonly experienced (Guala, 2017). How norms are commonly experienced, according to Bicchieri, are as scripts where one simply applies rules automatically to a situation at hand. Scripts prescribe activities at "varying levels of specificity that people automatically engage in" (Bicchieri, 2017, p. 132). Instead of having to consciously calculate what others approve of or expect of us, we often rely on previously developed schematic understandings of what one ought to do in a particular situation. A notable example comes from the "Saleema" campaign where uncut women in Sudan were associated with the schema of "ghalfa" meaning impure and promiscuous (Bicchieri & McNally 2018). The campaigners re-branded uncut women as "Saleema", meaning whole, intact, healthy, pristine, among other connotations of purity and completeness.

This means Bicchieri has an account of the proximate psychological mechanisms that cause people to follow norms. But I find it implausible that this psychological story of norm adherence can capture the kind of motivations involved in following norms tied to social identity. Providing new knowledge to interpret reality is not the same as socially identifying with a group. The former involves modifying what it means to be uncut which has effects on whether people approve of women being uncut or not. The latter involves developing an identity around

group membership which may include emotional bonds to other group members or identification for what the group stands for, as is the case with XR members. There is no substantial affective component to schema change nor is there a substantial role that identity plays. Schema change mostly involves categorizing situations and people differently. This might involve a re-interpretation of what one's body reflects about oneself, but that is not the same as developing a new understanding of oneself. Schemata are different than social identities.

The specific cases of norms that are problematic for Bicchieri's account are those tied to group membership. More specifically, some norms become personally meaningful and motivated by a desire to perceive oneself as a member of a certain group or occupant of a particular role. These identity norms, despite being motivated internally, still have a social or external provenance such that they cannot plausibly be called a personal norm. A natural resolution is that they are a type of norm which is both personal and social. I believe Bicchieri's existing framework cannot accommodate these kinds of norms and as such we should be open to social identity norms as a subtype of social norm. In the following section I move to another account which seems better equipped to handle the identity-backed cases I've argued for.

2.5 The limits of role norms

In this section I ask whether philosopher Charlotte Witt's account can handle social-identity norms. I will describe Witt's view of how norms are rooted in social positions and roles. I will show that her view can handle one of the problems that Bicchieri's account cannot, which is that Witt acknowledges social norms are often personally meaningful and still social. I show ultimately that Witt cannot account for social identity norms because Witt is committed to the idea that people adhere to norms without actually endorsing them. For Witt, norms are simply

what one ought to do as a product of one's social situation and position – not their attitudes about the norm.

Bicchieri's account of social norms is perhaps the most well-known and influential in the global development arena, but Philosopher Charlotte Witt offers a competing account of social norms that is worth discussing in this context because she acknowledges that norms are rooted in personal identity (2023). Witt argues that social normativity is rooted in the external world of roles, positions, and identities, and are independent of individuals' attitudes or actions. It is by virtue of someone's specific social role that they experience norms as personally binding. Furthermore, Witt offers a different perspective on how to change norms. Norm change according to Witt comes from those who develop expertise in their social positions, like the veteran doctor who has the requisite techniques to excel at the standards of their profession. These people know how to perform their social position well and also know why they follow the norms they do. This vantage point allows them to be critical of norms which do not serve the position well. For instance, shifts in the norms of medical practitioners to include triaging is now almost universally standardized; this wasn't always the case as prior to the 18th century most suggest medical care operated on a first come first serve basis.

Witt's view seems to be able to handle the objections which Bicchieri's cannot. Recall that one problem with Bicchieri's view is that it typifies norms either as personal or social, but that identity norms such as those of non-complacency which XR members endorse are plausibly both. This is not a problem for Witt because her account can handle such an ambiguity. Witt argues that norms are always personal because individuals experience norms as relevant precisely because of their identity or role. However, the norms which define the relevant role are not simply up to individuals to decide. For instance, what counts as being a good carpenter or

flautist, Witt says, is always determined by local contexts and local standards of excellence. Norms attached to roles are therefore socially learned techniques that require others' recognition (Witt 2023, p. 48).

The other criticism of Bicchieri's view is that it cannot account for the internally determined, self-derived aspect of norm following. An important feature of norm following according to the identity view is that people care about norms because they are constructs by which they understand themselves. When XR activists follow the norms of non-complacency such as marching in the street, occupying buildings, eating organic, and so on, they do this because they care about validating their group membership not just to others but also for themselves. These norms (of non-complacency) are followed because people identify with them and subscribe to the XR group. Another factor that likely adds to XR members wanting to subscribe to their norms and perceive themselves as a member is that group members develop emotional bonds to each other. This leads members to care more about their place within the group and their social identity (Tajfel & Turner, 1979). In short, people develop pro-attitudes towards their identity norms because they want to be recognized as part of the group, which leads them, XR members in this example, to care about acting sustainably, or to feel positively about participating in demonstrations, handing out flyers, etc.

Witt cannot account for this aspect of norm following because Witt explicitly denies the view that people adhere to norms because they self-legislate them or have pro-attitudes towards them. Witt's focus is on why people are bound or obliged to follow norms attached to social roles and how the source of this obligation is rooted in the social world. This is clear because a hallmark of social normativity for Witt is that people find themselves obliged to norms because they experience them as *relevant* due to their identity or role. This stands in contrast to being bound to

a norm because of an attitude one has about the norm. It is enough to have the "experience of finding oneself responsive to the norm, which is quite different from affirming or having a pro-attitude toward it", for a person to find themselves answerable to the norms of their social position (Witt, 2023, p. 32). However, for my view of social identity norms it is crucial that people do have such pro-attitudes towards it. People voluntarily self-ascribe the norms of their group and follow norms because they genuinely care about belonging to the community. This is a way to distinguish the category of norms from Witt's category.

Witt has some resources to respond to this objection, namely that people engage in a kind of self-ascription insofar as they pick and choose role models to copy. Students who devote themselves to education see teachers as role models who they aspire to be like, and they willingly imitate them. Witt argues people engage in self-creation by trying to be like their role models, which suggests that people care in some fashion about the norms they endorse because they see roles as reflections of who they are. The problem with this response is that it doesn't explain our central case of XR members' adoption of new norms. While XR members likely have role models, such as Greta Thunberg perhaps, they do not simply emulate these role models when they adopt their new normative lifestyles.

On the identity approach, individuals desire to belong to the group and role, whether it be as an XR member, vegan, climate activist, and so on. Copying on Witt's view does not carry such connotations of an individual who wants to identify with the group, but simply what one ought to do as part of the group. These are significantly different. A strong indicator of XR membership is perceived injustice against one's social group, which is especially strong in youth since they feel they are burdened by older generations' actions (Bond et al., 2020). Many XR members identify as such precisely because they understand themselves as belonging to a group which is adversely

affected. This goes beyond wanting to copy someone because you identify with their role to seeing oneself as a member of a group. And being a member of a group involves an affective component to norm following which includes people developing attachments to other group members, wanting to be recognized by these members, and caring about this social identity because it becomes central to who you are. So, Witt may acknowledge that adherence to norms involves some forms of self-legislation and self-creation, but she denies that it revolves centrally around processes of self-understanding, namely that agents care about norms primarily because they identify with them. Witt's view therefore differs substantially from the identity view.

What about changing social norms though? Witt offers different resources than Bicchieri to change norms. Recall that the problem with Bicchieri's account of norm change is that it doesn't involve an affective component tied to self-understanding. In the case of XR members they feel strongly about their bonds with the group and care about how norms reflect who they are. Witt argues norms change due to agents who develop expertise in roles such that their expertise allows them to be critical of the norms associated with their roles. A typical case that seems to fit Witt's view is that of parenting. Parents who have experience often rely less on general rules of how they ought to train their kid and start to adapt. Instead of following the rule of checking on their kid every time they cry, they recognize that some cries are significant and worth checking out while others are not. More generally, Witt's view seems to explain well cases where there are techniques associated with roles such that someone can be a better or worse parent, carpenter, or flautist. Self-understanding plays a part here because clearly someone identifies with these roles to an extent to want to practice it enough that they develop novel techniques and ways of excelling at the role.

This view of norm change does not seem to encapsulate how identity norms change though. Identity norms change because membership confers positive affect, and identification becomes a way that people desire to understand themselves. It's not clear what the role is that people are becoming critical of in the case of XR members and non-complacency. Perhaps there is a superordinate role category of being a human or global citizen and people excel more or less at this role. However, this hypothesis was actually tested alongside reasons for why people are driven to collective action as XR members. The data show there was no statistically significant tie that identification with humanity or 'global identification' played a part in driving people to collective action. Perceived injustice and XR identification were statistically significant drivers (Furlong & Vignoles, 2020). Furthermore, many people who join XR are youth and have likely not had the time or experience to hone techniques associated with their social roles and develop critiques of their associated norms. It is much more likely that the reason they adopt new norms is amenable to a different, identity-based explanation: membership feels important to them. It makes them feel significant, like they belong to a cause that is bigger than themselves, and they develop bonds to other group members which pushes them to adopt their norms. While Witt offers interesting and novel resources for thinking about how to change norms tied to roles, she cannot offer a robust framework for making sense of why people follow identity norms and how such norms change.

2.6 Conceptual foundations for norm change

In this chapter I argued that there are multiple subtypes of social norms and picked out social identity norms as an underappreciated type. Supporting multiple subtypes of social norms should prove useful for the purposes of changing norms since reasons for norm adherence can differ drastically. For instance, it will be insufficient to change what people expect of each other if

adherence to a norm is additionally determined by a person's social identity. In such cases, change may require offering different avenues for people to develop social identities that are commensurate with the desired change, e.g., being an XR member or a member of a 'green' affinity group if the goal is pro-environmental change. Neither social identities nor social expectations may be effective in altering norms which are seen as functional and crucial to the existence of social roles. Changing these norms may require insiders to develop alternative norms that are equally functional but less maladaptive or to appeal on moral grounds that such norms are unfair or unjust. Simply knowing whether a collective behavior is a social norm may not be enough then, in which case we should ask what kind of social norm it is and to develop interventions accordingly.

Chapter 3 – On the permissibility of propaganda for reducing meat consumption

3.1 Introduction

The Labradormiut are Indigenous people who reside in northern Labrador, Canada. They are incredibly resilient, tight-knit, and direct descendants of the Early Inuit who moved to the east coast from Alaska around 1000 CE. The Inuit travelled east due to the migration patterns of whale and harp seals. Recently, climate change has shortened the seal hunting season, which worries some of the Inuit because it threatens their culture. Seal hunting is important not just for the food and seal skin, but because it is part of Inuit tradition, which offers a reason for the community to gather and teach lessons about patience, determination, and respect for the land.

The Florida Cattlemen's Association would likely sympathize with the Inuit. Floridian cattle ranching began about 500 years ago. Like the Inuit, they are a tight-knit community who are family-focused, and their traditions around cattle ranching are an important part of their history as well as the values they pass down to future generations. And, also like the Inuit, they perceive a threat to this livelihood. But, as Floridian governor Ron DeSantis put it, the threat Floridian cattle ranchers face is not from nature, but from other humans. The threat that they perceive is from those who wish to sell cultivated meat. In May 2024, Ron DeSantis signed off on legislation which banned the sale and manufacturing of synthetic meat grown in a lab using stems cells from animals. At the signing of the legislation, Ron DeSantis stated, “I recognize the threats… take your fake, lab-grown meat elsewhere. We’re not doing that in the state of Florida.” Florida Cattleman’s Association Vice-President Pat Durden spoke after DeSantis and cemented the idea that cultivated meat was not for Floridians: “Farming is core to our state’s identity.” He then repeated the widely recognized advertising campaign slogan: “Beef. It’s what’s for dinner.” (FOX, 2024).

If you put yourself in either of these people's shoes, you can probably imagine how it might feel to face such a threat. Most of your life revolves around these traditions. Your family history, skills, know-how, and values are all there. You wake up and go to sleep thinking about what you need to do on the farm tomorrow or how the hunt might go. You keep track of the changing seasons by thinking about when you can start and stop hunting, or by how big the calves will be. Now imagine what you would feel if these traditions were threatened. It's likely you would want to protect these traditions and do whatever you could to make sure they carry on for future generations. So, how would you react if someone told you that these traditions were morally wrong?

You'd probably feel that you need to protect what you know. Many of your core moral values of patience, hard work, and community come from these practices. Plus, people have been practicing these traditions for centuries. You might think, Ron DeSantis is right: "We're not doing that here". This kind of reaction is known as moral backlash. Backlash to moral persuasion occurs when an attempt to change someone's behavior or beliefs using appeals to morality makes one more likely to engage in the behavior in question. This resistance can manifest as defensiveness, rejection of the persuasion, or counterarguments which reinforce the original beliefs or behavior. For instance, DeSantis and Floridian cattle ranchers would likely respond to arguments that eating meat is wrong by suggesting they are defending traditional American values, freedom of choice, and rural life. So, it's not just about meat but about a way of life that they perceived is being threatened. They might host "freedom BBQs", campaign against "woke" diets, or introduce bills to protect meat in school lunches.

This chapter is a critical examination of attempts to morally persuade people eating meat is wrong. I defend the empirical claim that we should not try to persuade people eating meat is

morally wrong because it is ineffective: it leads to backlash and resistance, which can cause more suffering and harm than if people were not morally persuaded at all. Moral backlash is when people engage in a behavior more after others are trying to persuade them that those practices wrong. For instance, eating more meat after others persuade them it is wrong to do so. Efforts to shift norms around eating meat must contend with the fact that such practices are often deeply embedded in tradition and identity.

