Social norms and the environment

Karine Nyborg

Social norms affect environmental quality. But what exactly is a social norm? Environmental economists studying the topic draw on diverse scholarly traditions, and may not have the same phenomenon in mind when using the concept. For example, ‘social norms’ may refer to common, but not necessarily socially approved, behaviors; to internalized ethical rules; or to one of several equilibria in a coordination game. I first discuss some of the definitions used in the environmental economics literature. Then, I outline a simple framework for analysis of voluntary contributions to public goods. Using this framework, I illustrate differences and similarities between altruism, moral norms, and social norms, and discuss implications for environmental policies. In particular, when a social norm represents one of several stable equilibria, policy can potentially invoke abrupt and dramatic behavioral changes.

Keywords: Social norms, moral motivation, voluntary contributions to public goods, tipping, environmental policy.

1 Department of Economics, University of Oslo, P.O.Box 1095 Blindern, NO-0317 Oslo, Norway. Email: karine.nyborg@econ.uio.no.
1. Introduction
Thirty years ago, few environmental and resource economists considered social and moral norms explicitly in their analyses. In accordance with the tradition in the field, researchers focused their attention on markets and formal institutions such as property rights, direct regulation, and market-based policy instruments. Ostrom (1990) demonstrated forcefully, however, that the traditional approach, while highly useful for a range of purposes, must have been missing something: in their path-breaking work, Ostrom and her collaborators found that even in the absence of individually assigned property rights and external enforcement, many local communities did, contrary to neoclassical predictions, manage their commons in sustainable ways.

Humans are social animals; social relationships are crucial for our well-being and ability to perform (Lieberman 2013). Given this, it would be more surprising if social interaction did not fundamentally affect human behaviors than if it did. Moreover, there are several reasons why informal mechanisms like social and moral norms may be of particular relevance to environmental and resource economists.

First, when it comes to management of environmental resources, market failures are prominent. Air and water quality, biodiversity and pristine wilderness are public goods; open-access fishery resources and pastures are common pool resources; pollution and traffic jams represent external effects. As indicated by Ostrom’s (1990) work, passive resignation is not the only possible response by people affected by market failures. Even in the absence of formal institutions, informal mechanisms may arise (Arrow 1970): for example, social and/or moral norms against littering in the wilderness, keeping a too large grazing herd, or polluting the drinking water. Where markets fail severely, informal institutions may be more likely to prevail.

In particular, informal mechanisms can play an important role when environmental externalities are accompanied by asymmetric information and limited third party verifiability. If a formal regulator cannot verify that a polluting activity has taken place, emission taxes or fines are of little help. If the activity can potentially be observed by the polluter’s peers, however, expected social disapproval may provide a strong disincentive. Even when the behavior can be observed by no-one but the polluter herself, internalized norms can affect behavior.

Second, if informal institutions are present, economic analyses not recognizing this fact may yield misleading conclusions – for example by suggesting policy advice undermining existing informal arrangements (Ostrom 1990, 2000).

Third, whether or not a social norm initially emerged as a response to an environmental problem, behaviors induced by such norms can have substantial environmental impacts. A social norm increasing the social status of those driving large, gas-guzzling cars is likely to affect air pollution. A social norm requiring food served at meals to be plentiful is likely to make food waste plentiful too. In cold climates, a social norm requiring sparse indoor clothing would increase energy demand for residential heating. Understanding the mechanisms of norms, and their interplays with markets, incentives and formal regulations, may be crucial for changing behaviors in such cases.
Although social norms function at the local level, they can spread between groups, eventually affecting behaviors on large scales (Nyborg et al. 2016). Prominent examples include changed norms for indoor smoking (Nyborg and Rege 2003a), the abandonment of foot-binding in China (Mackie 1996), and changed fertility norms (Dasgupta and Ehrlich 2013). The growing popularity of vegan and vegetarian diets in some western societies (Mintel 2014) may seem to be an emerging example.

In recent years, environmental and resource economists have increasingly turned their attention to informal institutions. A variety of informal mechanisms have been studied, such as social norms and social interaction, moral motivation, fairness norms, and reciprocal preferences (e.g., Czajkowski et al. 2017; Nyborg 2017; Dasgupta et al. 2016, Carlsson et al. 2016; Schlüter et al. 2016; Richter and Grasman 2013; Dasgupta and Ehrlich 2013; Videras et al. 2012; Allcott 2011; Cardenas 2011; Viscusi et al. 2011; Brown and Hagen 2010; Asheim 2010; Johansson-Stenman and Konow 2010; Lange et al. 2010; Kallbekken et al. 2010; Brekke et al. 2010; Bouma et al. 2008; Halvorsen 2008; Nyborg et al. 2006; Rege 2004; Brekke et al. 2003.) Topics that have been studied include, for example, eco-labeling, recycling, energy saving, common pool resource management and corporate social responsibility.

In everyday speech, concepts like ‘altruism’, ‘social norms’ and ‘moral norms’ are sometimes used interchangeably. Indeed, such phenomena often interact and/or coincide, making them hard to separate in practice. For analytical purposes, conceptual distinctions can still be useful. Nevertheless, different scholars apply different definitions, and do not even always seem to have roughly the same phenomenon in mind when referring to a ‘social norm’.

Nyborg et al. (2016, p.42) define a social norm as “a predominant behavioral pattern within a group, supported by a shared understanding of acceptable actions and sustained through social interactions within that group”. While game theorists may find this definition too vague, social psychologists may find it unnecessarily complex. Below, I begin by discussing alternative definitions of social norms used by environmental economists, with particular emphasis on concepts and ideas from these two scholarly traditions (Section 2).

I then turn to the distinctions between various forms of informal institutions, emphasizing the differences and similarities between altruism, moral norms and social norms (Section 3).

Cooperation on public good provision is a central theme for environmental economists. Several issues mentioned above, such as recycling, energy saving or purchase of eco-labeled goods, are examples of voluntary contributions to public goods (or reduced levels of public bads). In Sections 4 and 5 below, I outline a simple, stylized formal framework aimed at analyzing such questions. While starting off from the well-known pure and impure altruism models of Andreoni (1988, 1990), I use this framework to indicate, through minor modifications of the assumed preference structure, some of the central differences and similarities between altruism, moral norms and social norms, pointing out policy implications for environmental policies. Section 6 concludes.

Within the scope of a single journal article, providing a comprehensive review of the economics of social norms and the environment is not feasible. Consequently, I have chosen to focus on the ideas mentioned above, skipping rather lightly over other important and interesting aspects of the issue. Moreover, several excellent recent reviews on this and related topics are already available, and the interested reader is referred to those. Surveys on social norms and pro-environmental behaviors
have been provided by Farrow et al. (2017) and Carratini et al. (2017). Postlewaite (2011) and Young (2015) elaborate on economic analysis of social norms in general, while Videras (2013) discuss the environmental economics of social networks. The closely related field of environmental behavioral economics is surveyed by Croson and Treich (2014), Carlsson and Johansson-Stenman (2012) and Brekke and Johansson-Stenman (2008). Field experiments in resource and environmental economics is discussed by List and Price (2016) and Brent et al. (2016), whereas lab evidence on cooperation in public good game experiments is reviewed by Chaudhury (2011). Finally, OECD (2017) presents an overview of behavioral policy interventions in the environmental and energy sectors.

2. Definitions and concepts

Social norms have been studied by scholars from a number of disciplines, including sociology (e.g., Bourdieu 1979, Coleman 1990), psychology (Miller and Prentice 2016, Cialdini and Goldstein 2004, Cialdini et al. 1991), anthropology (Mauss 2011, Geertz 1973), philosophy (Biccieri 2006, Elster 1989), biology (Gavrilets and Richerson 2017, Ehrlich and Levin 2005), and economics (Young 1998a, 1998b, 2015, Brock and Durlauf 2001, Lindbeck et al. 1999, Akerlof 1980). However, combining insights from different disciplines can be very demanding, one obstacle being the substantial variations in professional language and methodology. We use similar words for different ideas; we use different words for similar ideas; and even when having quite similar phenomena in mind, we tend to reason very differently about them.

Thus, interdisciplinary communication requires more than simple word-by-word translation. Scholars from different disciplines are asking different fundamental questions, requiring them to distinguish concepts along different lines. Different concepts may reflect different reasoning. Just like a language evolved in a cold climate can have numerous words for “snow”, conveying finely differentiated meanings hardly understandable to someone from the tropics, scholars working within different traditions may need different ideas and distinctions to be reflected in their professional language.

Psychologists’ analyses, for example, focus on individual perceptions, actions and responses to the social environment, which requires due attention to contextual and individual heterogeneity. Economists, on the other hand, often neglect heterogeneity altogether. This is partly due to our interest in equilibria: while the individual is influenced by society, society consists, after all, of individuals, implying that equilibrium analysis must handle the feedbacks both ways – which often becomes too complex without gross simplifications. Evolutionary ecologists share economists’ interest in equilibria, especially the idea of multiple equilibria and concepts like tipping points and regime shifts. However, they are less preoccupied with markets, agents’ expectations, and policy design. Anthropologists, whose primary interest lies in cultural differences, are also less policy-oriented than economists.

2.1 Social norms in social psychology

Environmental economists’ thinking about social norms has been substantially influenced by social psychology (see Farrow et al., 2017). Aronson et al. (2005) define a social norm as “[t]he implicit or explicit rules a group has for the acceptable behaviors, values, and beliefs of its members” (p. 259). Cialdini et al. (1991) point out, however, that the word ‘norm’ “can refer either to what is commonly done – that is, what is normal – or to what is commonly approved – that is, what is socially sanctioned” (p. 202). To reflect this, Cialdini et al. (1991) distinguish between descriptive social
norms, which guide behavior via perceptions of how most others behave, and injunctive social norms, guiding behavior via perceptions of how most others would approve or disapprove of the behavior. These distinctions have become central to psychologists’ discourse on social norms. Further, Cialdini et al. (1991) define personal norms as norms that guide behavior via perceptions of how a person would approve/disapprove of her own conduct.

If a person believes that most others reuse their hotel towels, this is a descriptive norm, which will typically make her more inclined to reuse her own towel (Cialdini and Goldstein 2004, Goldstein et al. 2008). Her reasons to follow others’ example may lie in mere conformism, i.e., a wish to fit in by being or acting like others; she may interpret others’ behavior as a signal of the associated individual payoffs (if it’s clean enough for them, it’s probably clean enough for her too), or as an indication of the behavior’s social or moral appropriateness (Nyborg et al. 2006, Brekke et al. 2010).

A descriptive norm might thus signal the presence of an injunctive norm. Nevertheless, psychologists emphasize that the two do not necessarily coincide (Cialdini et al. 1991, Goldstein et al. 2004). Even if a person believes that others think she should be reusing her towel, she does not necessarily believe that they are reusing theirs. Even if she thinks others are reusing their towels, she does not necessarily expect them to think she should be doing so too.

In a given situation, several (possibly conflicting) norms may be present simultaneously. Psychologists emphasize that in such cases, a given norm is more likely to influence behavior if it is salient in the individual’s mind (Reno et al., 1993). Thus, when psychologists study the effects of social norms on environmental behaviors, they may for example vary the information provided to individuals about others’ behavior (“the majority of guests in this room reuse their towels”, Goldstein et al., 2008), or by introducing contextual clues reminding the individual of some injunctive norm (letting a confederate insert a single piece of litter into a clean environment, Reno et al. 1993).

Perhaps one way to summarize is that in this literature, a social norm mainly describes a behavior that is normal – either in the sense of being common, or in the sense of being commonly approved of.

2.2 Social norms in game theory
This differs, however, from a common understanding of the term ‘social norms’ among game theorists (see Young 2015, 1998a, Aoki 2011, Gintis 2010, Bicchieri 2006, Binmore 2005, Sugden 1989). For example, even if someone believes that the majority of guests staying in her hotel room before her reusing their towels, reusing one’s towel would hardly be defined as a social norm in the game theoretic sense.