The chapter also advances a substantial normative claim: we should not morally persuade people eating meat is wrong on rational grounds. Instead, we should use propaganda or nonrational means to persuade people of the ideal of not consuming meat. This is because the grounds of moral judgements about what is acceptable to consume are systematically nonrational. If it is often not even possible for agents to exercise their rational will when it comes to such judgements, and a great deal of suffering stands to be prevented by people not consuming meat, then we ought to persuade people on nonrational grounds given it will be more effective in persuading people to not consume meat.

This chapter is structured into six sections. In §3.2, I introduce the central topic: the use of moral persuasion to promote social change, with a focus on encouraging reduced meat consumption. In §3.3, I explain why moral persuasion can provoke backlash, particularly when people perceive it as a threat to their personal freedom or identity. In §3.4, I explain why this backlash is especially strong in the case of meat consumption. Many people adhere to norms around eating meat that are closely tied to identity, and efforts to change these norms are often seen as infringing on their freedom and self-image. In §3.5, I show that despite empirical claims against the efficacy of moral persuasion-style interventions, rationalists would argue for the usage of moral persuasion due to their rational methodological constraints. In §3.6, I argue that

moral judgements about what is acceptable to consume are systematically nonrational, and in §3.7 argue this means we should use nonrational means of persuasion, i.e., propaganda. In §3.8 I consider some responses by potential critics of my views.

3.2 Moral persuasion and social change

Moral persuasion is a tool for driving social change. At its core, moral persuasion involves appealing to ethical principles—such as justice, fairness, equality, or dignity—in order to encourage individuals and communities to reassess their beliefs, attitudes, or behaviors. Moral persuasion seeks voluntary change by trying to alter what people see as right or wrong. For instance, activists who want to promote climate change awareness might appeal to the harm that single-use plastics, flying, and lawn watering cause. By appealing to responsibility, intergenerational justice, and other moral values, these activists seek to morally persuade others that how they currently behave is unsustainable and morally impermissible, thereby aiming to change their behaviors and views.

Throughout history, social movements have relied on moral persuasion to challenge certain norms. The civil rights movement, the women's suffrage movement, environmental activism, and LGBTQ+ advocacy, have all used ethical appeals to expose injustices. Activists often frame their causes not merely as political or legal issues, but as moral imperatives that demand action and conscience-driven reform. Moral persuasion can occur in many contexts. For example, Rachel Carson's *Silent Spring* was put in *The New Yorker* prior to it becoming a book. Greta Thunberg's public address or speech where she condemned others for climate inaction became well-known and appealed to many: "how dare you?". Visual media like *An Inconvenient*

Truth and social media hashtags like #ClimateJustice are also currently used to rapidly spread moral calls to action.

One of the most empirically backed and psychologically plausible accounts of how to morally persuade people to eat less meat comes from philosophers Joshua May and Victor Kumar (2023). Their argument is that part of why people continue to consume meat is because they do not really believe it is wrong. People justify their consumption of meat by rationalizing that factory farming would continue without them, or that other people do it so it must not be that bad, or that eating meat is natural and humans have done it since the dawn of time. In other words, people remain unconvinced that there is anything wrong with eating meat. So, to reduce meat consumption, according to May and Kumar, we need to convince people that it really is wrong and that one ought not to eat meat.

May and Kumar's analysis focuses on factory farming, arguably the easiest case to make given the extreme suffering that occurs in this method of meat production. They offer two different methods to change people's ideas about eating factory farmed animals, to be used as appropriate. The first method is to provide people evidence that factory farm animals can suffer. They argue that we can do so through 'consistency reasoning'; we can try to persuade people that chickens and pigs in factory farms suffer by pointing out that they treat their own pet as if it can suffer. And that there is a psychological similarity between pigs and dogs and chickens, meaning minimally that they can all suffer. Because they all can suffer, there is no morally relevant distinction that supports treating factory farm animals so much worse than one's own pets. But since that's what you're doing when you eat factory farm meat, i.e., treating factory farm animals much worse than you would your own pet, then you ought to not eat factory farm meat.

The second method involves using *social proof* to convince people that eating meat is wrong. The particular argument for social proof stems from the more general idea that people typically take cues from respected and trusted figures in their community (Henrich & Gil-White, 2001). This general idea can be applied to how people reason in particular about whether or not to eat factory farm meat. If an elder, professor, or respected manager is not eating meat, then perhaps there is something valid to the arguments or reasons behind why they don't eat it. The aim is that individuals use others' behavior as *evidence* to decide what they should do – a kind of epistemic reliance on social cues. Kumar and May provide empirical evidence that social proof works from domains like climate friendly behavior such as persuading hotel guests to reuse their towels or reducing household energy usage (Goldstein, Cialdini, & Griskevicius, 2008; Schultz et al., 2007). Social proof worked in these cases by providing messages about what other hotel guests or our neighbors do, thereby offering evidence for what we should do.

By using these strategies in parallel, namely consistency reasoning and social proof, people should be convinced of the basic moral arguments why they shouldn't eat meat. Or so that is May and Kumar's view. Both techniques qualify as moral persuasion because they involve appealing to ethical principles (minimize suffering) to in some sense tell others what to do (eat less meat) or what they should believe (it is wrong to eat factory farm meat). Social proof works by directing people's attention towards what trusted others believe or think, which is not in itself a form of moral persuasion. But May and Kumar intend to use social proof to direct people towards the ethical arguments for why one should not eat meat, suggesting that ultimately, they hope people are persuaded by the reasons against the consumption of meat – and not merely changing their behavior due to the fact that other people whom they respect are not eating meat. Consistency reasoning is more clearly a form of moral persuasion given that it

operates by suggesting people ought to change their behavior (eat less meat) because it is a discrepancy given their moral beliefs (that one ought to minimize animal suffering).

Moral persuasion certainly seems a promising route to getting people to consume less meat. The point of these moral persuasion-style interventions is to establish a connection between one's purchasing factory farm meat and complicity in causing animal suffering. Whether it be through documentaries, short videos, or social media content, these interventions intend to expose the realities of animal suffering on factory farms (Mathur et al., 2021). In a recent meta-analysis of interventions to reduce meat consumption, 16 of the 22 were moral persuasion-style interventions which appealed to animal welfare (Green, Smith & Mathur, 2024, p. 7). Thus, moral persuasion seems to be used quite often and is taken to be an important route to getting people to consume less meat.

3.3 Backlash to moral persuasion

While persuasion seems promising, it is not without its difficulties. In particular, trying to persuade people that they are doing something wrong can often lead to moral backlash. People can perceive that they are trying to be manipulated and ignore the attempts. They may also double-down by developing further reasons to justify their behavior or views. Social psychologists call this *backlash reactance*.

Psychological reactance was first coined and theorized by psychologist Jack Brehm in 1966 and developed further significantly by Sharon Brehm and Jack Brehm in 1981. The theory states that agents will attempt to recover their freedom or resist social influence when they perceive a potential restriction in their freedom (Brehm, 1966). Reactance itself is described as an unpleasant motivational arousal that occurs when people experience a threat or loss of freedom (Brehm & Brehm, 1981). This unpleasant motivation of reactance is caused by a

perceived threat to their freedom and attempts at persuasion are frequently perceived as threats to one's freedom. Reactance is not synonymous with moral backlash but is rather an adjacent, psychological phenomena that can contribute to and act as a catalyst for backlash in general and moral backlash in particular. After all, a persuader's intent is to change your behavior or opinion, which is often perceived by people as someone telling you not just what you can or cannot do. And on some views moral persuasion is even stronger than what you can or cannot do. When one person morally persuades another, they are in some sense telling them to *do it* (Darwall, 2006).

An important aspect of reactance theory is the magnitude of the reactance. Depending on how strongly an agent feels their freedom is being threatened, they will…. Reactance response is a function of three factors: *centrality*, *proportionality*, and *pressure*. These originate from Jack Brehm's original proposal of the theory in 1966, but the specifics of these factors, especially centrality, have been expanded by social psychologists and philosophers since then (e.g., Sittenthaler et al., 2015; Rosenberg & Siegel, 2016; Rees et al., 2023). Let me unpack these three factors.

The first factor that augments the magnitude of reactance is the *centrality* of the practice to the individual. One way to gauge how central a practice is to someone is how much they identify with it. If you are an avid soccer fan and play each week, then a politician who wants to replace local soccer fields with shopping plazas will likely induce reactance. Given the centrality of identity to the human experience, the construction of personal identity is a prolific research area in the psychological and social sciences (Crocetti et al., 2025). But one's moral self or self-image is perhaps the most significant part of our personal identity. A moral self-image is a collection of traits that concern morality such as virtuousness, honesty, empathy, and courage. Psychological and autobiographical memory are important parts of understanding who we are,

but research shows that people perceive moral traits to be part of their 'fundamental' self. For instance, subjects express a reluctance to enhance certain traits such as empathy and kindness over others like memory, self-control, or concentration, on the grounds that doing so would fundamentally change who they are (Riis, Simmons, & Goodwin, 2008). When subjects were asked what traits would fundamentally change a person who is aging, morality was rated as the most significant factor that would change that person's true self (Strohminger & Nichols, 2014). Thus, people seem to care deeply about the role morality plays in who they are; moral self-image is certainly of central importance. It is likely therefore that people will display reactance when their moral self-image is threatened by others.

The second factor is *proportionality*—the proportion of freedom eliminated. When individuals perceive that a large portion of their freedom is being restricted, their level of reactance tends to increase accordingly. The more significant the perceived loss, the stronger the emotional and motivational pushback, as people are more sensitive to sweeping limitations than to minor constraints. For instance, making recycling mandatory might seem to require a greater change in behavior than not allowing littering, and as such mandatory recycling would be perceived as a greater threat to one's freedom therefore inducing more reactance than banning littering.

The third factor is *pressure* or the magnitude of the pressure to comply. Psychological reactance tends to intensify when individuals feel they are being heavily coerced or manipulated. The stronger and more overt the pressure—such as aggressive persuasion, authoritative commands, or emotional appeals—the more likely people are to resist, even if they might have agreed under less forceful conditions. Moral persuasion raises reactance in individuals who perceive that the pressure to change is high. Reactance arises not necessarily from opposition to

the message itself, but from the perceived threat to autonomy. Subtle attempts to influence are less likely to trigger reactance, whereas high-pressure tactics can backfire, leading individuals to reaffirm their freedom by rejecting the imposed behavior or belief outright. For instance, accusing and confronting someone about how shamefully large their carbon footprint is, is a high-pressure form of moral persuasion. It is likely to cause more reactance than persuasive efforts which come in the form of self-critique or an invitation to reflect about one's own views or behavior given the persuader's personal experience: "I used to drive a lot but then I started realizing how harmful it was to the environment! Then I realized I could start caring about the environment by taking the train more often.". Feeling pressured is not just a matter of how people persuade us but also who persuades us. We feel more pressured by those who seem to express moral superiority than those who seem to simply offer us information. So, while these are all attempts at persuasion, some are more likely to cause reactance than others due to differences in pressure.

These three factors, centrality, proportionality, and pressure, jointly shape the intensity of psychological reactance. When a freedom that is central to a person's identity is threatened, the reaction is more emotionally charged and deeply rooted. This effect is amplified when the restriction is perceived as sweeping or disproportionate, and when the attempt to enforce compliance is seen as overly forceful. Together, these elements help explain why some restrictions provoke strong opposition while others are more easily accepted: it is not only *what* is being threatened, but *how much* is being taken away, and *how forcefully* it is being imposed.

Any project that seeks to achieve social norm change should aim to avoid reactance. This is simply because reactance decreases the likelihood that persuasive efforts will be successful

and that they will have any desired influence. Even worse, reactance can lead to backlash such that someone becomes more likely to engage in the behavior targeted for change.

3.4 Reactance, moral backlash, and meat consumption

In this section, I show that attempting to persuade people eating meat is morally wrong is likely to affect all three of these factors which cause reactance and therefore can lead to substantial backlash. First, eating meat is often central: it is important to people's identity and plays a significant role in their traditions. Thus, persuading people to not eat meat can come as a threat to their freedom to express their identity. Second, what people eat is something that they typically have had the freedom to choose for most of their lives. Asking individuals to change that can require a considerable upheaval of daily routines and therefore be perceived as a threat to previously held freedoms. And finally, people often feel pressured to change when they persuaded to eat less meat by appeals that highlight the suffering of animals on factory farms. It is likely therefore that many people will display reactance due to such persuasive efforts. This reactance will likely trigger backlash and cause people to continue eating meat or eat more meat than before such persuasive attempts. Each of these factors which augment the magnitude of reactance are worth looking at in turn as they apply to persuading people eating meat is wrong.

The first is factor is centrality. For many people, eating meat is central because it is important to their identity at some level. This suggests that eating meat is a type of norm people follow because they believe they ought to as a member of the group, i.e., a social identity norm. Recall from the previous chapter that a social identity norm is an expectation about how one ought to act which stems from identification or self-categorization with a social group.

We can see different ways in which meat eating can be a social identity norm for different groups. Religious traditions are interwoven with meat: Muslims typically consume

Halal meat and Jews consume kosher meat. Meat dishes can be part of national identity too and has been documented in several countries such as Denmark (Randers & Thogersen, 2023), Australia (Lockie, 2001), and the United States (Joy, 2009). For some eating meat is a sign of masculinity (Ruby & Heine, 2011). More local identities can drive meat consumption too. In Quebec, tourtiere and poutine are ways of preserving the Quebecois identity in face of a growing Anglophone influence. Giving up on these traditions is difficult partially because people want to remember what the food symbolizes. The kind of food we eat is tied up with who we are. Family traditions like grandmothers' recipes are part of 'what we do around here' too. The dishes we eat, some of which include meat, feel important and central to who we are. Thus, persuading people that these traditions are in some sense wrong will likely cause reactance and backlash.