According to Young (1998a, p. 821), “[a] norm is an established and self-reinforcing pattern of behaviour: everyone wants to play their part given the expectation that everyone else will continue to play theirs. It is, in short, an equilibrium of a game.”

Young (2015, p. 361) points out that norms 1) are self-reinforcing, sustained by positive feedback loops; 2) typically evolve without top-down direction through a process of trial and error, experimentation, and adaptation; and 3) govern interactions that have multiple equilibria, making them contingent on context, social group, and historical circumstances. This view of social norms, thus, is closely linked to equilibrium analysis, based on ideas such as multiple Nash equilibria, self-reinforcing mechanisms and coordination.
To see this, it may be instructive to consider the following one-shot two-player game. Two car drivers, unable to communicate and driving in opposite directions on a narrow and heavily curved road, are considering whether to keep right or left. If they keep to different sides of the road, they will collide; if they choose the same side, each will reach their destination safely. Their payoffs can be specified as in Table 1 below:

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left</td>
<td>5, 5</td>
<td>-10, -10</td>
</tr>
<tr>
<td>Right</td>
<td>-10, -10</td>
<td>5, 5</td>
</tr>
</tbody>
</table>

**Table 1: Drivers’ coordination game.** Numbers to the left are the row player’s payoffs, numbers to the right are the column player’s payoffs.

This game has two pure strategy Nash equilibria: that both keep right, and that both keep left. It does not matter which side they choose, as long as they choose the same one, hence there is no reason to consider one equilibrium as intrinsically better than the other. If they play the game only once, however, and have no other information, there is nothing in the game itself helping them coordinate successfully; while a mixed strategy is as good as anything in such a situation, it will not grant them a particularly high expected payoff. However, if each knows that it is customary to drive on the left in the country they are in, and has reason to expect the other to know this too, the natural solution is for both to drive on the left (Sugden 1989).

Commonly known rules about which side of the road to drive on is an example of a convention, defined by Young (1998a, p. 823) as “equilibrium behavior in a game played repeatedly by many different individuals in society, where the behaviors are widely known to be customary. Note the importance of knowledge: the behaviors must not only be customary, they must be known to be customary, or else the behaviors are not in fact self-enforcing.” A convention, however, may be self-enforcing even without social sanctions. If driving on the left is the local convention, it is in your interest to follow it, even if no-one will be shouting at people who do not.

Consider now the following example. Two families let their cattle feed on the same grasslands. Each family is considering whether to increase the size of their own herd. If each family does so, overgrazing will result, and each family is worse off than before. The worst possible case, however, is that the other family increases the size of their herd, while one keeps a small herd oneself. Assume that payoffs in material terms can be represented as follows, constituting a Prisoners’ Dilemma game:

<table>
<thead>
<tr>
<th></th>
<th>Small herd</th>
<th>Large herd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small herd</td>
<td>4, 4</td>
<td>-2, 5</td>
</tr>
<tr>
<td>Large herd</td>
<td>5, -2</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

**Table 2: The overgrazing game.** Numbers to the left are the row player’s payoffs, numbers to the right are the column player’s payoffs.

If played a finite number of times, this game has only one Nash equilibrium, namely that both families increase their herds, leading to overgrazing and unproductive grasslands. However, imagine
that if a family is the only one to increase the size of its herd, it will face social sanctions, for example by being socially excluded from the village. Assume that the value of these sanctions for the family equals 3. Then, the game turns into a coordination game in total benefits:

<table>
<thead>
<tr>
<th></th>
<th>Small herd</th>
<th>Large herd</th>
</tr>
</thead>
<tbody>
<tr>
<td>Small herd</td>
<td>4, 4</td>
<td>-2, 5-3</td>
</tr>
<tr>
<td>Large herd</td>
<td>5-3, -2</td>
<td>1, 1</td>
</tr>
</tbody>
</table>

Table 3: The overgrazing game with social sanctions. Numbers to the left are the row player’s payoffs, numbers to the right are the column player’s payoffs.

The overgrazing situation is still a Nash equilibrium: if the other family increases its herd, the best response is to follow suit. Nevertheless, there is another pure strategy Nash equilibrium as well: namely, that each family keeps the small size of their herd. This situation is self-enforcing – due to the social sanctions threatening those who deviate.

Again, if this game is played only once, with no external information available to the parties, there is nothing in the game itself telling the families how to coordinate (Gintis 2010, Sugden 1989). However, let us return to Young’s definition of a convention, modifying it slightly to yield a possible definition of a social norm: A social norm is equilibrium behavior in a game played repeatedly by many different individuals in society, where equilibrium behavior is sustained through social sanctions, and where the behaviors and the associated social sanctions are widely known to be customary.

Thus, if the traditional practice in the village is to keep small herds, and to ostracise those who break this rule, we might say that a social norm of keeping small herds is in place (is being enforced). If the tradition is, on the other hand, to keep the herd size one sees fit and not to sanction those keeping large herds, there is no social norm in place restricting herd sizes, and the village population is playing the non-cooperative Nash equilibrium.

In the herd example, I ignored the question of how the sanctioning of deviators came about. Social sanctioning – whether done by the other family or a third party – may be a choice too, and if so, that choice should also be considered part of the game. Indeed, a social norm will often be part of a larger pattern of rules and expectations in a society, where the various rules support each other to form a Nash equilibrium in the larger game (Basu 2005, Ch. 5.2). If I keep a too large herd, you will sanction me; if you do not sanction me, your neighbor will sanction you; if he does not, his reputation is reduced, weakening the chance that his daughter is allowed to marry my son; and so on. The point is that such networks of expected behaviors support and reinforce each other, making the entire pattern of behaviors a Nash equilibrium. (Gintis, 2010, argues that social norms ought to be described as correlated equilibria rather than Nash equilibria, reflecting the ‘choreography’ role of joint expectations and rules. To keep things simple, I will be ignoring this distinction here.)

### 2.3 Social norms: descriptive, injunctive, and self-reinforcing

Recall, now, the definition of a social norm provided in the introduction: “a predominant behavioral pattern within a group, supported by a shared understanding of acceptable actions and sustained through social interactions within that group” (Nyborg et al. 2016, p.42).
While developed with the intention of being understandable to an interdisciplinary audience, this definition is closely linked to the game theoretic tradition.

For a behavioral pattern to constitute a social norm according to this definition, the norm must be descriptive as well as injunctive (a predominant behavioral pattern within a group, supported by a shared understanding of acceptable actions). A behavior that is common, but not necessarily generally viewed as more acceptable than the alternatives – opting for a clean hotel towel in a context where this appears to be usual default choice – is not a social norm according to this definition. The same holds for uncommon, but widely socially approved behaviors, like spending one’s vacation cleaning up litter in a natural park.

Even if a behavior is common as well as socially approved, it is not a social norm, according to this definition, unless it is enforced through social interaction within the group. That is, if I abstain from littering in the street for fear of social disapproval, I may be obeying a social norm; if I abstain from littering when alone and unobserved in the wilderness, I am probably rather obeying an internalized moral norm.

Finally, while less explicit about Young’s (1998a) requirements of repeated play by many different individuals and wide knowledge of the custom, the last part of the definition by Nyborg et al. (2016) does hint at the idea of norms as equilibrium behaviors: the behavior is not simply present, but sustained; and it is sustained at least in part through social interactions such as social approval and disapproval. The latter distinguish a social norm from conventions as well as phenomena such as network externalities and internalized moral norms.

3 Altruism, moral norms, social norms

So what, then, a moral norm? Is it the same thing as altruism? And does it matter for predictions, and for policy implications, whether a norm is social or moral? Let me offer some general remarks, before turning to a slightly more formal framework below.

Like ‘social norms’, the term ‘altruism’ is used differently within different fields and disciplines. In economics, the term is usually taken to mean that the altruist is at least partly internalizing the utility of someone else, or certain aspects of someone else’s utility, in her own utility function. However, since economists tend to have different ideas in mind when they speak of ‘utility’, interpreting the above statement is not necessarily straightforward. Standard neoclassical consumer theory usually considers ‘utility’ simply a numerical representation of the individual’s revealed choices.

In welfare economics, on the other hand, utility is often interpreted simultaneously as a representation of revealed choices and as a corresponding, welfare relevant measure of well-being, life satisfaction, or the like. Given the latter utility concept, an altruist can be thought of as someone whose personal well-being is increasing in (certain aspects of) the well-being of someone else. This dependency is reflected in the altruist’s behavior: when acting to maximize her own well-being, the altruist will, at least to some extent, take into account the effects of her actions on the other’s well-
being. This seems to capture phenomena such as true love: when a child is in severe pain, for example, the parent is herself desperate; when my beloved one is happy, I share his happiness.

Moral motivation, the way I think of it, is not primarily a question about being personally affected by others’ well-being, but about doing what one finds to be ethically right. Whereas the altruistic parent feels better when the child’s pain is relieved, regardless of the reasons for the latter, the primary concern of a morally motivated person is to fulfill her moral obligations, not the other’s well-being as such. Thus, altruism appears to correspond to Amartya Sen’s (1977) concept of ‘sympathy’, while moral motivation corresponds to his concept of ‘commitment’: “If the knowledge of torture makes you sick, it is a case of sympathy; if it does not make you feel personally worse off, but you think it is wrong and you are ready to do something to stop it, it is a case of commitment” (Sen 1977, p. 95).

A moral norm, thus, is a rule of ethically appropriate behavior, enforced by the individual herself though inner feelings such as guilt, conscience, and cognitive dissonance. Hence, this seems to correspond reasonably well to psychologists’ concept of personal norms (Cialdini et al., 1991).

As emphasized above, a social norm is at least partly enforced through extrinsic social forces such as others’ social approval, disapproval, inclusion or exclusion. A social norm may well coincide with a moral norm, however, in the sense that a behavior may be encouraged both by internalized moral views and external social sanctions. It is also easy to envisage situations where social and moral norms are at odds: if the social norm dictates genital mutilation of girls, a parent may mutilate her daughter to prevent social exclusion, while still silently considering the practice morally wrong.

Dasgupta et al. (2016) point out that socially interdependent preferences can be ‘socially-directed’ and ‘socially-embedded’: socially-directed preferences reflect that one cares about others; socially-embedded preferences that one is influenced by others. Altruism and moral motivation are thus examples of socially-directed preferences, while social norms are enforced through socially-influenced preferences.

However, these distinctions are not clear-cut. To the extent that altruism and/or moral norms coincide with social norms, behavior is shaped by socially-embedded as well as socially-directed preferences (without anyone necessarily being able to distinguish between the different parts in practice).

Furthermore, social and moral norms are intertwined in several ways. First, moral norms are socially learned. Not only do prevailing social norms affect which rules people internalize as moral norms themselves: social norms may even determine which ethical values parents actively try to install in their children (Tabellini 2008). In a trusting and trustworthy society, for example, an honest child can do well; in a society of cheaters, an honest child will be continually exploited and consequently suffer. Also, when in doubt about one’s personal moral responsibility, people may look to others’ behavior for guidance (Nyborg et al. 2006, Brekke et al. 2010). And of course, social and moral norms may possibly have developed as responses to the same problem.

Even in the very short term, where internalized moral norms may be considered fixed, morally motivated behavior is influenced by social interaction. In particular, the prevalence of conditional cooperation is thoroughly documented; for example, even in anonymous environments, a substantial share of individuals are willing to contribute more to a public good the more others contribute.
(Chaudhury 2011, Ledyard 1995). In a public good game experiment, Hauge (2015) asked her subjects explicitly what they considered to be the morally right behavior in the experiment; when conditional answers were allowed, the majority of subjects reported that the morally right behavior was to contribute an amount equal to the average contribution in one’s group.