Trying to persuade people that eating meat is wrong can threaten something else of central importance: people's moral self-image. In a study that measured defensiveness of moral self-image, behavioral intentions to reduce meat consumption and support for vegetarianism or veganism were lower after reading information about how eating factory farm meat contributes to violence and that purchasing meat contributes to animal suffering (Shulam et al., 2021). For most people, it is reasonable to think that contributing to unnecessary animal suffering would tarnish their moral self-image. Those who consider themselves empathetic or kind would want to defend the integrity of their moral self against such claims that eating factory farm meat is inconsistent with being a kind of empathetic person. In this particular study, people defended their moral self-image by supporting human-animal hierarchies and de-moralized the issue by suggesting eating meat is normal and natural, i.e., motivated reasoning. Another moral persuasion-style study shows similar results. The goal of this other study was to measure the moralization of meat eating (Feinberg et al., 2019). Participants who prior to the study ate meat

and did not view it as a moral issue were given educational material about the state of animal suffering on factory farms, and more generally why eating meat should be a moral issue. This was a longitudinal study lasting several months where participants were shown videos about animal suffering on factory farms. After each information session, researchers collected data about the participants' intentions to eat less meat and whether eating factory farm meat was a moral issue. While some participants came to view meat eating as a moral issue, the majority of participants in each trial de-moralized the issue by stating that they did not find the issue of eating meat to be moral. It is likely that one reason participants who de-moralized the issue is that to moralize it would threaten their moral self-image. Accepting that the consumption of factory farm meat contributes to unnecessary suffering is to also accept that one is not as kind or empathetic as one once thought. Thus, resisting these claims by denying that animals really do suffer, or that one contributes to suffering when one purchases meat, or that there is actually anything wrong with factory farming, are all ways that people defend their moral self-image and permit themselves to continue consuming factory farm meat.

The second factor is proportionality. Persuading people to eat less meat is likely to cause reactance because it affects a freedom which is significant. If I tell you that you should no longer eat meat because it is unethical, and this means you must change some or every one of your meals, then you will likely perceive my moral persuasion as a significant threat to your freedom. Recall that proportionality is a result of how much prior freedom one has had in this domain. As economist Robert Wright points out, "most consumers have long had extensive freedom to choose what foods to consume, and how much of each food to consume" (Wright, 2012, p. 138). This suggests that by having someone tell you what you cannot eat will be perceived as a

significant *loss* of freedom and that one can expect a great deal of reactance in response to such persuasive efforts.

It is worth contrasting how meat consumption is different from other cases of moral progress when it comes to the question of freedom and its proportionality. May and Kumar suggest that one of the most promising cases of moral progress is: "the large and rapid decline in anti-gay attitudes in the U.S. and other countries over the past few decades" (2022, p. 376). May and Kumar's argument is that once people recognized there was no morally relevant difference that supports treating people differently based on their sexuality, then people stopped holding these attitudes. So, if people can come to recognize that animals suffer and there is no morally relevant difference between pets or one category of animals and factory farm animals, then people will come to care about factory farm animals too and eat less or no meat as a result.

Putting aside the fact that institutions purposely design meat to hide the fact that it comes from a living creature and thereby attempt to mask people feeling empathy for animals, there are differences in these freedoms. For contemporary people in the West, eliminating anti-gay attitudes requires may require little change. This was likely not true at the beginning of the historical process when attitudes toward gay people started to change. Some people disowned or would disown their gay relatives, and some people avoided gay neighborhoods. So, for these individuals, changing their attitudes towards gays requires a reasonably big shift to their behavior and freedoms. But while in the United States many people know someone who is gay, and while there are many gay neighborhoods, accepting a gay person or visiting one of these neighborhoods is not likely a daily, effortful change. The consumption of meat on the other hand is a daily, multi-faceted behavior. Changing daily meal choices involves changing what people purchase, where they shop, what food they prepare, and obviously, what they eat.

The final factor which will affect reactance is pressure. When someone perceives they're being manipulated, such as when emotionally charged tactics are used like making someone watch or read about the horrors of factory farming, the higher reactance is likely to be. But interestingly the pressure one feels to comply is a factor of not just how one is persuaded but who does the persuading. In a study that examined people's intentions to consume meat in response to messages about fewer people consuming meat, researchers found that when the communicator was labelled as a vegan activist, people were less likely to intend to eat less meat. The experimenters' explanation is that "participants perceived the vegan activist as morally superior. The activist's message was interpreted as implying that individuals who eat meat are less moral than those who do not eat meat (Boenke et al., 2022, p. 8). This poses a problem to moral persuasion: those who seek to persuade people to eat less meat by telling them it is morally wrong can be perceived as thinking they are superior, thereby leading people to reject what they say.

Does this mean that moral persuasion as a method to stop eating meat is doomed? I show in the next section that some philosophers would still advocate for moral persuasion, and then in the section after the next I will argue why I think we should not use moral persuasion here.

3.5 Rational moral persuasion

In this section, I show that May and Kumar endorse a rationalist view of how to achieve moral progress in general and reduce meat consumption in particular. Despite the backlash involved with using the method of moral persuasion for reducing meat consumption, May and Kumar insist that we ought to rationally convince people that eating meat is wrong. This is due to a methodological constraint of theirs: we need to achieve progress on rational grounds. Applied

to the topic at hand, this view states that we ought to get people to eat less meat by morally persuading them because eating (factory farm) meat is wrong.

May and Kumar are rationalists when it comes to moral progress. This puts important methodological constraints on how they think we ought to convince people eating meat is wrong. Recall that May and Kumar highlight two cognitive mechanisms - social proof and motivated reasoning - that can, under the right conditions, get people to eat less meat.[1] Their research shows that these mechanisms which typically stand in the way of convincing people eating meat is wrong can actually be harnessed to convince people of these arguments. This last point is important: May and Kumar want to make progress on getting people to eat less meat, but in the right way, which for them means through *rational moral persuasion*. Moral persuasion in general is the attempt to convince someone of a moral opinion through the provision of reasons as opposed to through force or coercion. But there are different sorts of moral persuasion. Some are rational and others are nonrational. The goal in *rational moral persuasion* is that we should convince people of moral beliefs through rational deliberation, i.e., the provision of reasons that anyone ought to accept. The goal of *nonrational moral persuasion* is that we should convince people of moral beliefs through nonrational means, which might involve emotional appeals or other ways which do not require rational justification. What I want to show in this section is that May and Kumar are committed to rational moral persuasion.

May and Kumar draw on cognitive mechanisms like motivated reasoning and social proof not as direct ways to get people to eat less meat but rather to convince people that eating

[1] In a forthcoming paper by Joshua May and Victor Kumar write, "How is moral progress… to be achieved?... Moral judgements matter. And if individuals are to make better moral judgements, reasoning is key at all levels of society." They also write on the topic of meat consumption that a key problem in preventing moral progress is that "most consumers remain unconvinced by the ethical arguments", hence they believe the role of moral reasoning is key in convincing people of these arguments and getting them to eat less meat (May & Kumar, 2022, p. 371).

meat is wrong. Recall that motivated reasoning is the cognitive bias where people use reasoning to reach a desired conclusion - as opposed to objectively evaluating information. May and Kumar think that "personal connections with animals can give people motives to attend to evidence in support of the basic argument against eating meat (2022, p. 376). Notice that what they think should be doing the work here of getting people to eat less meat is still ultimately the arguments against it. Their hope is that personal connections will motivate people to pay closer attention to arguments such as: if it is wrong to cause unnecessary harm to a pet then it is wrong to do the same to factory farm animals. Ideally, people will say yes and will be convinced that it is irrational to hold such inconsistent views and therefore one ought to stop eating factory farm animals.

What about social proof? Again, the idea behind social proof is that it is a mechanism where people use others' behavior as *evidence* to decide what they should do. Here May and Kumar recommend that activists ought to co-opt prestigious figures and trusted members of the community to flaunt their vegetarian ways. This is because they think such social proofs will send a signal to others that there is something *right* about the arguments against eating meat. It could be that people are following what these community members are doing because they want to fit in, they admire them or are using evolutionary heuristics to figure out what they should do, but May and Kumar's aim is not for people to be persuaded to eat less meat because of these reasons. Instead, May and Kumar's aim is to attend to the arguments against eating meat because people they care about are not eating meat. They think it should be the basic ethical arguments that change people's attitudes and behavior about eating meat - not simply that they were influenced in any of these nonrational ways. The problem with getting people to eat less meat is

that they are "not fully convinced that they morally ought to reduce their meat consumption", and as such May and Kumar's goal is for people to accept the validity of these ethical arguments.

Given the empirical research I provided on moral backlash, I believe we ought to be skeptical that continuing to persuade people eating meat is wrong will lead to less people consuming meat. It is likely that many of these attempts will backfire and act as motivation for people to consume more meat in the future. In the next section, I look at additional reasons for why we should relax the methodological constraint of rationality.

3.6 Nonrational moral persuasion

We know that empirically speaking persuading people eating meat is wrong faces the problem of backlash. But what is the alternative to rational persuasion? Talk of non-rational means is likely to trigger charges of propagandizing, which I will address below. But first we need to appreciate there are reasons to doubt that rational persuasion is even a possibility in debates related to meat eating.

Philosopher Cora Diamond (1976) puts forward the view that the reason why we eat meat is not rational or based on logical consistency in her essay on 'Eating Meat and Eating People'. Diamond arrives at this view through a criticism of a typical argument that proponents of vegetarian or veganism often make. The argument is this: we should not eat meat because it is logically inconsistent with a respect for other creatures' right to life and utilitarian principles against not causing suffering. So, if you believe that other animals can suffer, then there is no logical reason to treat your pet well but eat factory farm animals. This view is essentially the one May and Kumar hold: that getting people to eat less meat should be done through rational argumentation on grounds of logical inconsistency.

Diamond argues that not eating animals out of a concern for not causing them suffering misses the point for why we do not eat animals. If not causing suffering was the only concern of vegetarians, then there should be: "nothing queer, nothing at all odd, in the vegetarian eating the cow that has obligingly been struck by lightning" or that a "vegetarian should be perfectly happy to eat the unfortunate lamb that has just been hit by a car" (Diamond, 1976, pp. 468-472). Many vegetarians would not want to eat a lamb or cow even if no suffering was involved at all. If a desire to not cause suffering was the sole motivation in judgements about what is acceptable to eat, then it should also be perfectly reasonable to eat other humans as long as it does not cause distress. But of course, it is highly taboo to eat humans and only in extreme circumstances is it somewhat permissible to do so. This suggests that it must be something else which grounds our moral judgements in what one should or should not eat.

Diamond argues that the moral permissibility of what we consume does not stem from not wanting to cause suffering, so what else could be the basis of our judgements about what is acceptable to consume? Diamond's argument is that these judgements are made on the basis of categories and concepts which are socially learned, and that experience and imagination play an important role in which creatures we sympathize with, which in turn shapes our concepts. For example, we conceptualize some non-human animals as pets and pets are not for consumption. 'Non-consumable' animals include hobby animals, pets, and exotic animals (Pongrácz & Dobos, 2023). One reason we might categorize some animals as non-consumable is because of a fellow feeling that we or our ancestors have cultivated. We sympathize with certain creatures more than others because we spend time near them, learn about them, and because we see ourselves in them. We classify other non-human animals as farm animals or wild animals. We have concepts

within 'consumable animals' such as bacon or pork for pigs. The way the consumable and non-consumable animal categories are formed is not rational. We have fellow-feeling for some animals and label them not for consumption. At the same time, we have labels for animals which we do consume that that deliberately downplays or disables the possibility of feeling sympathy. One way this is done is by dissociating meat from their animal origins through meat's presentation or, again, through the categories we use to label meat (Benningstad & Kunst, 2020). The important idea is that to understand why humans eat meat, and how to persuade them not to eat meat, we first need to appreciate how we make judgements about what things are acceptable for consumption. And to understand this we need to appreciate how it is that we classify creatures for consumption, which depends on many nonrational factors such as how much we can or cannot sympathize with other creatures, as well as existing concepts and classifications.

We need not persuade others that eating meat is wrong on rational grounds because rational grounds are not the basis upon which humans generally make judgements about what they should consume. Such attempts are utterly unpersuasive because they don't speak to the roots of how people decide what to consume. And, further, we know that we should not tell some people how to behave or what to believe more generally because that will likely cause backlash. But then how should we persuade people to not eat meat?

An alternative is to persuade people on nonrational grounds either through appeals to emotions or through shifting the categories people classify animals with. An example of the latter comes from a piece of 'vegetarian propaganda' (1969, p. 59) called the 'Dutiful Vegetarian': "Eat the creatures killed for sale,\ But never pull the pussy's tail.\ Eat the flesh from 'filthy hogs' \-But never be unkind to dogs." Diamond likes this poem because it shows there is hypocrisy and confusion behind how we treat and think about different sorts of animals. But it doesn't seek to

justify to people who eat meat why they are morally wrong, or how they should behave. Instead, it presents a sort of challenge to think critically about why we categorize hogs as bacon but dogs as pets.