Reciprocal preferences (Rabin 1993, Dufwenberg and Kirchsteiger 2004, Sobel 2005, Falk and Fischbacher 2006, Cox et al. 2007) provides one plausible explanation for the prevalence of conditional cooperation. Reciprocity can be defined as the preference to repay kind intentions by kind actions and mean intentions by mean actions (Nyborg 2017). It can help explain why social sanctions and approval are often provided even in anonymous contexts, and at a net material cost to the sanctioner (Fehr and Gächter 2002, Rabin 1993). Reciprocity provides a possible link between moral and social norms: while the moral norm determines which behaviors are perceived as “kind” or “mean”, the emotional urge to reciprocate triggers the social reaction needed to sustain the social norm. (Of course, “kindness” or “meanness” might also be judged based on other criteria than obedience of moral norms; others’ failure to honor one’s expected (if unfair) privileges being only one example.)

4 A simple formal framework: altruism and morality

Norms can influence a large range of behaviors – fertility, overconsumption, gender roles, diet restrictions, improper language, and so on. Also, a multitude of approaches to formal modelling of social and moral norms exists in the economics literature (e.g., Akerlof 1980, Sugden 1984, Benabou and Tirole 2006, Michaeli and Spiro 2017). Nevertheless, in the following, I will narrow my focus to consider a very simple and stylized framework for analysis of voluntary contributions to public goods (see also Nyborg and Rege 2003b).

Consider a society with \( N > 1 \) identical individuals, who have identical preferences and an equal, exogenously given income \( Y > 0 \). \( G \) is the supply of a pure public good, for example a stable global climate. If they so wish, individuals can contribute public good – for example by purchasing climate offsets when travelling by air.

Each individual \( i \) shares her income between her own private consumption \( x_i \) and her contributions towards the public good, \( g_i \):

\[
(1) \quad Y = x_i + g_i
\]

The total supply of the public good is given by an initial, exogenously determined supply \( G^0 \), plus contributions from each individual:

\[
(2) \quad G = G^0 + \sum_{j=1}^{N} g_j
\]

Assume that each \( i \) considers the total contributions from others, \( G_{-i} = G^0 + \sum_{j \neq i} g_j \), exogenously given.

4.1 Self-interest, pure and impure altruism

Now, let us first regard the case where individual \( i \)'s utility \( U_i \) is given by

\[
(3) \quad U_i = u(x_i) + v(G),
\]
where $u$ and $v$ are concave and strictly increasing functions, and linear separability is assumed for simplicity. This can be viewed as representing narrowly self-interested preferences (Bergstrom et al. 1986). If there is an interior utility maximum, $i$ contributes until the following first-order condition is fulfilled (primes denote derivatives): $u' = v'$. However, if the public good supply by others is sufficiently high, the individual prefers a corner solution, contributing nothing at all: if $v'(G_{-i}) < u'(Y)$, the marginal increase in her own public good benefits secured by her own contribution justifies no contribution whatsoever. Consequently, if $G^0$ is sufficiently high, no-one contributes anything voluntarily. To make individuals contribute in this case, a regulator would have to force them, through some form of direct regulation, or to subsidize contributions by a sufficient high amount.

This is, of course, the classical free-rider problem. In this economy, an interior Pareto optimal contribution level by each $i$ must satisfy the first order condition for welfare maximum, $u' = Nv'$, implying a strictly higher contribution level than the utility maximizing condition $u' = v'$. Likewise, for no individual contributions at all to be Pareto optimal, we must have that $Nv'(G^0) < u'(Y)$.

The preferences specified in eq. (3) need not represent a purely self-interested person, however. Alternatively, they can describe what Andreoni (1988) called pure altruism – a concept often referred to by environmental economists. The difference lies in the interpretation of $v(G)$: while the narrowly self-interested person cares about his own access to the public good, the pure altruist cares about others' (possibly also his own) access to the public good.

Nevertheless, since the two models are formally equivalent and only the interpretation differs, the free-riding results above hold for the purely altruistic person as well as the narrowly self-interested. Of course, if a narrowly self-interested and a pure altruist have identical preferences for their own public good benefits, the altruists’ willingness to contribute is higher than the egoist’s. For example, the altruist contributes a positive amount at higher levels of others’ contributions than the egoist does. Nevertheless, pure altruism is behaviorally indistinguishable from narrow self-interest with a strong taste for the public good. If contributions are costly, the pure altruist always prefers others to secure the provision.

Moreover, pure altruism and behaviorally equivalent models yield quite implausible equilibrium predictions, such as full crowding-out of public sector supply, no voluntary contributions by anyone but the very richest individuals, and neutrality of taxes (Warr 1982, Roberts 1984, Bernheim 1986, Bergstrom et al. 1986, Andreoni 1988). Consequently, purely altruistic preferences is hardly a plausible explanation of substantial voluntary contributions to public goods by substantial parts of the population (see Nyborg and Rege 2003b).

Motivated by these shortcomings, Andreoni (1990) proposed the idea of ‘impure altruism’, which can be represented by a modification of eq. (3) as follows:

$$U_i = u(x_i) + v(G) + h(g_i)$$

where $h$ is an increasing and concave function, and $h(g_i) \geq 0$ is a private good, the ‘warm glow’ of giving. Unlike the benefits arising from the environmental quality level, $v(G)$, the warm glow cannot be gained through others’ contributions; hence, there is no free-rider problem associated with this
part of the utility function. Consequently, the impure altruist may well choose to contribute even if $G_{-i}$ is high.

Students often find the words ‘pure and ‘impure’ confusing. While the impure altruist contributes at least in part to grant himself a nice feeling, so does, in a certain sense, the pure altruist (at least if one interprets utility as well-being); and as pointed out above, pure altruism is in fact behaviorally equivalent to narrow self-interest. Nevertheless, these phrases have become quite established in the literature.

4.2 Moral norms
In the simple representation of the impure altruist model above, ‘warm glow’ is obtained through the very act of giving: the more one contributes, the better one feels. Viewed as moral reasoning, this would of course be rather unsophisticated. Wouldn’t a ‘warm glow’ plausibly depend on other factors as well, such as the (believed) social importance of the cause, or one’s perceived moral obligation to do something about it? It turns out that in some cases, taking individuals’ ethical reasoning explicitly into account can matter for behavioral predictions as well as policy implications.

Following Brekke et al. (2003), let us make a slight reinterpretation and respecification of eq. (3). While still assuming that $i$ obtains a private benefit from contributing, let us now interpret this as the benefits from keeping a self-image as morally responsible:

\begin{equation}
U_i = u(x_i) + v(G) + I_i
\end{equation}

where $I_i$ is $i$’s self-image, determined by

\begin{equation}
I_i = -a(g_i - g^*)^2,
\end{equation}

where $a > 0$, and $g^*$ is the morally ideal contribution for $i$, as perceived by the individual herself (since everyone is identical, individual differences in the view of $g^*$ is suppressed in the notation).

With this specification, $i$ can only improve her self-image by coming closer to what she truly considers to be morally right. Note also that whenever there is a discrepancy between what $i$ is doing and what she thinks she should be doing, she experiences a self-image loss; the best possible self-image equals zero. Thus, this corresponds closely to what psychologists call cognitive dissonance (see Aronson et al. 2005, Ch.6).

Brekke et al. (2003) assume that $g^*$ is determined endogenously by the individuals herself by appealing to a general ethical rule or moral norm. Several alternative moral norms may be relevant here, reflecting different views about matters such as individual moral obligations, fairness and reciprocity, judgements of social welfare and so on. One example would be the conditionally cooperative view a morally ideal contribution mentioned above (Hauge 2015): “I should contribute an amount equal to the average of others’ contributions”. Brekke et al. (2003), however, assume that the individual asks herself: “What would happen if everyone acted just like me?” More precisely, they assume that the morally ideal contribution $g^*$ is determined endogenously in the model as the contribution that would have maximized social welfare if made by everyone. This should be understood as a purely hypothetical question; the individual may not expect anyone to follow her example, but is using a rule inspired by the Golden Rule or Kant’s Categorical Imperative to
determine her own moral responsibility in the situation at hand. That is, if \( i \) is, for example, a utilitarian, ascribing to a social welfare function of the type

\[
W = U_1 + \cdots + U_N,
\]

then, \( g^* \) is defined as

\[
g^* = \arg \max_{g_i} W, \text{ s.t. } g_i = g_j \text{ for all } j = 1, \ldots, N.
\]

One may discuss, of course, whether the self-image part of utility ought to count in welfare evaluations. There are arguments in each direction. The benefits of a good self-image are hardly less real than any other benefits; nevertheless, it is far from obvious that a morally motivated person would consider the satisfaction of doing good a relevant input in the definition of good itself. However, this question does not matter for the welfare maximization used to define \( g^* \): in that exercise, \( g_i = g^* \) by definition, hence \( I_i = 0 \) (for every morally motivated individual \( i \) included in the evaluation).

This type of model can be viewed as reconciling utility maximization with true ethical reasoning. While the morally motivated is admittedly contributing in order to obtain a better self-image, the only way to achieve that benefit is to get closer to the behavior one find to be morally ideal.

Nevertheless, a morally motivated person is no angel. She trades off self-image benefits against other goods, always contributing strictly less than \( g^* \). Thus, although she may refer to Kant when determining the moral ideal action \( g^* \), her actual behavior is not categorical in the Kantian sense.

Why go through all this, not simply keeping with the impure altruism model? Since \( g^* \) is endogenous, any changes affecting \( g^* \) may affect behavior too. This can change the impacts of environmental policy in ways not predicted by the impure altruism model.

First, increased efficiency of contributions, for example due to better curbside collection systems to facilitate recycling of household waste, can make morally motivated individuals increase their contributions (Brekke et al. 2003). Second, economic incentives such as green taxes can have quite different effects depending on whether they are interpreted as prices, covering the associated social costs, or as punishments, i.e., symbolic reminders that the behavior is wrong (Brekke et al. 2003).

Third, since a heavy moral responsibility (a high \( g^* \)) is a burden, the morally motivated may strategically avoid information (Nyborg 2011). Hence, in certain cases, information campaigns can be an effective means to increase contributions – simply by providing unwanted information.

Fourth, Asheim (2010) shows that in the context of asymmetric information about the severity of an environmental problem, moral motivation provides a rationale for skepticism towards information provided by the government (such as climate change skepticism). Asheim’s argument goes as follows. The government knows that the morally motivated contribute towards a better environment, but not enough (\( g_i < g^* \)). By strategically selecting which information to convey, the government can exaggerate the magnitude of the problem, thus aiming to make the morally motivated increase their contributions all the way to the first-best \( g^* \) (assuming that the government shares the social welfare views of the morally motivated). However, rational individuals foresee this, and are thus skeptical to the information. Unless the government can make a credible commitment always to convey all
available information, the problem cannot be solved simply by promising to be unbiased in the information provision, since citizens cannot verify whether the promise is being kept.

5 Social norms

The analysis of moral norms above was based on an interpretation of the ‘warm glow’ as the individual’s approval of herself. Another possible interpretation is that the ‘warm glow’ represents social approval from the individual’s peers.

5.1 A social norm interpretation of warm glow

Let preferences now be given by

\[ U_i = u(x_i) + v(G) + s_i, \]

where \( s_i \) is the social approval \( i \) receives from others. For simplicity, assume now that each \( i \) considers her own potential impact on the total public good supply to be negligible, hence regarding \( G \) as exogenous. Moreover, let the choice of \( g_i \) be a discrete one: either the individual contributes to the public good, denoted \( g_i = 1 \), or she does not contribute, where \( g_i = 0 \). Assume \( Y > 1 \).