Challenging existing classifications is merely one nonrational way to persuade people. And admittedly I think that this kind of nonrational persuasion leads to an inference that someone is bad or wrong for eating meat, which might just as likely cause reactance and backlash. Another option is to try and draw on imagination and sympathy to move certain animals out of the category of objects to be consumed. If we can cultivate sympathy and fellow feeling for animals and recognize the extent to which our lives are similar to those of other creatures, then we might cultivate some moral appreciation for them that excludes them from the consumable category. This might also involve recognizing just how dissimilar animals' lives are to us, and to cultivate astonishment or awe at how beautiful, or delicate, or ferocious other animals are. Whether it is out of sympathy and fellow-feeling or awe, the idea is to use emotional appeals to evoke the idea of animals as living beings which are fellow creatures worthy of moral concern (Aigner, 2021, p. 209).

The specific type of nonrational persuasion that seems to make the most sense in the context of reactance is simply to tout the advantages of a vegetarian or vegan diet. Several initiatives like Slow Food and The Vegan Society promote legumes, especially beans, as an affordable, sustainable, healthy, and ethical plant protein option (Davies, 2017). By offering people information about the benefits of changing their diet towards plant-based foods, these activists can promote behavior change without challenging people on grounds of irrationality or moral wrongdoing.

The key point here is not to argue for a specific strategy within the category of nonrational persuasion, but to show that we are not confined to rational moral persuasion. We have other options and in the case of persuading people to eat less meat, we should use them. Diamond provides us reason to think that when it comes to persuading people to not eat meat on rational grounds is insufficient. Perhaps still, though, many will feel the lingering concern that this amounts to objectionable propaganda. In the next section I embrace that objection and argue that propaganda is indeed sometimes acceptable.

3.7 Propaganda is permissible to reduce meat consumption

One view of moral persuasion is that it needs to be done on rational grounds. May and Kumar think that what is morally true is also rational, so convincing people of what is morally true needs to be done on logical grounds. Peter Singer offers another view: what matters when we persuade people of our moral opinions is simply that they acquire the right beliefs and not how they get there. When it comes to persuading people to eat less meat this means we can and should use nonrational means (e.g., emotional appeals) to persuade people eating meat is wrong. I explain that this means in the case of getting people to eat less meat it is okay to use propaganda.

Peter Singer offers an alternative account of how we should achieve moral progress. Singer's view is that instead of focusing on how individuals arrive at their moral beliefs, we should prioritize persuading them of the correct moral conclusions. He contends that the process by which people come to hold moral beliefs is less important than ensuring that the beliefs themselves are ethically sound. This perspective emphasizes the outcome of moral reasoning over the methodology employed to reach it. The reason Singer believes that it is okay to

persuade people by using false premises or logically defective arguments is that he thinks our moral intuitions are systemically false (Singer, 2005; de Lazari-Radek & Singer, 2012). Singer's view is that we should assume "moral judgements we intuitively make are likely to derive from discarded religious systems, from warped views of sex and bodily functions, or from customs necessary for the survival of the group in social and economic circumstances that now lie in the distant past" (Singer, 1976, p. 516). How do we decide the validity of a moral theory then if all moral theories are based on false intuitions? Singer draws on philosopher Henry Sidgiwck to explain:

> Sidgwick would admit that, in the end, we have nothing to fall back on other than careful consideration of whether some fundamental principle really is intuitively clear and certain (though we may be able to support our intuitions to some extent by a comparison with common sense morality), but he does not define validity for a moral theory in terms of agreement with our ultimate intuitions or in terms of a match with our particular judgments (Singer, 1976, p. 514-515).

If we cannot rely on our intuitions to provide a valid moral theory, then we must search instead for undeniable fundamental axioms. For instance, the axiom of impartiality says that whatever action we judge to be right for ourselves is also right for all similar persons in similar circumstances. The key part here is that Singer thinks we can then compare these axioms with people's moral judgements and use particular ones to support the validity of the ethical axioms. And the larger takeaway is this: when 'we' know that certain ethical axioms are true, and that most people's ethical intuitions are false, then it is fine to draw on false premises, or appeal to people's misguided intuitions, or to follow a chain of 'argument' that is logically defective, so long as the end result is their behaving in ways that fit the true moral theory.

One reason to support this view is the idea that getting people to behave in a way that fits a true moral theory justifies persuading them on nonrational grounds, but for Singer it is justifiable to use nonrational means even more so given that the basis of people's ethical judgements is already systematically false.

I agree with Singer here when it comes to moral judgements about eating meat. People's judgements about whether one ought to eat meat are systematically false. The other ingredient I need is to show that eating meat is wrong is a moral truth - one which would justify me or others in persuading people of. Since this is not an article about the moral permissibility of eating meat but rather how we should persuade people to not eat meat, I will focus on the easier case of consuming factory farm meat and take it as rather self-evident that most ethicists and ethical theories agree that factory farming is immoral (May & Kumar, 2022; Schwitzgebel & Rust, 2014; DeGrazia, 2009). The current way non-human animals are treated on factory farms includes intense suffering, and devastating effects of climate change on humans and other animals (Eshel et al., 2019). So, if eating meat is morally wrong and our moral judgements about the acceptability of eating meat are systemically false, then we are in the same position that Singer arrives at when it comes to moral persuasion in general: if we can be sure of moral truths, then what matters is we get people to follow them, and if that means using nonrational means, then we should.

Another way to understand my argument is that I am advocating for the use of propaganda in the case of getting people to not consume meat. In philosopher Jason Stanley's *How Propaganda Works*, Stanley defines propaganda as follows:

> **Supporting Propaganda**: A contribution to public discourse that is presented as an embodiment of certain ideals, yet is of a kind that tends to increase the realization of those very ideals by either emotional or other nonrational mean (Stanley, 2016, p. 57).

Here are examples of propaganda that Stanley points to. Using frightening public health warnings to make people afraid to smoke with the goal of getting people to not smoke and supporting the ideal of public health. Another pertinent one is appealing to past wrongs against a group to strengthen ethnic pride and self-identification (Ibid). We can apply these forms of propaganda to getting people to eat less meat too: making people afraid of the health risks of eating meat, or discussing how great it is to be a vegetarian, or how people in the past have wronged animals to make people feel good about being a vegetarian. 'Learning to be a dutiful carnivore' by Jane Legge is another piece of vegetarian propaganda, so are posters that state "our love of flesh is killing the planet". When it comes to getting people to eat less meat, these forms of persuasion are permissible.

3.8 Response to critics

One objection to using vegetarian propaganda is on empirical grounds and the other on normative grounds. I'll begin with the empirical. One might object that propaganda can cause just as much reactance as moral persuasion, if not more. Propaganda is often seen as manipulative or deceptive, and when people recognize that they are being manipulated, they may feel their autonomy is being undermined. This can provoke a strong backlash, especially if the propaganda is blatant or perceived as coming from an untrustworthy source. Moreover, propaganda campaigns often come from activists or governments, which may already be viewed with suspicion. If the audience detects that their opinions or behaviors are being subtly

engineered, they may feel even more controlled than when confronted with direct moral arguments leading to significant reactance.

Propaganda should produce less reactance than moral persuasion because it avoids telling people what to do (at least directly) and that they are wrong. Moral persuasion appeals explicitly to a person's sense of right and wrong, which can threaten their self-concept and trigger defensive resistance. Recall that persuasion on some accounts does not simply tell someone they are wrong but also tells someone what to do. Propaganda, in contrast, works by promoting another course of action through any kind of emotional or nonrational appeal. For instance, one way to influence others is to appeal to the health and cost benefits of vegetarian dishes, or the dangers of eating certain types of meat. It may be that by learning about those who we wish to influence we can appeal to their history with vegetarian or veganism, which might offer ideas about what types of protein-rich, non-meat foods might replace meat. Because propaganda of this form doesn't overtly threaten autonomy or identity, it bypasses the mechanisms that typically lead to reactance. Thus, while both forms of persuasion can provoke resistance when detected, propaganda is less likely than moral persuasion to elicit immediate, identity-protective pushback.

While Stanley might not wholly be a Kantian, someone who adopts a Kantian perspective would likely argue it is immoral to use propaganda to persuade people. Propaganda is a nonrational means of persuading people of ideals - in this case of vegetarianism. This is incommensurate with the Kantian moral principle that one ought to never treat other people as means but as end in themselves. What this means here is that to persuade someone they should be vegan requires a provision of reasons; reasons which any rational agent ought to recognize as true. But insofar as propaganda is a method which excludes rational deliberation to convince

people of vegetarianism, Kantians would judge it a moral violation (Rawls, 1971; Hill, 1972). Propaganda bypasses the rational will and as such we shouldn't do it; propaganda is immoral.

I am generally inclined to agree that we should not bypass other agents' rational deliberation. In the following chapter I agree with Kant that one ought to respect others as rational agents, which precludes the usage of shame as a tool to change behavior in *most* cases. But choosing to consume other animals is not most cases. Vegetarianism is a limit case for moral rational persuasion because of the nature of how systematically nonrational people's judgements are. Judgments about the consumability of animals are systemically non-rational, not because they lack complete reason, but because they are shaped less by coherent logical, moral deliberation than by entrenched social identity norms, conceptual categories of consumption, and selective applications of fellow-feeling or sympathy. People frequently distinguish between animals that are “edible” and those that are “companions” or “wild,” but these distinctions are not grounded in morally relevant differences between species (something which May and Kumar agree with). At the same time, the capacity to sympathize with animal suffering is often restricted to animals perceived as near or “like us,” while others are kept at a distance through factors like the industrialized process which masks meats' animal origins. And of course, social identity norms create attachments to certain dishes for emotional and personal reasons - whether it be Grandma's pot pie or a Sunday roast.

My view is that the extent to which most humans can exercise their rational will when it comes to making moral judgements about consuming animals is minimal. Even if rational persuasion is effective to some limited extent, I think that ultimately we should be convinced by a utilitarian stance when it comes to how one ought to engage in persuasion here. The amount of good that can come from people eating less meat in general, for both humans and non-human

animals, I think should outweigh greatly the good that can come from persuading people on rational grounds eating (factory farm) meat is wrong. This is because I think persuasion on rational grounds is not even possible in most cases when it comes to getting people to consume less meat.

In this chapter I defended the empirical claim that morally persuading people to not eat meat is likely to cause backlash. I also defended the normative claim that nonrational persuasion or propaganda is permissible in the case of getting people to eat less meat. In the next chapter of this dissertation I examine the moral permissibility of using shame as a tool for norm change. I argue the worry that shame often disrespects agency and rationality, it is not a necessary feature of shame and that shame can be permissible when it aims at moral persuasion. This reflects my broader commitment to the importance of respect in social change.

Chapter 4 - Shaming is permissible when it aims at moral persuasion

4.1 Introduction

Shaming is a tool for social change in development contexts. In the Global North, people shame others for taking too many flights, for using too much water, for eating meat, and more. In the Global South, shaming has become part of NGOs' strategies to change norms around sanitation. An estimated 673 million people practice open defecation (OD) each day. OD is the practice of defecating outside such as in fields, rivers, alleyways, streets, and other "open" spaces. This contaminates water and food sources which leads to several severe health complications and life-threatening diseases such as diarrhea, typhoid, cholera and hepatitis. The goal of the United Nations is to eliminate OD completely and ensure global sanitation by 2030. But progress has not been as fast as one might have hoped. Even in places where latrines are constructed, their usage can be infrequent or rare. Since any contamination of water or food sources can cause disease, communities must be unanimous in their decision and usage of latrines. It became clear that eliminating OD was not just an infrastructural problem but also a cultural one (Adjibolosoo et al., 2020). With progress not fast enough to meet target levels of sanitation, an approach which gained widespread prominence is the 'Community-led Total Sanitation'. In 2011 there were an estimated 44 countries which adopted this approach. A major part of this approach is getting people to feel shame and disgust, but also to put in place practices of shaming each other to stabilize latrine usage (Kar, 2011).

The need for tightly adhered to sanitation norms is not unique to the Global South. By one estimate, the probability of experiencing a pandemic during one's lifetime is 17% but could grow to 44% in the next 20 years (Williams, Jones, Welch, & True, 2023). It is likely that shaming will be used more in situations like pandemics and environmental disasters when change is

deemed urgent. This is because shaming (the action) is an effective motivator of behavioral change due to its connection to self-esteem or self-worth. When an individual feels ashamed (the emotion), they perceive a devaluation of self, which can motivate them to change their behavior out of fear of social rejection.

The recent COVID-19 pandemic led to a resurgence people being shamed for masks, vaccinations, social distancing, handshaking, and more. This also led to a resurgence of the ethical debate about whether we ought to be shaming each other. Some researchers argued that shaming is often overlooked and intensified by government public health responses in the context of pandemics (Cooper et al., 2024). It is true that when large-scale cooperation is needed, those who dissent can be seen as shameless and selfish. We might think that 'water shaming' can make our neighbors stop wasting water, or that 'flight shaming' can make people stop travelling by plane so much. But other times shame can seem a morally unreliable way to punish (alleged) wrongdoers. So, should we be shaming people more often, or is shame a pernicious tool in social and moral development?

The view I defend in this paper is that shaming as a tool to change others' behavior or views is sometimes permissible. My argument is a response to philosopher Martha Nussbaum's view that shaming is incompatible with a commitment to human dignity because shaming others expresses a contempt for others and a devaluation of their worth. I show there are important exceptions to shaming's impermissibility such as when we shame someone for not being open to moral persuasion. In cases like this, shaming can act as a reminder that our interlocutors ought to engage with us as reasoners. Calling someone out for not being open expresses a respect for their rational agency, meaning that they should have control in deciding whether our reasons are relevant to changing their behavior or views. Instead of stratifying or denigrating as shaming

often does, shaming here functions to include people into a moral community that they close themselves off from. So, while in all other cases shame acts as illegitimate coercion, in this case shaming stems from a legitimate duty to call out for now engaging in this practice of rational exchange.

This chapters contains six sections and a conclusion. In §4.2 I distinguish shame the emotion from shaming the action. The focus of this paper is on shaming as a tool to change one's behavior or views – not primarily the emotion of shame. In §4.3 I consider some of the classical objections to using shame as a means for promoting adherence to the law. This section will draw heavily on the work of Martha Nussbaum, who argues that shame is inconsistent with human dignity. I will develop this point in two ways. First, using shame expresses a distrust of others' capacity to considers reasons for action by using esteem as leverage to change their will. Second, shaming expresses a disregard for others' agency in choosing how to act by forcing someone to act out of concern for social approval. In §4.4 I argue that despite this shaming is not necessarily morally problematic if it supports human dignity and that this is indeed the case with shaming someone to be open to moral persuasion. In §4.5 I argue that shaming is morally permissible in such cases because it expresses a respect for the norm violator, namely, that they deserve to be part of rational, moral discourse. In §4.6 I consider how moral standing undercuts the duty to be open to moral persuasion. This is relevant because for shaming to be permissible, it must be the case that someone is an accurate judge of when someone is closed off from persuasion, which often presupposes personal or cultural familiarity. Indeed, shaming *can* be morally permissible in developmental contexts, but only when one properly judges another to be closed to persuasion.

4.2 Shame and shaming

First, what is shame? Much philosophical discussion of shame as an emotion takes inspiration from John Rawls. Rawls characterizes shame as the feeling that occurs after a loss of self-respect or self-esteem (Rawls, 1971). Dan Zahavi (2012) argues that what is central to shame is that one experiences a "global" devaluation of one's self-conception. When you feel ashamed you feel ashamed about your whole person such that it makes you question who you are. This distinguishes shame from the less intense emotion of embarrassment. Feeling embarrassed is to regret what one has done or some aspect of oneself with it encroaching on one's self-conception. But once one starts to question one's self worth, Zahavi would argue we are talking about shame.

Rawls and Zahavi share a focus on how shame leads to a devaluation of self-worth without much regard to the role others play in shame. But there is reason to think that shame's sociality is important. Philosopher Heidi Maibom (2010) argues that those who experience shame are committed to living with others and experience a devaluation due to their failing to live up to public expectations. Shame can be a sign of a mature moral agent recognizing that they fall short of certain social ideals. But it may also be that one can experience shame even when one disagrees with those ideals or people's specific criticisms of when we fail to live up to them (Calhoun, 2004). This is important because it speaks to those experiences where we disagree with our shamer's judgement that one ought to be ashamed but still end up feeling ashamed because we recognize that they have standing to criticize us.

The act of shaming differs from the emotional experience. If shame is the feeling one has when you fail to meet certain expectations or standards, then shaming can be understood as the actions that intend to bring about this emotional experience or awareness of not meeting the

expected norms. But *shaming* also carries with it that the supposition that there is a reason why someone ought to be ashamed. So, shaming here must be understood not merely as a means to notify someone that they do not live up to group standards, but also that this act is performed with the intention that target of shaming will rectify their behavior or views out of concern for their self-worth. Examples of shaming include making people with DUIs drive around with bumper stickers stating their offense, posting pictures of shoplifters outside local stores, and creating an online list of people who don't pay their taxes. These diverse acts have different levels of threat to dignity or autonomy.

How is it that shaming can undermine a commitment to human dignity? In the next section, I will spell out how shaming can undermine dignity by disrespecting others' vulnerability, rationality, and agency.

4.3 Shaming is generally morally impermissible

Shaming those who violate norms risks significant harm and is not an unproblematic act. In this section, I will argue that shaming should be avoided as a means of changing people's behavior for two reasons—shaming a norm violator fails to respect their dignity and undermines their agency. To defend this claim, I draw on Martha Nussbaum's argument in *Hiding from Humanity* that shaming law breakers is inconsistent with a commitment to uphold human dignity (Nussbaum, 2006). I differ from Nussbaum by showing that shame need not necessarily be incompatible with a respect for those who we shame and human dignity more generally. This section establishes the conditions under which shame could be permissible, namely when it does not disrespect a norm violator's agency and rationality.

Nussbaum argues that shaming often undermines a commitment to human dignity. This is first and foremost because shaming can express a contempt for the target. For example, in England, under the poor act of 1697, the poor were made to wear a badge of blue or red cloth once they received aid. This deters individuals from seeking out relief until they were desperate but essentially humiliates them for being poor. Nussbaum objects to shaming in such a case because it makes someone feel bad not merely for what they've done but for who they are. This violates a commitment to the equal status of individuals, which dignity requires, by suggesting that people who violate norms are not just wrong but bad people.

Nussbaum's view then is that shaming is unacceptable when it humiliates, expresses a superiority, and causes another person or people's degradation of self-worth. Nussbaum argues that shaming symbolizes a contempt for someone or some group as well as a contempt for human reality. By shaming those who we see as different, wrong, or failing up to community standards, we reject the idea that it is natural for humans to be incomplete, needy, and 'under construction'. But all humans *are* imperfect, needy, and fail. When we shame others, we express contempt for this fact and strive for the unattainable ideal of completion, perfection, and needlessness. Shaming, as Nussbaum sees it, is an "infantile demand for omnipotence and the unwillingness to accept neediness", which is ultimately, "a way of hiding from humanity that is both irrational in the normative sense, embodying a wish to be a type of creature one is not, and unreliable in the practical sense, frequently bound up with narcissism and unwillingness to recognize the rights and needs of others" (2006, p. 15). Shaming would then be always the wrong thing to do, because it is an expression of the shamers' inability to face human incompleteness and neediness – a neediness which cannot be overcome or erased from human nature.

When it comes to shaming people to change their behavior, Nussbaum's worry that shaming undermines human dignity is clearly seen. But shaming is meant to change people's behavior, it can also undermine their agency. Consider traditional views such as Kant's that we ought to treat persons always as end in themselves and never merely as means. This means we should respect others by recognizing them as rational and free rather than as irrational, manipulable, or simply as means to our own ends. When we treat others as moral persons we respect these features of their humanity, which means we do not see them as means to our own ends, or as people incapable of responding to reasons.

Kantian philosopher Colin Marshall suggests that treating other as moral persons involves (at least) two distinct kinds of respect: "rationality respect" and "agency respect" (Marshall, 2024). First, rationality respect requires that we treat others as people whose actions stem from reasons and are owed reasons in certain situations. Certainly, trying to persuade someone to act differently or hold different views is one of those occasions. If I believe you should change your conduct or your moral views, then it is disrespectful of me to treat this as an occasion where reasons are optional. I must provide you reasons sincerely in a way that respects your capacity to reason, e.g., by not lecturing you or providing reasons that are superficial. If I lecture you then I do not respect your rationality, just as blindingly obvious reasons may also disrespect your rationality.

What respect for agency amounts to depend on what agency is. While accounts differ, a moderate view is that agency is any exercise of one's will to fulfill one's intentions. For example, your intention to stay at home on the weekend to finish your homework. It would be disrespectful of your agency to cajole you out to the bar and concerts all weekend long if I know your planned intention is to get your homework done. While not all our views and intentions

deserve respect, we ought to give respect to others' agency if their aims and plans do not harm others. For instance, if I know that you frequently spend your days playing video games instead of finishing your homework, then it would be disrespectful of me to hide your controller. Doing so would express a distrust in your capacity to be in command of how you use your time to accomplish your goals.

Similarly, when someone shames a person to change their views or behavior, they are not demonstrating trust in the person to make rational decisions on their own. Shaming, when defined as intending to bring about a negative evaluation, merely coerces someone into changing their views, in order to regain the shamer's their esteem. Holding someone's esteem hostage to persuade them of something is not respectful of their rationality. Shame bypasses consideration of determinate reasons for holding a certain view, which means that it does not offer the persuaded a chance to change their view due to the consideration of reasons why they ought to. This is true even if I offer you reasons to change your view. For example, say I try to convince you that littering is always wrong. I offer you several reasons why this is the case, but I see that you are still reluctant to give up your view. If I shame you by telling you I no longer want to hangout with someone who litters, after providing you reasons to change your views, then I still disrespect you and your agency. I did not respect your capacity to consider the reasons I provided you on their own terms. Instead of allowing you to make up your mind and hoping that the reasons I offered will convince you, I resorted to shame. This expresses a distrust of your capacity to make the 'right' decision based on the reasons I provided. Once I use shame, I coax you into believing littering is wrong, as opposed to letting you make up your mind. Even if you change your views now, I tarnish the possibility that it was your choice. Instead, you and I will both know that shame possibly played a role instead of these reasons, i.e., that you changed your

views to garner my approval instead of changing them because you found the reasons convincing.

But while shaming is generally morally unacceptable in the above ways, Nussbaum gestures towards ways that it could be permissible. That is, shaming might be permissible if it does not humiliate or coerce. We can add to this that for shaming to be acceptable it must also express an equal worth by seeking to reintegrate and repair equality between persons rather than denigrate another's self-worth. So, instead of shaming *always* being morally impermissible, one might object that shame isn't going to be morally problematic in cases where people are not open to moral persuasion. This is because trying to get someone to be open to persuasion seeks to show them that they are in fact a worthwhile participant in moral discussions. Furthermore, because the practice itself is one of rational exchange, shaming someone to be open doesn't coerce them. I think there is something right about this objection, and on this point, I disagree with Nussbaum that shaming necessarily undermines a commitment to human dignity. On the contrary, shaming can be used to try and establish equal status and respect among people. In the next section I will begin to unpack this objection by showing why people have a duty to be open to moral persuasion.

4.4 Everyone has a duty to be open to moral persuasion

Moral persuasion is the attempt to change someone else's views about moral matters through the provision of reasons. Two features are key to moral persuasion—the topic on which one is being persuaded is a clearly moral one (persuading you to take Vitamin D for your health does not qualify) and the means of persuasion must be through considering reasons for action. Proffered reasons are meant to 'stand on their own' and be the source of why one changes their views regarding a certain moral topic. Cajoling, coercing, using rhetoric, or even changing one's

views for no reason at all do not therefore count as moral persuasion. Of course, shaming would not quality as moral persuasion, either.

Philosopher Regina Rini (2018) argues that there is a general obligation to be open to moral persuasion. “Being open” to moral persuasion involves engaging with reasons in a sincere manner. Sincerity here requires that the possibility of changing one’s views is on the table. It also means that one actually engages with reasons provided to them and whether they may be relevant to the views that the persuader seeks to change. To see that engagement with reasons and sincerely being open to change are two distinct components of being open to moral persuasion consider the following two examples of persuading someone to not litter.

The first case is one where the person is open to changing their views but does not sincerely engage with your reasons. You approach the litterer with reasons for why they shouldn’t litter: it is bad for the environment, every bit makes a difference, that no one is exempt from helping despite many people littering, and it is not good for local wildlife either. Imagine again giving the litterer reasons why they should not litter, and they respond with: “None of that matters because people litter all the time, so my contribution makes no difference to climate change either way”. One way of interpreting this response is the litterer was sincerely open to moral persuasion, but since they bluntly contradict one of your reasons, they weren’t actively listening and engaging with the reasons you provided.

The second case is one where the person is not open to changing their view but does sincerely engage with your reasons. Imagine now the litterer responds: “You make a good point about everyone should help and that it must be bad for the local wildlife. But none of that really matters since I am going to keep thinking littering is fine no matter what you say.” We may interpret this case as one where the person was closed to being persuaded but open to engaging

with the reasons they were offered. While they stated they would hear you out, it was not a sincere statement that they might reconsider their views in light of the reasons you offered them. The litterer excluded the possibility of changing their views and thereby failed to be open to moral persuasion.

These cases show that sincere engagement and willingness to change one's view are distinct parts of being open to moral persuasion. Judging whether someone is open to moral persuasion requires careful interpretation. It is difficult to know at the best of times whether someone is sincere in being open to change their views or if they weigh the reasons provided. But there are also clear enough cases to suggest that it is not unreasonable to know when people fail to be open to moral persuasion.

It is important to note that being open to moral persuasion doesn't require that the interlocutor change their views. They may respond: "Look, I know littering harms the environment and animals and that I may have reasons to think that what I do makes a difference. But I still believe I have bigger fish to fry and that the other activities I engage with in my life are more important or make a greater impact than the little amount I litter." In this case the litterer seems sincerely open to moral persuasion. They show that they have understood the reasons you (the persuader) offered, and they seemed sincere in their attempt to consider how your reasons might be relevant to changing their own views. But they state that those reasons were not successful in persuading them. Being open to moral persuasion does not require one to change their views, just that they are open to being persuaded and sincerely consider the reasons offered.

With this account of what it means to be open to moral persuasion on the table, we can now turn to the argument that there is a duty to be open in this way. The obligation to be open to

moral persuasion that applies to all free and rational agents is grounded in our nature as free and rational agents who interact with one another. Philosopher Stephen Darwall (2009) calls this the "fact of reason", meaning that implicit in engaging in moral persuasion is a type of mutual recognition about each other's ability to engage with reasons. By giving you reasons with the intention of persuading you of a different view, I engage with you and recognize you as a person capable of understanding these reasons and evaluating whether they should change your views on the topic at hand. And if you are open to my moral persuasion, then you return a kind of respect which I already offer to you. It is due to this relation between us as two rational and free agents that certain obligations and norms apply to how we ought to treat other and whether our conduct is respectful or not. The duty to be open to moral persuasion is what is required to respect each other as rational agents.

Rini argues that this obligation to be open to moral persuasion is an imperfect duty in the Kantian sense (2018, p. 5). There are limitations to how much and how often we can engage in moral persuasion as we are finite beings. As finite beings we must also attend to others matters than our moral views. It is certainly valuable that part of what it means to be human is to participate in a moral community. But it is simply not possible that we spend all our time entertaining others' persuasion or engaging in persuading others.

Part of what makes the duty to be open to moral persuasion imperfect is that there may be certain relations between individuals that interfere with this duty. For instance, Rini notes that we do not owe it to a figure like R.M. Hare's fanatic to be open to their persuasion given that we know the duty will not be reciprocated (Hare, 1963, p. 105). The fanatic, according to Hare, is a person so attached to a moral conviction that they are unwilling to entertain reasons or views which come into tension with it. This plausibly interferes with the duty to be open to moral

persuasion since it is mutual and reciprocal. I cannot go around only persuading people of my own views but never listen to their reasons for why I should hold their views. There is a limit to how open we should be to someone who persuades us that has no interest in reciprocating.

In addition to avoiding wasting time on fanatics, some have thought that membership in a moral community might be another principal way to limit our obligation to be open to moral persuasion. Perhaps shame's propriety depends on the relationship between the persuader and those who they persuade more generally. For instance, hypocrites plausibly violate the reciprocal nature of persuasion because they make an exception of themselves. Instead of being open to persuading each other and endorsing an equality between us, hypocrites make an unjustified exception of themselves and thereby eliminate their standing to blame (Fritz & Miller, 2015). This, like the fanatic, plausibly violates the nature of reciprocal openness and thereby limits the obligation to be open to such a person's persuasion as well.

Given that people have a duty to be open to moral persuasion, how should we respond to those who fail to live up to this duty? Is shame a viable response in this situation? In the next section I argue that it is indeed in some such cases, and not others, that it is permissible to respond to such failures by shaming the person who flagrantly violates the duty to be open to moral persuasion.

4.5 It is permissible to shame those who are closed to moral persuasion

While Nussbaum provides an argument that shaming is never morally permissible, the case of those closed to moral persuasion offers an interesting challenge to that universal prohibition. I will argue that under circumstances in which an interlocutor is closed to moral persuasion, shaming can avoid all of Nussbaum's objections. Shaming can sometimes be non-

coercive, as well as respect human dignity and autonomy, when it notifies others that they fail to be open to moral persuasion. In such cases, it is in fact a duty to attempt to reintegrate those who violate norms of being open to moral persuasion. When a rational agent is closed to moral persuasion, the act of shaming becomes a communicative act that notifies others that they failed to fulfill this duty. It also acts as 'proleptic reason' for the person to be open to moral persuasion, namely, it gives the person being persuaded a reason to be open out of concern for social approval (Williams, 1989). The approval of the persuader offers a temporary reason to be open until an adequate respect for the duty to be open is acquired.

Recall what the duty to be open to moral persuasion is. It involves a sincere engagement with reasons which the persuader provides and a willingness to change one's views in light of the reasons offered. Now consider a case where your neighbor engages in domestic abuse.[2] You try to offer your neighbor reasons to treat their partner better and they are not interested in hearing you out at all. Eventually they say, "There is nothing you can say which will change my mind". You can see that even if they were to let you provide them reasons why they should stop abusing their partner that they are unwavering on their views about the topic.

This plausibly constitutes a failure to be open to moral persuasion. The domestic abuser is adamantly unwilling to engage with your reasons, and the abuser remains closed to the possibility that they will change their views from jump. The question is the following: we know that it is everyone's obligation to be open to moral persuasion, so how should we respond to such

[2] I have adapted this case from an influential one offered by Bernard Williams in: Internal reasons and the obscurity of blame. In: *Making Sense of Humanity: And Other Philosophical Papers 1982–1993*. Cambridge University Press; 1995:35-45.

an egregious failure on behalf of people like the domestic abuser, namely those who are firmly closed to moral persuasion?

The response I will argue for is that in such a case it is permissible to respond by shaming the person. In this case, we may shame the domestic abuser not directly for how they treat their wife, but because they are blatantly closed to moral persuasion. Before I argue why shame is appropriate here, it is important for my argument to keep in mind the distinction between notification and coercion (Darwall, 2009). Stephen Darwall argues that coercion undermines respect for other agent's capacity for reasoning whereas notification does not.

When someone threatens you by saying that something bad will happen if you do not oblige, then they create a reason for you to act in a specific way. But this can be done without the proper authority or legitimacy as is the case when a school bully threatens to hurt you if you do not give up your lunch money. The bully has no right to your lunch money and as such the reason they provide to you is irrational and illegitimate. When there is legitimate authority that all parties involved recognize, or should recognize, then a threat is a notification (Darwall, 2009, p. 49-52). For instance, if you have the authority to sanction someone, as a teacher does to the bully if they take someone else's lunch money, then informing the bully they will experience consequences if they do take the lunch money is not an irrational threat or coercion but rather a notification.

The fact that being open to moral persuasion is everyone's duty and that a norm violator fails to fulfill this duty puts the legitimate authority in an interesting position. They have a duty to respect the norm violator's dignity, agency, and rationality, even when the norm violator is not currently exercising their rationality. What might a legitimate authority do in such a context? One option is to avoid pressuring the person into being open, but another option is to attempt to

shame them into being open. The latter is preferable because one might think it is in fact a duty to shame the person into participation Shaming someone for being closed tells them that they ought to be included in the moral community and expresses a kind of respect for their opinion. It is a responsibility humans have to each other to include each other in the social practice of moral discourse, so shaming here, as an invitation to join in rational exchange, can be seen as fulfilling a responsibility to include others in a tradition of respect.

It is important to note however that judging someone to be closed to moral persuasion is not always easy. There is a conceptual distinction between flagrant and slight violations of the obligation to be open to moral persuasion, but in practice things are more complicated. This comes down to the degree to which we judge our interlocutors to be earnestly open to and engaging with the reasons provided during persuasion. The domestic abuser who patently refuses to hear you out or has made up their mind to not change their views in advance is a flagrant violator. But, again, the duty to be open to moral persuasion is imperfect. Individuals do not have unlimited time and effort to engage with all those who persuade them to change their views, which means that it may be difficult to tell apart cases where someone currently has little mental resources or time to engage with a persuader's reasons from cases where someone habitually puts off considering a persuader's views. This will typically require keen judgement and knowledge of our interlocutors. Egregious violations of the duty to be open to moral persuasion are liable to be shamed, i.e., we can shame those who continually fail to engage or are rarely open to changing their views in light of moral persuasion.

If shaming is permissible, then it must not be coercive. It also must be respectful of others as free and rational agents as well as of human dignity more generally. It is possible for shaming to satisfy these moral desiderata when it is in response to the failure of a duty like moral

persuasion. That is because shame here acts as a notification that the persuader is treating the norm violator as if they are a rational agent who can consider reasons for action even when they are not currently doing so. Shaming others is therefore compatible with respecting particular individuals who we interact with by seeing their worth and actively working towards including them rather than express contempt at their shortcomings. In the sense which I currently discuss, shame is an invitation to feel guilt.

This last point suggests that shaming is compatible with Nussbaum's moral concerns, which is that shaming violates a respect for human dignity based on the equality and vulnerability of individuals. When shaming is used as an inclusive force to help someone participate in moral discourse, it expresses a recognition that morality is a social practice which is open-ended and incomplete. Reintegration is key to the pursuit of valuing humans as moral creatures because humans to be moral we must practice it by persuading one another. While shaming might typically act as an exclusionary force that expresses a childlike urge to do away with neediness, shaming here acknowledges that humans are needy and need each other. Thus, shame in this instance is respectful and compatible with human dignity.

It will help to illustrate these points with the case of the domestic abuser. I attempt to persuade the abuser that what they are doing is wrong, but they blatantly refuse to hear me out or really consider their views in light of the reasons I provide. I then tell them that they have a duty to be open to moral persuasion, which amounts to the provision of another reason why they should hear me out. It may be that they lend me their ear in response to hearing explicitly about this duty, but if they stick to what they have already shown, which is that they are not open to being persuaded, then they leave no pathway that the provision of reasons will succeed in traversing. In such cases, they are a suitable target to be shamed. But why shame?

Bernard Williams (1995) argues in "Internal reasons and the obscurity of blame" that the desire for moral esteem does motivating work. For instance, in the case of the domestic abuser, they may see nothing wrong with abusing their partner but still seek out the approval of others. While one would hope that the abuser shares their views, it is possible to use your own approval and esteem of them as a temporary substitution. It may seem that drawing on an individual's concern for their esteem to get one's foot in the moral door is inappropriate. Perhaps we should keep giving the domestic abuser reasons why they ought to be open to moral persuasion instead of coercing them with shame to do so. But in cases where people flagrantly violate this duty, it is unlikely that they will be responsive to additional reasons for why they should be open to being persuaded. Avoiding disapproval or gaining approval can act as motivation to lead someone to change their views or conduct in the absence of directly acquiring reasons for why they should hold a certain opinion or conduct themselves in a specific manner. Since the abuser does not care to hear your views about the matter at hand, they may be more willing to lend you their ear if your approval is on the line. There is a wealth of evidence which supports the hypothesis that humans generally care about their reputation and will go great lengths to garner social approval (Baumeister & Leary, 1995; Bastian & Haslam, 2010).

We can now appreciate why shame is a suitable tool to reprimand someone who seriously violates their duty to be open to moral persuasion. It is suitable because it leverages humans' general desire for approval or esteem to get them to care about their lack of openness to moral persuasion. It respects them as persons with whom they can communicate and who can change their opinions. In other words, shaming respects them as someone who is not the fanatic, or someone not worth bothering with. As the duty to be open to moral persuasion is an imperfect one, we can be sure that those who never live up to it are failing a duty that they owe to their

interlocutor. Shame gets our foot in the door to open those individuals up to moral persuasion. It acts as a "proleptic invocation of a reason", i.e., a temporary reason for the violator to fulfill their duty to be open to moral persuasion until they have a direct regard for this duty (Williams 1995: 41). In the case of the domestic abuser, they are closed to moral persuasion and lack motivation to engage with your reasons. The hope is that by suspending our approval for this person, they come to fulfill their obligation out of concern for their esteem. Once they begin to fulfill their duty, then, ideally, they will come to regard it as a worthy duty and no longer need the motivation of others' esteem.

The legitimacy of shaming in such cases comes from the harm to oneself and to the interlocutor when they blatantly refuse your attempt at persuasion. This harm creates a duty, namely, a duty to notify this person that they fail to fulfill their obligation to be open to moral persuasion. Since everyone has a duty to be open to moral persuasion, when there are instances where someone habitually or seriously violates this duty, we have the authority to call them out and tell them that we are owed their sincere engagement. Shame is generally impermissible because there is no duty to shame someone for their direct offense of harming a third party. Nussbaum shows us this when she argues that shaming undermines human dignity. But when someone harms you by refusing your attempt to persuade them, it does not undermine their dignity to shame them. Instead, it creates a duty you did not previously have to shame them. It would be a challenge to their dignity to give up on them and walk away and call them crazy or a fanatic much more than it would be to shame them into considering reasons for their actions.

One might worry that shame seems an odd tool to use when it can stifle second-personal engagement. After all, those who feel ashamed typically want to hide from sight. Upon discovering that he has killed his father and married his mother, Oedipus exclaims: "O, I adjure

you, hide me anywhere/ Far from this land, or slay me straight, or cast me/ Down to the depths of ocean out of sight." If shame can lead our interlocutor to withdraw further (from persuasion), then why try to use it to get them to open up?

It is certainly true that shame can lead to withdrawal, but when those who are shamed acknowledge shame's legitimacy, then shame can become internalized leading to guilt. The distinction and transition from shame to guilt are important for understanding how shame does not simply turn people away from being open to persuasion. When someone feels the emotion of shame it is typically third-personal and anonymous: shame is what *anyone* must do in such and such a situation. In this case, any interlocutor who is firmly closed ought to be open to being persuaded. If we shame someone for failing to live up to this duty, they may judge that the persuader has no standing to call them out, perhaps because they do not know each other well, or because they come from different societies which share different evaluative frameworks about what being open to moral persuasion looks like. But of course, they might also feel that the criticism is fair, and they have not been open to persuasion as they ought to have. Now the shame they feel becomes guilt. Guilt, Stephen Darwall argues, is "to feel as if one has the requisite capacity and standing to be addressed as responsible… it acknowledges the authority to make a demand, that is, to address a second-personal reason for acting." (2009: 71-72). If I shame the domestic abuser and make them feel guilty for their lack of openness, then that they feel is that *they* owe it to *me* to be open to being persuaded. The moral debt can be repaid by letting me persuade them of my views. This vindicates the usage of shame and addresses the worry that shame stifles second-personal engagement.

4.6 Exceptions to the duty to be open to moral persuasion

Does moral standing somehow limit the obligation to be open to moral persuasion? There are several ways that the relationship between interlocutors can plausibly place restrictions on whether someone owes someone else openness and consequently whether one has a duty to shame someone for their lack of openness. If one is skeptical that there are standing norms, i.e., that standing norms do not exist or that there are no facts about free and rational persons which could exempt them from moral criticism, then the duty to be open to moral persuasion is unchanged and it is always permissible to shame others for being. But it is worth considering how the existence of certain standing norms might interfere with who owes who openness to moral persuasion, and consequently, how this affects the permissibility of using shame in such circumstances.

Standing norms regulate who is permitted to perform some speech act or interfere with others' affairs. What is special about standing norms is that they are second-order norms which are *ad hominem* - they refer to facts about the intervener and not the intervention itself (Herstein, 2017). The standing norm most often discussed is non-hypocrisy. That is, if one has standing to blame others then they must not be guilty of having done the same or similar things as the wrongdoer. And this can be pushed further to the idea that to have standing is to be appropriately committed to the value under consideration (Todd, 2017). For instance, imagine a persuader is not committed to the value in question, say that we ought to keep our local parks litter free, then it is possible that there is no obligation to be open to their persuasion if it happens to be about how you should not litter in your park.

Recall R.M. Hare's character the fanatic. They are someone who happily promotes their own views, but are so utterly convinced by them, that they have no time for others'. Plausibly, why

we do not owe it to the fanatic to be open to their persuasion is because they are hypocrites about it (persuasion). Fanatics expect others to be open to being persuaded without expressing an openness themselves. Since the fanatic has no intention in changing their own views, they plausibly lack a commitment to the value of participating in rational exchange, and as such lack the right standing norm to shame anyone for their lack of openness.

Notice that the fanatic is someone whose unbudging moral views plausibly stem from a lack of willingness to change. But part of moral persuasion is not just an openness to changing one's views, but also an engagement with the reasons the persuader offers. This is why someone changing their views just to appease the persuader does not count as moral persuasion. We might call someone who is willing to change their views but fails to engage with others' reasons when persuaded the flippant. They, like the fanatic, lack the proper respect for the duty of being open to moral persuasion. Since they do not seriously engage with others' views, they plausibly also lack the standing to demand others be open to their persuasion. Although it is clearly permissible to shame the fanatic or the flippant for their lack of engagement or openness to persuasion.

This brings us full circle to the permissibility of shame. Shame is an acceptable response to someone who we *appropriately* judge to be closed to moral persuasion. Everyone ought to be open to moral persuasion, but judging *correctly* whether someone is open or closed is difficult indeed. It might be that you just happen to catch someone at a bad time, or it might be that they purposefully deflect your persuasion and actually fail to be open. What the existence of standing norms might show is this: just as persuaders need to know their interlocutors to judge whether they fail to fulfill their obligation, so too do those who are being persuaded.

Judging whether someone is closed to persuasion or not might depend on the familiarity with those who we persuade in two senses. First, violations of openness do not always take the form

of someone downright refusing reasons but dodging or deflecting them subtly over time. In such a case, some relationship is required between interlocutors to judge whether someone's deflections are habitual and representative of a more serious closure to persuasion. Another sense of familiarity is cultural rather than personal. While the exchange of reasons is universally human, its form is not. Some familiarity with how openness and reason-giving are enacted culturally is therefore important to judge whether someone is closed. Storytelling, metaphor, and proverb might not be recognized as reason-giving by someone who is used to persuasion as an explicit exchange of reasons. So, whether one has the authority to shame another for their lack of openness must suppose some familiarity with the form reason exchanging takes.

Does this mean that shame can never be morally performed by a stranger? Not necessarily. Just because someone is unfamiliar with who those who they persuade, does not mean that there is a complete inability for them to judge someone who is closed to persuasion. A complete lack of perspective taking, reflecting, or paraphrasing, could be signs that your interlocutor fails to engage with your persuasion. As could interrupting or not acknowledging your interlocutor addressing you. Ultimately, it is an open question how much of being open or closed to persuasion is culturally specific or universal. This will affect how culturally familiar one might have to be to morally shame someone for being closed.

What does this mean for the cases of shaming in developmental contexts such as eliminating open defecation? I believe that shaming can be an acceptable tool in developmental contexts. Shame can be used in cases where people are obviously closed to persuasion. In some cases, familiarity with the culture or person might be required to properly judge if someone is closed to persuasion. But if one can be reasonably assured that those who they engaged are closed, then we can be assured that shame is permissible and moreover a duty to reintegrate the person as a

valuable, contributing member of the moral community. This offers valuable information about how aid workers approach persuading communities they are unfamiliar with, as well as how one might advise community leaders to address their own community. Shaming is not inherently bad or unacceptable; there are times when it serves a noble purpose of helping others see their value in participating in moral discourse. But judging whether someone is genuinely closed to persuasion is likely a fickle art that depends on context. Aid agencies would do well to approach the usage of shaming delicately and to reserve it for situations that call for persuasion but where interlocutors lack openness.

4.7 Conclusion

In this chapter, I put forward the novel view that shame is permissible as a response to individuals who fail to fulfill the duty to be open to moral persuasion. Being closed to moral persuasion creates a duty for the persuader to shame their interlocutor. This is because the intention of shaming in such a case is to integrate the person as a participating member of the moral community, which means shaming here is done with respect for human dignity. This offers a principled approach to when shaming is acceptable more generally. If shaming is reintegrative and respects human dignity, then it may be acceptable. A caveat on this is that if norms of standing exist which exclude one from having to be open to persuasion, then it would also follow that shaming is no longer permissible. No corresponding failure to fulfill the duty to be open to moral persuasion would occur and, in turn, no obligation on behalf of the persuader to shame their target would be present either. Future research from anthropology or sociology which refines our understanding of how persuasion occurs and how openness is enacted may shed light on what counts as being closed and whether one is in a position to shame.

Chapter 5 - Conclusion to the dissertation

My dissertation offers important additions to the standard social norm literature. One lesson is that not all social norms are the same; some social norms are personally meaningful because they are tied to social identities and the desire to self-categorize in terms of social groups. I also make significant progress on contemporary ethical questions regarding the permissibility of tools of social change, particularly shame and propaganda. Shame and propaganda can be permissible in the right circumstances – which are counterintuitive conclusions given orthodox views that one ought not to use propaganda or shame to change norms.

How does the social identity approach offer a contribution to the standard social norm literature? My analysis of social identity norms offers an important corrective to Bicchieri's analysis. It is a corrective because Bicchieri's framework for analyzing types of norms involves a division between norms followed for personal reasons and those followed for social reasons. But social identity norms are both and as such it is unhelpful to begin categorizing norms on this assumption that norms are wholly either social or personal because they can be both. One instance in which I have shown paying attention to the personal dimension of norm adherence matters is when norm change threatens the freedom to express one's identity or as a threat to one's identity. Aside from this discrepancy in how one diagnoses and measures what norms exist, I take my framework to be compatible and a useful addition to Bicchieri's framework. Some norms are wholly contingent on what others do or believe, but others are more personal and depend on an individual's values or identity.

Another contribution I make is to show that the standard social norm literature has it backwards when it comes to why many people follow social norms. The social identity approach

states that we do not follow norms simply by following what others do or believe. Instead, the groups we want to belong to and see as part of who we are play a vital role. Social groups play an important role in one's self-image, which offers a psychologically realistic interpretation of why people follow social norms. That my account is psychologically realistic makes it different than Bicchieri's rational reconstruction and useful for norm change interventions where it is crucial that we understand how people think about norms.

On this point I also differ from Charlotte Witt's recent account of social role norms – norms which people follow by virtue of the role they occupy. My account differs from Witt's here because her goal is not to provide a psychological account of why people follow norms, but rather to discuss why norms have their normative force to begin with. The origins of normativity can be traced back to social roles and their institutions. What one ought to do is a function of the role they must play within society is the basic idea. Since the account does not seek to account for the psychological dimension of norm adherence, norm change is focused on a revision of roles and institutions. There is no doubt that over history roles and institutions change dramatically due to critiques about what they ought to be – for instance, perhaps the role of women and men are the most dramatic recent examples here. But it is unclear how this insight can be applied to questions about to deliberately change norms, or if it even applies to certain norms. For instance, how can being critical of social roles or institutions get people to want to use less energy, or to eat less meat? Some roles people are critical of and cause norms surrounding the role to change, for instance gender roles in society, but it is not clear that there is any role attached to certain norms. For instance, is there a role of a person who eats meat or is it just something someone does to fit in, or what people are used to doing around here? I think eating meat is often a social identity norm – something one does around here *and wants to* as part

of the group – rather than a role norm one has to conform to. Put differently, while some norms are adhered to simply because we occupy a social role, others we care about and want to identify with.

More generally, the social identity perspective is an important contribution to the social norms literature because it offers us this robust psychological perspective on why people follow norms – one we do not get from the choice theoretic perspective in Bicchieri's research or institution-focused perspective from Witt. There is a wealth of evidence on why humans categorize themselves according to social groups and this is relevant to norm adherence and norm change. One important future direction for norm change will be to look at the influence of *prototypical* in-group members. These aren't necessarily formal leaders or trendsetters in Bicchieri's vocabulary because they don't have any more power or status than the rest of the group. Rather, they are individuals who are seen as embodying the group's core values or characteristics. Changing their actions can lead to a modification of perceived in-group consensus, or what one ought to do as a member of the group, which can be a powerful driver of group-level change. People are more likely to adopt a new norm if they believe it's already widely practiced or approved of by their relevant in-group.

Another future direction is to apply the theoretical tools of social identity theory to norm change. For instance, one reason that people desire to self-categorize and identify with social groups is because it distinguishes them from others. The desirability of in-group norms is likely moderated by if they perceived as distinct to that group. For instance, it was found that the acceptability of bullying or fairness was moderated by whether the norm was unique to an in-group. When in-group norms (bullying or fairness) were perceived as too similar to those of other groups, one might perceive group distinctiveness to be under threat and find the norm in

question to be less desirable (Ojala & Nesdale, 2010). We know according to social identity theory that individuals are motivated to maintain a positive self-image by differentiating themselves from others and their social groups.

An important limitation of my dissertation is how norm adherence and change is complicated by multicultural societies where people hold multiple social identities. But the norms which are acceptable for some social groups are highly unacceptable to others. For instance, eating vegetarian in some groups is expected and in others blasphemy. This means handling tensions in norms across groups and advising how people to navigate this tension will be crucial for understanding how to promote adaptive norms such as getting people to eat less meat or consuming less alcohol. This raises interesting questions about the ethics of norm change. For instance, in cases where modifying social identities is difficult, should we expect people to change norms and violate what one ought to do as a member of a certain group? Or should we find it acceptable for people to adhere to norms contingently depending on their social group?

The next natural step for research on the permissibility of propaganda for reducing meat consumption is how we should be persuading people on nonrational grounds. This is in part an empirical question: it is likely that some propaganda will cause just as much backlash as rational persuasion. How exactly propaganda should be used is also an important ethical question. One way forward that seems ethical and effective is to try and undo propaganda that promotes meat consumption, or to focus on using propaganda to promote vegetarian dishes by appealing to its health or environmental benefits without suggesting there is something morally wrong with current practices. What seems clear is that it will be difficult to clearly distinguish between responsible persuasion and misleading propaganda. Can vegetarians use any kind of propaganda,

even if it knowingly spreads misinformation about animal treatment, how dangerous factory farming is for the environment, or how dangerous consuming meat is? These are important questions which I do not explore. To address these issues, future research should aim to clarify ethical boundaries and create guidelines for responsible nonrational persuasion.

Research on rational persuasion also stands to benefit from further philosophical research. It is especially important to speak to develop further resources to think about whether it is permissible for 'outsiders' to shame locals. To answer such questions requires a principled approach to what constitutes an outsider. And developing an answer is important for downstream questions about judging whether someone is open to persuasion. If someone judges poorly that another person is closed, because the customs surrounding moral persuasion are different than what they are used to, then they may feel (wrongly) that that person ought to be shamed.

Shame that aims at moral persuasion is unique because it respects others' agency and rationality. I focused on this because it is an interesting case of when shame is compatible with a commitment to human dignity. When shame is constructive it can express a need for the other. When shame is destructive it expresses a resentment and disapproval of neediness altogether. And shame that aims at moral persuasion does not just say "I need you to engage in this practice" but also aims at reparation between the two parties. Theoretically there are other cases where shame is permissible. For instance, shame might not aim at moral persuasion as such, but as a way to achieve reconciliation between conflicting groups or classes. We need to be careful here that shame is not deemed permissible just because it undermines narcissism and aims at moral ideals liberals tend to favor. Even if the ideals shame is used to promote are noble, we should be weary that shame is not a substitution for rational deliberation.

Bibliography

Adjibolosoo, S. V., Ofori, B. D., Adongo, P. B., Afranie, S., & Yirenya-Tawiah, D. R. (2020) 'The influence of culture on open defaecation in some basic schools in selected districts in Ghana: A preliminary study'. PloS one, 15(10), e0239491. https://doi.org/10.1371/journal.pone.0239491

Aigner, A. (2021). "Don't push away your emotions – Cora Diamond and animal ethics". In *Justice and food security in a changing climate*. Leiden, The Netherlands: Wageningen Academic. https://doi.org/10.3920/978-90-8686-915-2_30

Barth, F. (1969). Ethnic Groups and Boundaries: The Social Organization of Culture Difference. Oslo: Universitetsforlaget.

Bastian, B., & Haslam, N. (2010) 'Excluded from humanity: The dehumanizing effects of social ostracism'. Journal of Experimental Social Psychology, 46(1), 107–113.

Baumeister, R., & Leary, M. (1995) 'The need to belong: Desire for interpersonal attachments as a fundamental human motivation'. Psychological Bulletin, 117, 497–529.

Bichierri, C. (2016). *Norms in the wild: How to diagnose, measure, and change social norms*. New York: Oxford University Press.

———. (2006.) *The grammar of society: The nature and dynamics of social norms*. Cambridge: Cambridge University Press.

Bichierri, Cristina., & Peter McNally. (2016). "Shrieking Sirens. Schemata, Scripts, and Social Norms: How Change Occurs." *PPE Working Papers*, Philosophy, Politics and Economics, University of Pennsylvania.

Bicchieri, Cristina, Ryan Muldoon, and Alessandro Sontuoso, "Social Norms", *The Stanford Encyclopedia of Philosophy* (Winter 2023 Edition), Edward N. Zalta & Uri Nodelman (eds.).

Benningstad, N. C. G., & Kunst, J. R. (2020). "Dissociating meat from its animal origins: A systematic literature review." *Appetite*, *147*, 104554. https://doi.org/10.1016/j.appet.2019.104554Bicchieri, SEP

Boenke, L., Panning, M., Thurow, A., Horisch, J., & Loschelder, D. (2022). "Who can nudge for sustainable development? How nudge source renders dynamic norms (in-)effective in eliciting sustainable behavior." *Journal of Cleaner Production*, *368*, 133246.

Boyd, R., & Richerson, P. J. (2005). *The origin and evolution of cultures*. Oxford University Press.

Brehm, J. (1966). *A theory of psychological reactance*. Academic Press.

Brehm, S., & Brehm, J. (1981). *Psychological reactance: A theory of freedom and control*. New York: Academic Press.

Brennan, G., Eriksson, L., Goodin, R., & Southwood, N. (2013). *Explaining Norms*. Oxford: Oxford University Press UK.

Calhoun, C. (2004). An apology for moral shame. Journal of Political Philosophy, 12(2), 127-146.

Cancian, F. (1975). *What are norms? A Study of beliefs and action in a Maya community*, Cambridge: Cambridge University Press

Chudek, M., & Henrich, J. (2011). Culture–gene coevolution, norm-psychology and the emergence of human prosociality. *Trends in cognitive sciences*, *15*(5), 218-226.

Cislaghi, B., & Heise, L. (2019). Using social norms theory for health promotion in low-income countries. *Health promotion international*, *34*(3), 616-623.

Cooper, F., Dolezal, L., & Rose, A. (2024) 'Shame-sensitive public health'. Journal of Medical Humanities, 46, 59-73

Crocetti, E., Pagano, M., De Lise, F., Maratia, F., Bobba, B., Meeus, W., & Bacaro, V. (2025). Promoting adolescents' personal and social identities: A meta-analysis of psychosocial interventions. *Identity*, *25*(2), 167-196.

Darwall, Stephen (2006) The Second-Person Standpoint: Morality, Respect, and Accountability. Harvard University Press.

Davies, A., & Evans, D. (2019). Urban food sharing: Emerging geographies of production, consumption and exchange. *Geoforum*, *99*, 154-159.

DeGrazia, D. (2009). Moral vegetarianism from a very broad basis. *Journal of Moral Philosophy*, *6*(2), 143-165.

Diamond, C. (1978). Eating meat and eating people. Philosophy, 53(206): 465-475. https://doi.org/10.1017/S0031819100026334

Ellickson, P. L., Bell, R. M., & McGuigan, K. (1993). Preventing adolescent drug use: long-term results of a junior high program. *American journal of public health*, *83*(6), 856-861.

Feinberg, M., Kovacheff, C., Teper, R., & Inbar, Y. (2019). Understanding the process of moralization: How eating meat becomes a moral issue. *Journal of personality and social psychology*, *117*(1), 50.

Fritz, K. & Miller, D. (2015) 'Hypocrisy and the standing to blame'. Pacific Philosophical Quarterly, 99(1), 118-139.

Gilbert, M. (2008). Social convention revisited. *Topoi*, *27*, 5-16.

Goldstein, N. J., Cialdini, R. B., & Griskevicius, V. (2008). A room with a viewpoint: Using social norms to motivate environmental conservation in hotels. *Journal of consumer Research*, *35*(3), 472-482.

Green, S. A., Smith, B., & Mathur, M. (2025). Meaningfully reducing consumption of meat and animal products is an unsolved problem: A meta-analysis. *Research Square*, rs-3.

Guala, F. (2017). Review of Cristina Bicchieri's Norms in the wild: how to diagnose, measure, and change social norms. Oxford: Oxford University Press, 2017, xviii+ 239 pp. *Erasmus Journal for Philosophy and Economics*, *10*(1), 101-111.

Haidt, J. (2012). *The righteous mind: Why good people are divided by politics and religion*. Vintage.

Haines, M., & Spear, S. F. (1996). Changing the perception of the norm: A strategy to decrease binge drinking among college students. *Journal of American College Health*, *45*(3), 134-140.

Hare, R. M. (1963) *Freedom and Reason*. Oxford: Clarendon Press.

Henrich, J., & Gil-White, F. J. (2001). The evolution of prestige: Freely conferred deference as a mechanism for enhancing the benefits of cultural transmission. *Evolution and human behavior*, *22*(3), 165-196.

Herstein, Ori J. (2017). Understanding standing: permission to deflect reasons. Philosophical Studies 174 (12):3109-3132.

Heyes, C. (2024). Rethinking norm psychology. *Perspectives on Psychological Science*, *19*(1), 12-38.

Hill, T. E. (1972). The kingdom of ends. In *Proceedings of the Third International Kant Congress: Held at the University of Rochester, March 30–April 4, 1970* (pp. 307-315). Springer Netherlands.

Hlobil, U. (2017). A handbook for social change: Cristina Bicchieri: Norms in the wild: how to diagnose, measure, and change social norms. Oxford: Oxford University Press, 2016, 264 pp, $29.95 PB.

Jansma, A., Van den Bos, K., & de Graaf, B. A. (2024). How climate protesters perceive injustice and justify breaking the law: Qualitative interviews with Extinction Rebellion. *Global Environmental Psychology*, *2*, 1-23.

Joy, M. (2020). *Why we love dogs, eat pigs, and wear cows: An introduction to carnism*. Red Wheel.

Kahan, D. M. (2015). The politically motivated reasoning paradigm. *Emerging Trends in Social & Behavioral Sciences, Forthcoming*.

Kar, K. & Milward, K. (2011) 'Digging in, spreading out and growing up: Introducing CLTS in Africa'. IDS Practice Papers, 1-64.

Kish Bar-On, K., & Lamm, E. (2023). The interplay of social identity and norm psychology in the evolution of human groups. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, *378*(1872), 20210412. https://doi.org/10.1098/rstb.2021.0412

Kelly, D., Westra, E., & S. Setman. (2025). "The Psychology of Normative Cognition." *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.) https://plato.stanford.edu/entries/psychology-normative-cognition/

Kelly, D., & T. Davis. (2018) "Social Norms and Human Normative Psychology." *Social Philosophy and Policy* 35, 54–76. https://doi:10.1017/s0265052518000122.

Lewis, D. (1969). *Convention: A philosophical study*. John Wiley & Sons.

Nussbaum, Martha C. (2004). *Hiding from Humanity: Disgust, Shame, and the Law*. United Kingdom: Princeton University Press.

Maibom, Heidi (2010). The Descent of Shame. *Philosophy and Phenomenological Research*, 80(3), 566-594.

Marques, J., Abrams, D., Paez, D., & Martinez-Taboada, C. (1998). The role of categorization and in-group norms in judgments of groups and their members. *Journal of personality and social psychology*, *75*(4), 976.

Marshall, Colin (2024). Political Persuasion is Prima Facie Disrespectful. *Journal of Moral Philosophy*, 1-34.

Mathur, M. B., Peacock, J., Reichling, D. B., Nadler, J., Bain, P. A., Gardner, C. D., & Robinson, T. N. (2021). Interventions to reduce meat consumption by appealing to animal welfare: Meta-analysis and evidence-based recommendations. Appetite, 164, 105277.

May, J., & Kumar, V. (2023). Harnessing moral psychology to reduce meat consumption. Journal of the American Philosophical Association, 9(2), 367-387.

Molinsky, A. (2007). Cross-cultural code-switching: The psychological challenges of adapting behavior in foreign cultural interactions. Academy of management review, 32(2), 622-640.

Ojala, K., & Nesdale, D. (2010). Group belongingness and intra-and intergroup attitudes in children. Griffith University, Queensland, Australia.

Phinney, J. S., Horenczyk, G., Liebkind, K., & Vedder, P. (2001). Ethnic identity, immigration, and well-being: An interactional perspective. Journal of social issues, 57(3), 493-510.

Pongrácz, P., & Dobos, P. (2023). What is a companion animal? An ethological approach based on Tinbergen's four questions. Critical review. Applied Animal Behaviour Science, 267, 106055.

Pong, V., & Tam, K. P. (2023). Relationship between global identity and pro-environmental behavior and environmental concern: a systematic review. Frontiers in Psychology, 14, 1033564.

Prentice, D. A., & Miller, D. T. (1996). Pluralistic ignorance and the perpetuation of social norms by unwitting actors. In *Advances in experimental social psychology* (Vol. 28, pp. 161-209). Academic Press.

Randers, L., & Thøgersen, J. (2023). Meat, myself, and I: The role of multiple identities in meat consumption. *Appetite*, *180*, 106319.

Rakoczy, H. & Schmidt, M. (2013). The early ontogeny of social norms. *Child Development Perspectives, 7*(1), 17-21. https://dx.doi.org/10.1111/cdep.12010.

Rakoczy, H., Warneken, F. & Tomasello, M. (2008). The sources of normativity: Young children's awareness of the normative structure of games. *Developmental Psychology, American Psychological Association (APA), 44*, 875-881.

Rawls, John (1971) A Theory of Justice. Cambridge, MA: Harvard University Press.

Rees, M. R., Smith, I. H., & Soderberg, A. T. (2024). Psychological reactance to leader moral hypocrisy. *Business Ethics Quarterly*, *34*(4), 634-661.

Riis, J., Simmons, J. P., & Goodwin, G. P. (2008). Preferences for enhancement pharmaceuticals: The reluctance to enhance fundamental traits. *Journal of Consumer Research*, *35*(3), 495-508.

Rini, Regina (2018). Abortion, Ultrasound, and Moral Persuasion. Philosophers' Imprint 18.

Rosenberg, B. D., & Siegel, J. T. (2018). A 50-year review of psychological reactance theory: Do not read this article. *Motivation Science*, *4*(4), 281.

Ruby, M. B., & Heine, S. J. (2011). Meat, morals, and masculinity. *Appetite*, *56*(2), 447-450.

Schwitzgebel, E., & Rust, J. (2014). The moral behavior of ethics professors: Relationships among self-reported behavior, expressed normative attitude, and directly observed behavior. *Philosophical Psychology*, *27*(3), 293-327.

Shulman, D., Shnitzer-Akuka, M., & Reifen-Tagar, M. (2021). The cost of attributing moral blame: Defensiveness and resistance to change when raising awareness to animal suffering in factory farming. *Plos one*, *16*(8), e0254375.

Singer, P. (1974). Sidgwick and reflective equilibrium. *the Monist*, 490-517.

Sittenthaler, S., Jonas, E., & Traut-Mattausch, E. (2016). Explaining self and vicarious reactance: A process model approach. *Personality and Social Psychology Bulletin*, *42*(4), 458-470.

Smerdon, D., Offerman, T., & Gneezy, U. (2020). 'Everybody's doing it': on the persistence of bad social norms. *Experimental Economics*, *23*, 392-420.

Sripada, C. S., Stich, S., Carruthers, P., Laurence, S., & Stich, S. (2006). A framework for the psychology of norms.

Stacey, T. (2022). The Truth Won't Save Us. In *Saving Liberalism from Itself* (pp. 159-171). Bristol University Press.

Stanley, Jason. *How propaganda works*. Princeton University Press, 2015.

Strohminger, N., & Nichols, S. (2014). The essential moral self. *Cognition*, *131*(1), 159-171.

Suls, J., & Green, P. (2003). Pluralistic ignorance and college student perceptions of gender-specific alcohol norms. *Health Psychology*, *22*(5), 479.

Tajfel, H. (1979). Individuals and groups in social psychology. *British Journal of social and clinical psychology*, *18*(2), 183-190.; 1981

Tajfel, H., Turner, J. C., Austin, W. G., & Worchel, S. (1979). An integrative theory of intergroup conflict. *Organizational identity: A reader*, *56*(65), 9780203505984-16.Westra & Andrews, 2024

Thomas-Walters, L., Scheuch, E., Parkes, I., Sabherwal, A., O'Dell, D., & Shreedhar, G. (2023). Barriers to Climate Activism: The Vital Role of Social Identification.

Todd, P. (2017) 'A unified account of the moral standing to blame'. Noûs, 53, 347-374.

Williams, Bernard (1989). Internal Reasons and the Obscurity of Blame. In William J. Prior, Reason and Moral Judgment, Logos, vol. 10. Santa Clara University.

Williams, A., Jones, C., Welch, V., & True, J. (2023) 'Outlook of pandemic preparedness in a post-COVID-19 world. NPC vaccines, 8(1), 178.

Witt, C. (2023). *Social goodness: The ontology of social norms*. Oxford University Press.

Wright, R. E., & Palmer, J. C. (2012). Reactance and behavioral change: Marketing healthy lifestyles. *Journal of Applied Business and Economics*, *13*(4), 136-143.

Zahavi, Dan (ed.): The Oxford Handbook of Contemporary Phenomenology. Oxford: Oxford University Press, 2012.