As an example, let \( G \) be water quality in a lake, while the \( N \) individuals are farmers whose farms are situated around the lake. Consider a situation where each farmer can choose to emit its wastewater uncleaned into the lake. However, at a private cost 1 to be covered by the farmer, each farmer can alternatively choose to utilize a central sewerage system, directing the farm’s wastewater to a cleaning facility outside of the village. If the farm does not use the central system, \( g_i = 0 \).

In contrast to the case of internalized moral motivation, contributions must be observable to one’s peers for social approval to matter for the contribution decision. Hence, assume that wastewater emissions into the lake are easily visible by passers-by, and, moreover, one can easily identify which farm the emission originates from.

Assume that farmers who emit their wastewater into the lake neither receive nor give social approval; farmers utilizing the central system, however, exchange social approval when they meet. Let \( a \) be the share of farmers who use the central system. Each \( i \) considers her own impact on the share \( a \) as negligible. Let the social approval received by farmer \( i \) be given by

\[ s_i = aKg_i, \]

where \( K > 0 \) is a fixed constant.

Farmer \( i \) will then prefer using the central system if and only if the social approval benefits of doing so outweighs the lost consumption benefits due to her private cost. This holds if

\[ a \geq \frac{[u(Y) - u(Y - 1)]}{K}, \]

i.e., if the share of farmers using the central system is large enough.

Under these conditions, will there be multiple Nash equilibria? Consider first the situation in which no-one uses the central system, i.e., \( g_i = 0 \) for all \( i \), so \( a = 0 \). In this case, for eq. (11) to hold, we
must have $0 \geq \frac{[u(Y) - u(Y-1)]}{K}$. This cannot hold, since $u$ is strictly increasing and $K$ and $Y$ are all strictly positive. Intuitively, if no-one else uses the central system, there is no-one to provide social approval to somebody who does, thus no-one wants to use the central system. Hence $\alpha = 0$ is a Nash equilibrium.

Next, consider the case where $\alpha = 1$, i.e., everyone uses the central system, contributing $g_i = 1$. In this case, (11) implies that a farmer prefers contributing (using the central system) whenever

(12) \hspace{1cm} K \geq u(Y) - u(Y - 1).

Consequently, if the parameter $K$ is large enough, the social approval gained by contributing outweighs the lost consumption benefits – given that everybody else is contributing too. Hence, $\alpha = 1$ is also a Nash equilibrium if (12) holds.

Again, if the game is played only once, there is nothing in the game itself to tell the farmers how to coordinate. However, consider the case where there is a local tradition of emitting one’s wastewater uncleaned into the lake. Since the model is static, one cannot, within the formal framework, interpret this statement in an explicitly dynamic sense. Nevertheless, it can be understood as indicating a focal point (Schelling 1960): among the possible equilibria, one stands out as a more natural point to coordinate on.

If the tradition is to emit one’s wastewater uncleaned into the lake, and this is interpreted as a focal point, farmers have little reason to do anything else than following the tradition. Thus, there would be no social norm enforcing central cleaning of one’s wastewater in this economy.

If, on the contrary, the local tradition is to use the central cleaning system, it is the Nash equilibrium in which everyone uses the central cleaning system which would constitute the natural focal point. That is, there is a social norm for cleaning one’s wastewater: “everyone wants to play their part given the expectation that everyone else will continue to play theirs” (Young 1998a, p. 821). (For a more formal discussion, see Gintis 2010.)

Note that if (12) holds with strict inequality, there must be a third Nash equilibrium too, namely the situation where $\alpha = \hat{\alpha} = \frac{[u(Y) - u(Y-1)]}{K}$. When a share of exactly $\hat{\alpha}$ are using the system, (11) holds with equality, meaning that everyone is exactly indifferent between using the system and not using it. (This can occur either because everyone applies a mixed strategy, using the central system with probability $\hat{\alpha}$ and not using it with probability $(1 - \hat{\alpha})$, or because a share $\hat{\alpha}$ of the farmers are participating with certainty while a share $(1 - \hat{\alpha})$ are not.) However, this equilibrium is unstable in the sense that any small deviation making the share using the system marginally larger (smaller) than $\hat{\alpha}$ would make all farmers strictly prefer (not) to use the system. In this sense, $\hat{\alpha}$ is a tipping point in the model.

Social norm models need not necessarily produce multiple equilibria. The crucial question is whether mechanisms in the model make the individual benefit of some behavior increase sufficiently in the share of others choosing the same behavior. In models of social status-seeking, for example, the reverse is often the case: the social esteem of driving a huge car, for example, may be lower the larger share of others are doing the same thing (see Dasgupta et al. 2016). Social norm models with
multiple equilibria are particularly interesting, however, because apparently minor changes in underlying parameters or policies can sometimes cause abrupt and dramatic behavioral changes.

5.2 Policy
Assume, now, that the village is initially in the no-cleaning Nash equilibrium, \( a = 0 \), in which \( G = G^0 \) and \( U_i = u(Y) + v(G^0) \). If \( a = 1 \), the lake would be cleaner: \( G = G^0 + N \), and each farmer’s utility would be \( U_i = u(Y - 1) + v(G^0 + N) + K \). That is, provided that the cost of central cleaning is sufficiently low compared to the benefits of a cleaner lake, the \( a = 1 \) equilibrium is Pareto superior to the \( a = 0 \) equilibrium (and may be so even if one chooses not to count the social approval benefits as relevant for welfare comparison).

Assume that the village council does not have the authority to require everyone to use the central cleaning system. However, the council chooses to introduce a subsidy \( q > 1 \) for each farmer using the central system.

In the model above, \( g_i \) denoted farmer \( i \)'s usage of the central cleaning system as well as her net cost of using the system. To study the impact of a subsidy, we need to be able to distinguish between the two. Let \( g_i \) now denote farmer \( i \)'s usage of the central cleaning system, such that \( g_i = 1 \) means that the farmer is using the system, while \( g_i = 0 \) means that she is emitting her wastewater directly into the lake. This requires that we also modify the budget equation (1): Now, assume that \( Y = x_i + 1 - q \) when \( g_i = 1 \), while \( Y = x_i \) when \( g_i = 0 \).

After implementation of the subsidy, what are the Nash equilibria in this economy? As before, a farmer prefers using the central system if the utility of doing so exceeds the utility of not doing so, taking \( a \) and \( G \) as given. This now implies \( a \geq \frac{[u(Y) - u(Y - 1 + q)]}{K} \).

This always holds, since \( u(Y) - u(Y - 1 + q) < 0 \), while \( a \geq 0 \). Thus, the only Nash equilibrium when the subsidy is implemented is that everyone uses the central cleaning system.

Next, consider the reverse change. Regard a village identical to the one discussed above, in which everyone is initially using the central cleaning system, and a subsidy \( q > 1 \) is implemented. The village council now wants to remove this subsidy. Will that make the farmers start polluting the lake? Not necessarily.

In the initial situation, with the subsidy, using the central cleaning system is the only equilibrium strategy. However, assume that \( K \geq u(Y_i) - u(Y_i - 1) \). Then, if the subsidy is removed, \( a = 0 \) becomes a Nash equilibrium too; but this does not imply that \( a = 1 \) ceases to be a Nash equilibrium. In fact, since the established practice in this village is to use the central cleaning system, it may well be the case that \( a = 1 \) still functions as a focal point – in which case no-one has reason to change their behavior, even though the subsidy is abolished.

The above is a static version of the argument provided by Rege (2004) in her explicitly dynamic model of environmental social norms. Assume that in the absence of policy, there are at least two equilibria, A and B, and that the economy is initially in A. A policy is then introduced, pushing the economy to equilibrium B. The economy may then stay in B even after policy reversal – since, after all, B is an equilibrium too. That is, in situations with multiple equilibria, a regulator can help
coordinating individuals in a preferred equilibrium. To pass the tipping point, a policy may have to be forceful – but it may not need to be permanent.

For a more formal analysis of policies’ effects on equilibrium selection, explicit dynamic modelling is of course required. A diversity of approaches have been proposed and used for this purpose (see, e.g., Young 1998b, Young 2015, Brock and Durlauf 2001, Gavrilets and Richerson 2017, Schlüter et al. 2016, Richter and Grasman 2013, Lindbeck et al. 1999).

The dynamics of processes potentially tipping the economy between equilibria depends, among other things, on how forward-looking and rational one assumes individuals to be. One approach that has been used in dynamic analyses of social norms is evolutionary game theory (see Weibull 1997, Young 2015, Young 1998b), a tool that allows the researcher to assume that individuals take payoffs of alternative strategies into account, but without necessarily being perfectly forward-looking. For example, Rege (2004), Nyborg and Rege (2003a) and Nyborg et al. (2006) assume that individuals learn their strategies through processes represented by the replicator dynamics from evolutionary biology (Taylor and Jonker, 1978). The replicator dynamics implies that the growth rate of a strategy is proportional to the current payoff of that strategy, as compared to the average current payoff in the population (Weibull 1997). A wide range of learning processes can give rise to such dynamics, for example keeping one’s current strategy unless its payoff falls short of some limit, in which case one experiments by mimicking a random other person (see Benaïm and Weibull 2003, Börgers and Sarin 1997, Weibull 1997, Gale et al. 1995).

To keep the scope of the present paper within manageable limits, I abstain from further discussion of dynamic modelling. It is worthwhile to note that even in the absence of explicit modelling of the dynamics of play, several policy relevant insights can be gained from a simple, static model like the one sketched above.

In particular, if there are multiple equilibria in the economy, policy can have an additional function (as compared to the case with unique equilibria): namely, to help coordinating the economy in a socially preferred equilibrium. The above analysis hints at two ways this may take place.

First, if a policy somehow succeeds in changing expectations about others’ behavior (changing the “focal point”), this can in itself be sufficient to change behaviors, even if material incentives are not changed.

Second, a new policy may result in the current situation no longer being an equilibrium, as in the case of a subsidy for central wastewater cleaning discussed above. The implementation of such a policy will likely result in the economy moving to another state which is still an equilibrium (the $\alpha = 1$ situation in the wastewater subsidy example). Once the tipping point has been passed, there may be no particular reason why a policy reversal would move the economy back to the initial situation, as long as the new state is an equilibrium both with and without the policy.

**Conclusions**

In the economics literature, models of social norms are usually explicitly dynamic and analytically quite complex. Here, I have chosen a somewhat different approach. My goal has been to provide an
overview relevant to environmental economists, to emphasize conceptual clarification, and to do so by means of a formal, but still non-technical presentation.

In particular, I discuss some important differences between social psychologists’ and game theorists’ approaches to social norms. I also apply a highly stylized, static framework to highlight the differences and similarities between pure and impure altruism, moral norms and social norms.

These distinctions can, at times, be crucial for environmental policy implications. For example, individuals striving to fulfill moral norms can change their behaviors in response to policies that indirectly, perhaps unintendedly, change individuals’ perceived moral obligations (Brekke et al., 2003). Such individuals may also, under certain circumstances, try to avoid information about the social consequences of their actions (Nyborg 2011). Combined with information asymmetries, moral norms can also lead to phenomena that are otherwise hard to explain within economic models, such as climate change skepticism (Asheim 2010).

Social norms, on the other hand, can involve multiple equilibria (Young 2015). In such cases, policies may contribute to tip the economy from one equilibrium to another, with potentially abrupt and dramatic changes in aggregate behaviors as a result (Rege 2004, Nyborg and Rege 2003a, Mackie 1996, Nyborg et al. 2016).

The highly simplified framework used above overlooks, of course, plenty of relevant complications, such as individual heterogeneity and distributional issues. Moreover, within the limited scope of this article, I have not analyzed even this exceedingly simple framework in depth. I still hope that this non-comprehensive review can inspire the reader to dig further into the literature on the social and moral norms of the environment.

References:


