

9. Virtue and the commons

Xavier Márquez

Introduction

Many environmental problems have the familiar structure of the dilemmas of the commons (Gardner et al. 1990), where any given individual may have reason to act in ways that result in the group being collectively worse off when everyone else acts in similar ways, so that recognisably suboptimal outcomes are produced for all if each person acts in accord with their 'private' reasons.¹ The climate change problem (Gardiner 2001, 2004), the problems caused by rapid population growth (Hardin 1968), the problems of sustaining fisheries (Ludwig et al. 1993), some problems of agricultural and forest land use, and many other environmental problems have all been argued to have this structure (see also the examples discussed and literature cited in Gardner et al. 1990).²

At the same time, the broad outlines of any solutions to these dilemmas are well known. Three kinds of solutions are possible. We might call the first kind, following Hardin (1968), *technical* solutions. Here the dilemma is resolved by directly mitigating the harm done by the behaviour, without changing the motivational structure that gave rise to the harm-producing behaviour in the first place. If we take global climate change as our example of a commons dilemma, 'geoengineering' or carbon sequestration schemes fall into this category; if the problem is the depletion of commonly owned aquifers, the construction of desalination plants would be a technical solution.

The second kind of solution comprises what we might call *external* approaches. These approaches attempt to resolve the dilemma by changing the external incentives to which the agents respond, for instance by creating institutions that reliably provide either external sanctions for 'egoistic' behaviour or positive

1 An earlier version of this chapter was presented at the Ethical Foundations of Public Policy conference and at the Victoria University of Wellington Philosophy Programme research seminar. The final version benefited from comments by participants in both places. Thanks also to Ben Thirkell-White for useful comments on an earlier draft and to Marcus Frean for a stimulating discussion on the subject.

2 As Elinor Ostrom and her collaborators have noted (Gardner et al. 1990), however, not all common resource situations are commons dilemmas. Some problems that *appear* to have the structure of the commons dilemma may be best described in some other way; see, for example, Gardiner's criticism of Hardin's depiction of the 'population problem' (Gardiner 2001).

rewards for ‘cooperative’ behaviour.³ The private reasons people might have to act in destructive ways are not changed, however; individuals are merely given potentially overriding reasons to act in a different way, so that if the institutions in question cease to operate, the individuals previously restrained by them may resume their previous (destructive) behaviour. Carbon taxes or cap-and-trade schemes fall into this category, as well as a wide variety of informal schemes that sanction some behaviours or incentivise others through social and peer pressure in different common resource situations.

Finally, the third category of solutions to commons dilemmas attempts to change the *internal* motivations of individuals so that they no longer engage in behaviour that is destructive of a commons, even in the absence of external sanctions or incentives. To use a relatively old-fashioned term, these approaches attempt to make the individuals *virtuous* rather than merely to discipline them. Any actions (for example, appeals to fair play or exemplary behaviour) or institutions whose aim (or one of whose aims) is to educate or build ‘character’ fall into this category.

These broad approaches are not necessarily independent of one another, and indeed one main purpose of this chapter is to explore how these approaches interact with one another. The other main purpose of this chapter is to examine the limits and possibilities of the *internal* approach to solving commons dilemmas, which a growing number of people, dissatisfied with external and technical approaches to these problems, have begun to promote. One needs more than incentives and sanctions or technical solutions, the claim goes; one needs a transformation of character, so that people are less likely to act in ways that harm the global commons to begin with; one needs a new ethical consciousness and new ‘virtues’ (Orr 2004; Barry 1999; Sandler 2007). Recent discussions about so-called ‘environmental citizenship’ or ‘ecological citizenship’ (Bell 2005; Dobson and Bell 2006; Hailwood 2005; Humphreys 2009; Mason 2009; Valencia Sáiz 2005; Barry 2006) also argue along similar lines that solutions to environmental commons dilemmas require specific forms of character, even if they do not always use the language of virtue. Here, the terminology of ‘citizenship,’ which denotes some bundle of rights and obligations possessed by the members of some community, is often paired, implicitly or explicitly, with a conception of the ‘virtues’ that articulate the appropriate dispositions and attitudes that the individual men and women who bear those rights and responsibilities should acquire. A common thread running through these discussions is that we can create political institutions that not only discipline our wayward (current)

3 These sanctions or incentives need not always be provided by external actors (for example, the state) in order to work, contrary to the ‘Hobbesian’ ideas of Hardin (1968) and Olson (1971); as Ostrom (2000, 1990) has argued over the years, communities can often develop rules and customary institutions that are quite capable of producing cooperation without the intervention of external agents. But the kinds of rules that Ostrom has described still represent a set of *external* incentives to behaviour in the sense described above.

dispositions (as corrupted and self-interested consumers, for example) but also (and more importantly) *educate* us and *transform* us into virtuous citizens. Instead of a politics that merely manages the deleterious effects on a particular commons of our (culturally created) dispositions, the rhetoric of virtue and citizenship strives for a politics that makes us *better* people, that is, more able to surmount our commons dilemmas without the constant use of explicit sanctions or incentives.

In this chapter, I suggest that the virtue–ethical approach to such problems, though not without merit, has important limitations. I begin by examining the meaning of ‘virtue’ in the context of commons dilemmas, and argue that we must understand such virtue as more than simply a disposition to restrain one’s use of a common resource. The ‘virtues’ appropriate to commons dilemma situations are dispositions to *actively* contribute to the solution of the commons dilemma.

But, as I argue, there are two kinds of virtue, which I call ‘robust’ and ‘conditional’. ‘Robust’ virtues are dispositions to contribute to the solution of the collective action problems presented by commons dilemmas that operate across a wide range of such dilemmas. On the other hand, ‘conditional’ virtues are dispositions to contribute to the solution of such dilemmas only under specific conditions, such as the fraction of other conditionally virtuous users of the common resource or the costs of contributing to the solution. Empirical evidence suggests that ‘robust’ virtues are rare, so cannot be relied on to solve the many commons dilemmas we confront today.

While conditional virtue is more common, I show that its ‘supply’ is an endogenous effect of the potential solutions to the commons dilemma that virtuous agents may promote. Moreover, since different potential solutions have opposed effects on this supply, virtuous agents must evaluate trade-offs between solutions that can deal with urgent commons dilemmas (which may over time decrease the supply of virtue) and solutions that increase the long-term supply of more robust virtue (which tend to be slow and unreliable).

Given plausible epistemic constraints on virtuous agents, this means that even virtuous agents will face severe collective action problems in deciding how to best contribute to a ‘difficult’ collective action problem (such as the problem of global climate change). This argument suggests there are sharp limits to any reliance on the inculcation of virtue for a solution to large-scale dilemmas of the commons. What virtue demands is not necessarily *more* virtue but more intelligent investment in various potential solutions to these dilemmas, including technical and external solutions.

Virtue in the context of commons dilemmas

The word 'virtue' has many meanings, not all of which are relevant to questions about the dilemmas of the commons. Moreover, as with any significant term in moral philosophy, the term is subject to a great deal of controversy. But we do not go far wrong if we define a virtue as a relatively stable affective and deliberative *disposition* or *character trait* (that is, an identifiable and consistent pattern of affective and deliberative responses to similar situations) that gives an agent *good reasons* to perform actions that systematically promote good ends in a variety of contexts. What is considered a good end (for example, the agent's natural flourishing) determines what counts as a good reason. A virtuous person is thus more *sensitive* or *practically responsive* to certain kinds of reasons for action than is a non-virtuous person, in the sense that a virtuous person will find certain kinds of (good-promoting) reasons more likely to motivate them to action than to motivate a non-virtuous person, and will be more likely to find such reasons *salient* in any given situation. A classic example is the virtue of courage. The courageous person, in contrast to the non-courageous person, has less sensitivity to *the potential for bodily harm* as a reason not to do something, and more sensitivity to *protecting the community* as a reason to do the same.⁴

Virtues thus have both 'cognitive' and 'affective' components that combine to enable an agent to 'respond appropriately' to a situation (Kamtekar 2004, p. 477), where 'appropriately' has a broad ethical meaning. Conversely, character traits are to be called virtues if by making the agent appropriately sensitive to certain kinds of reasons to perform actions in a variety of contexts, they systematically promote good ends (Sandler 2007; Driver 2001; Hursthouse 2001). We leave aside here the problem of the precise nature of those ends in general (for example, whether these ends are always agent-relative or should include 'noneudamonic' ends). We simply assume that in any account of virtue character traits that systematically promote such ends can be called virtues, precisely in virtue of the fact that possessing a virtue makes a person sensitive to good reasons for acting.

From this point of view, a 'virtue' in the context of a commons dilemma can be defined as a stable character trait that provides an agent with good reasons to act in ways that systematically preserve the resource held in common. However, though such virtues will usually include such traditional traits as self-restraint or moderation (using only those resources that are sufficient for one's real needs) and justice (giving each what is due to them, and taking only what one is due), they cannot be restricted to them. In particular, given the structure of a typical

4 At the extreme, a virtuous person would be perfectly sensitive to *good* reasons for action (whatever those may be) and perfectly insensitive to *bad* reasons for action (whatever those may be). In this sense, virtue is a kind of knowledge, as in the Socratic formulation.

commons dilemma, mere self-restraint is often impotent to preserve a resource; the common resource is preserved only if others do the same. But the virtuous person would not be satisfied with restraining their consumption of a resource in the face of general indifference if they knew that more could be done to preserve the resource; in fact, the virtuous person would actively seek to *foster cooperation* (very broadly speaking) with others so as to preserve it. Mere self-restraint would seem to be *part* of what it means to have a virtuous disposition with respect to the commons, but not the whole of it. Thus, if a virtue in the context of a commons dilemma is a disposition to contribute (broadly speaking) to the solution of the commons dilemma, then a virtuous person should *invest resources* in finding and promoting the adoption of a solution to the dilemma. Hence, a virtuous person would decide whether they should support and promote a technical, an incentive-based, or an educational solution to the problem. In other words, the virtuous person is an *activist*; they are sensitive to reasons to act in ways that promote the solution of the dilemma, ways that will normally go beyond mere self-restraint on their part.

Robust and conditional virtues

It is also important to avoid understanding virtue as a sensitivity to reasons to cooperate in commons dilemmas that is *independent* of context. Virtue theorists assume, plausibly, that virtues can be produced, with greater or lesser reliability, through either intentional educational processes or as a by-product of other interactions (for example, formal schooling, participation in democratic institutions, or the other usual forms of socialisation). They also, less plausibly, tend to assume that these habits of character are not narrowly tailored to specific contexts (that is, are not merely 'local character traits') but are capable of furthering good ends in a variety of such contexts (that is, are 'global character traits', stable across multiple contexts). On this view, a virtuous person is not only sensitive to reasons for cooperation in some particular subset of commons dilemmas, but is well disposed to perform such actions in most of them, so long as such actions are not otherwise impossible or incoherent from the point of view of the person's ultimate goals or values. But it is an open question whether such character traits *can* exist (Kamtekar 2004; Doris 2002; Appiah 2008; Annas 2005; Arpaly 2005; Solomon 2005); people who perform virtuous actions in some contexts may not perform virtuous actions in other contexts that differ only in trivial respects. More to the point, people who are well disposed to cooperation in some commons dilemmas may not be well disposed to cooperate in trivially different situations. Indeed, research on cooperation in the commons shows that such cooperation, though much more easily achieved in some contexts than a standard rational choice model would lead us to expect, can just as easily be

disrupted by relatively minor changes in incentives or beliefs (Camerer and Fehr 2006; Ostrom et al. 1999; Bicchieri 2008). Population growth, the introduction of new technologies, changes in sanctioning opportunities, framing effects that indicate which norm is operative, an inability to identify with other common resource users, all appear to easily turn responsible users of the resource into irresponsible ones, and 'virtuous' people into 'selfish' agents. This suggests that whatever character traits may underlie cooperation in commons dilemmas (or in general), they tend not to be 'robust' or 'global' but rather situation-specific. In other words, they are not 'virtues' in the strong sense of the term prominent in the work of some virtue ethicists (for example, Hursthouse 2001). Sensitivity to the right reasons for acting or not acting is *conditional*, not *global*, and *fragile*, not *robust*.

This 'situationist' challenge to virtue ethics (Harman 2009; Doris 2002) should be taken seriously, for it suggests that no solution to commons dilemmas can rely on the inculcation of 'global' virtues, be they green or otherwise; human beings do not appear to be well suited to the development of such 'robust' virtues. But this does not mean we cannot speak of virtue at all, or that the notion of virtue is useless for thinking about the solution of commons dilemmas. Instead, we should distinguish between the ethical ideal of 'robust' virtue, representing a (perhaps empirically impossible) consistent disposition to reason, feel, and act in ways that contribute to the solution of commons dilemmas *generally*, and more or less conditional virtues, representing dispositions to act in ways that promote the solution of *some* commons dilemmas in *some* circumstances. The virtues that are normally found among human beings are of the second kind, that is, they are 'conditional' or 'relative' virtues, which may of course be more or less robust (that is, they may approach the ideal for some individuals and some range of situations). Individuals may still look to the ideal of robust virtue as something that they should strive for even if they cannot quite achieve it, but we do not need to assume the possibility or actual existence of fully robust virtue in order to talk meaningfully about the many relatively virtuous dispositions that can be instilled in people with greater or lesser degrees of reliability.

On the other hand, we also need to be able to distinguish between actual virtue and the mere conditional willingness to cooperate that large numbers of people display in a wide variety of settings. As research on cooperation in the commons has shown, many people (somewhere between 40 per cent and 70 per cent of experimental subjects, with some variation due to cultural differences, learning, and the specific stakes involved) are 'conditional cooperators'. That is, they are disposed to cooperate (restrain their use of resources, contribute to the production of a common good, and so forth) even in the absence of purely external incentives (for example, in one-shot prisoner's dilemma or ultimatum games) but only if they understand well the consequences of general non-

cooperation and estimate that enough others will cooperate as well (Camerer and Fehr 2006; Henrich et al. 2001; Ostrom 2000; Chuah et al. 2009; Dawes 1980; Roth et al. 1991; Kurzban and Houser 2005; Fischbacher et al. 2001). The mere presence of conditional cooperators is not enough to induce high levels of cooperation (indeed, it is compatible with very *low* levels of cooperation, if self-regarding individuals trigger a cascade of non-cooperation). However, high levels of cooperation can be sustained in experimental settings if subjects are given the opportunity to communicate and sometimes ‘punish’ non-cooperators even at some cost to themselves, an opportunity that some fraction of conditional cooperators usually take (Camerer and Fehr 2006; Ostrom et al. 1992; Bicchieri 2008). These altruistic punishers (or ‘strong reciprocators’), that is, people who are willing to expend some of their own resources in sustaining cooperation, are essential to the maintenance of high levels of cooperation.

Yet though such ‘altruistic punishers’ display some of the features of virtuous individuals, it is clear that such people are not necessarily virtuous, even in the ‘conditional’ sense discussed above.⁵ Altruistic punishers are not (always) appropriately sensitive to the effects of their actions on actual levels of cooperation, and they do not (always) adequately consider whether means other than punishment would sustain cooperation. A virtuous person, by contrast, is characterised by a sensitivity to reasons for punishing others when punishment is appropriate, or to invest resources in other potential solutions to the commons dilemma when other solutions are appropriate. Therefore, the virtuous person would be able to support or promote, when appropriate, not just schemes to discipline non-cooperators, but also schemes to mitigate the harm of non-cooperation or to increase the number of virtuous individuals. Virtue would involve here both a *motivational* component, disposing the virtuous individual to *want* to achieve collectively optimal outcomes and to react appropriately in emotional terms to the failure of cooperation (a component that is present among altruistic punishers), and an *epistemic* component, indicating that the virtuous individual has better than average understanding of *which* means can best resolve the common resource dilemma in question.

Investments in cooperation and the supply of virtue

But which means can best resolve commons dilemmas? As mentioned above, there are only three generic kinds of solutions to these dilemmas: technical,

⁵ Such a character trait can be likened to the ‘natural virtue’ that Aristotle opposes to genuine virtue on the grounds that the former is a mere tendency, uninformed by practical reason to a significant degree (*Nicomachean Ethics* 1144b5-22); see Kamtekar (2004, p. 480).

external (incentive schemes), and internal (educational interventions). Virtuous agents will have reason to invest resources in developing or promoting one or more of them, according to their social position, their personal resources, and their individual judgements of the relative feasibility of the particular options on the table. Virtuous agents may agitate for particular technical solutions or incentive schemes (for example, a particular legal regime or a particular irrigation technology) or directly provide incentives to others to engage in particular sorts of behaviour (for example, they may punish others for violating rules in small-scale settings). They may engage in 'exemplary' behaviour in the hopes that others may imitate them (for example, they may engage in conspicuous non-consumption or conspicuous recycling). They may promote the creation of certain institutions in the hopes that they will build certain kinds of character (for example, they may promote certain kinds of participative institutions). They may 'raise awareness' (providing information to others). From a *static* perspective, virtuous agents do not have an *a priori* reason to prefer one of these types of solutions to the others (for example, to prefer 'punishment' to 'education' or to the technical amelioration of the problem), since the effect of any of them is, considered in isolation (and assuming equal effectiveness), equivalent: a collective good is produced or a commons is preserved. In economic language, they are perfect 'substitutes' for each other. But because human virtue is not robust, each potential solution has different effects on the fraction of (conditionally) virtuous agents in the population (the 'supply' of virtue), at least relative to the particular commons dilemma in question, and these potential effects need to be considered by virtuous agents when deciding how best to respond to a commons dilemma. Let us examine this matter in more detail.

Technical solutions

Technical solutions have the effect of decreasing the collective costs of non-cooperation in a particular commons dilemma. For example, assume for the moment that there existed some technical solution to a commons dilemma: some geoengineering scheme for global warming, such as the (still fictional) 'carbon-eating super-trees' that Freeman Dyson has promoted (Dyson 2008), or some form of renewable, carbon-neutral, and cheap energy. Moreover, assume that such a scheme were generally effective and cheaper to implement than taxing carbon. The presence of this technology would make collective 'defection' (that is, continued use of fossil fuels at current levels) less costly than otherwise, and thus would make it possible to solve the problem by means of the same motivations that gave rise to it in the first place. For a more realistic example, we might point to the spread of technologies of drip irrigation or desalination plants, which might (under some circumstances) mitigate the harm involved

in the overexploitation of aquifers in arid lands, allowing users to continue to consume water at levels that would have produced the overexploitation of the resource in the absence of the technology.⁶

The more effective a technical solution, the greater the incentives for non-cooperation regardless of any incentive or sanctioning schemes; indeed, under a sufficiently effective technology, even virtuous agents may ‘defect,’ since the dilemma may appear to be solved (the force of the reason to refrain for consumption being considerably weakened). Moreover, if virtue is a kind of habit, and thus a matter of *practice*, as virtue theorists have argued since Aristotle, then if an agent lacks consistent opportunity to practice virtue they will (slowly or quickly) lose it, at least for the particular common resource situation in question. Technical solutions will, thus, tend to decrease the existing supply of virtue in a particular commons dilemma through what economists sometimes call ‘moral hazard’, sometimes to such an extent that their effectiveness as solutions to these dilemmas may be undermined.⁷

A technical solution will also tend to affect negatively the *future* supply of virtue in the particular commons dilemma. For example, if a particular geoengineering scheme (for example, carbon-eating super-trees) mitigated the harm from carbon emissions, then there would be less incentive to *become* the kind of person who is spontaneously willing to adopt a less emissions-intensive lifestyle in the absence of external incentives. Similarly, by breaking the connection between harm and behaviour a technical solution may make it increasingly difficult to tell *what* exactly would constitute virtuous behaviour. In both cases the future supply of virtue decreases.

6 We ignore for the moment the possibility that in making defection less costly in some particular commons, a technical solution may create more problems in some other context or at some other time. It is clear, for example, that some geoengineering schemes (such as pumping sulphur into the atmosphere) would likely have bad side effects elsewhere (for example, disrupted rain patterns, a higher risk of fast warming if the scheme were stopped, increased risks of ocean acidification, not to mention various unwelcome political implications). See Rasch et al. (2008) for an overview of research into the possible climate effects of pumping sulphur into the stratosphere, and see Schneider (2008) and Keith (2000) for discussion of its political risks and moral hazard. Similarly, the use of desalination plants might accelerate the destruction of some marine ecosystems or, given their large energy use, contribute to global climate change.

7 For example, the use of biofuels, especially ethanol, in cars was initially conceived as a way to diminish the climate impact of the burning of gasoline, and indeed considered in isolation the burning of ethanol has less of an impact on the climate than the burning of gasoline. But the encouragement of ethanol production through legislative mandate in the United States seems to have induced an increase in land clearing for biofuel purposes, which *increases* the greenhouse gas emissions of ethanol production relative to the burning of gasoline (Searchinger et al. 2008). A (partial) technical solution to the climate problem was thus self-undermining when widely adopted.

Incentive schemes

Incentive schemes, by contrast with technical solutions, increase the benefits of individual cooperation and decrease the individual benefits of non-cooperation, making collective cooperation more likely. These incentives may be provided in a variety of ways: social and peer pressure, sanctions by specific users of the common resource (for example, altruistic punishment), sanctions by non-users of the resource (for example, the state), and explicit property regimes (for example, privatisation). An incentive scheme is not identical with *Leviathan*.

A perfectly efficient incentive scheme would be one that induces everyone (virtuous or not) to cooperate. Though the efficiency of an incentive scheme depends on many factors (for example, the monitoring technology available and material resources) it is worth noting such schemes tend to display ‘increasing returns’ to virtue, or at least have low virtue and high virtue equilibria. In other words, incentive schemes work better the greater the supply of virtue, since in that case fewer resources will need to be devoted to monitoring and sanctioning defectors. Moreover, such dependence may generate virtuous or vicious loops, where, for example, low-corruption equilibria may tend to induce more ‘virtuous’ behaviour (given the high costs of non-virtuous behaviour), which may reduce the need to monitor behaviour and sustain trust further. On the other hand, high-corruption equilibria may tend to induce less virtuous behaviour (given the low costs of non-virtuous behaviour), which may further reduce the amount of resources devoted to monitoring and punishing cheaters, leading to even more corruption.

Incentive schemes may have not only ‘motivational’ effects (affecting the payoffs for cooperation or non-cooperation of everyone, not only the virtuous), but may also have ‘epistemic’ effects (affecting the ability of the virtuous to do the right thing).⁸ Consider the problems a person who wants to minimise their carbon emissions while retaining some of the comforts of civilisation faces today. Should they use paper or plastic bags? Buy frozen or fresh vegetables? Use dishwasher detergent? The ‘right’ decision under current circumstances here depends on a wide variety of highly controversial and technical judgements about energy use throughout the entire lifecycle of the product (Goleman 2009). But if there existed an incentive scheme that put an appropriate price on carbon, the decision might become straightforward, as the price of the item would already embody the appropriate signal regarding its carbon content. A well-designed incentive scheme, thus, aligns both motivation and information, making non-virtuous action (both intentional and unintentional) *harder* than virtuous action, though of course an ill-designed incentive scheme would do

⁸ I first became aware of this point by reading Matthew Yglesias’s popular blog (<http://yglesias.thinkprogress.org>).

the opposite. Yet this very advantage of incentive schemes may induce a kind of ‘epistemic’ moral hazard. For example, a price on carbon may make it easier for the virtuous to act properly, but in the absence of appropriate price signals such behaviour may quickly go astray. By ‘externalising’ epistemic responsibility for the consequences of one’s actions certain incentive schemes may make virtuous character *harder* rather than easier to achieve over time, reducing its future supply.

Incentive schemes might affect the supply of virtue in a given commons in two other ways. On the one hand, since virtue is partly a matter of habituation, practice in cooperation, though at first prompted by external incentives or sanctions, may, if sustained for long enough, become internalised.⁹ We might call this the *socialisation* mechanism. Evidence from examples of successful management of commons dilemmas in small-scale settings (Gardner et al. 1990; Ostrom et al. 1992) suggests that a certain amount of socialisation is likely to occur when users of a common resource have a large role in designing the rules of their interaction and ample opportunity for communication. Moreover, there is good evidence that people do not like to be ‘suckers’: they may willingly cooperate in the solution of some commons dilemma only if they are assured (by some impartial mechanism, such as a law) that others will also cooperate. If people have evidence (for example, based on past rates of cooperation) that other people will cooperate, then they will be more sensitive to reasons to cooperate, or, in other words, reasons to cooperate will have more force for them.

On the other hand, a body of psychological literature on motivation (Frey 1994; Camerer and Hogarth 1999) suggests that the offer of external rewards can undermine internalised motivations under some circumstances. Virtuous people may become accustomed to cooperating primarily through external rewards, and so may over time lose their internal motivation (they may become *less* sensitive to the right reasons to cooperate in the absence of the ‘external’ reasons). Indeed, evidence from examples of failed management of commons dilemmas (also in Gardner et al. 1990) suggests that certain ‘external’ interventions (by the state, for example) can undermine ‘internal’ motivation (by shifting ‘the locus of control’) and lead to higher rates of opportunistic behaviour, especially in the absence of sufficient monitoring (see also Bowles 2008). We might call this the *corruption* mechanism. Whether socialisation or corruption dominates, however, cannot be determined *a priori*, since the net effect will depend on the specific details of the incentive scheme in question.

9 See Kuran (1997) who suggests a ‘hidden’ preference (in this case, the preference for defection, which incentive schemes make it appear as a preference for cooperation) can be transformed over time by social pressure into an internalised norm, that is, a true preference.

Educational interventions

Educational interventions counteract the corrosive effects of both technical solutions and incentive schemes by *directly* (rather than as a side effect) attempting to increase the supply of virtue (or rather, the number of virtuous individuals relative to a particular commons dilemma). We should understand such interventions very broadly. Exemplary behaviour, for example (consider the idea of ‘living in truth’ in Havel 1992), may enable other people to learn what virtue requires and to produce not only people who cooperate out of shame (that is, increase the number of cooperators) but people whose preferences shift towards cooperation (that is, increase the number of virtuous agents over time). People may become sensitive to the reasons for cooperation *made salient* by such exemplary behaviour. Certain institutions may also have long-term positive effects on the character of their users, and hence their creation would count as a kind of educational intervention. Thus, for example, some people advocate for the expansion of participatory and deliberative democratic institutions for their supposed benefits on the character of participants in them: such people, it is claimed, can take a broader, less selfish view of the problems confronting them and their communities (Warren 1992). That is, their participation in such institutions makes them more sensitive to the right sorts of reasons for action, though it is worth repeating that these benefits may exist only relative to the particular commons managed by those institutions, given the lack of robustness of human virtue.

In general, educational interventions can be usefully divided into interventions affecting the *motivation* of agents (turning some ‘selfish’ agents into ‘cooperative’ agents) and interventions affecting the *knowledge* of agents (increasing the availability of information about the right thing to do for individuals who are already motivated to act cooperatively), though of course some may do both at the same time. The second sort of intervention, which includes such activities as raising awareness of a problem and direct teaching about things to do to help, may not increase the proportion of genuinely virtuous individuals in the population, but may, nevertheless, encourage people disposed towards cooperation to do the right thing, so to speak. They may increase (or decrease) the *effectiveness* of virtue through their epistemic effects.

It seems plausible to think that this second sort of educational intervention (whose effects are mainly epistemic) may be generally more effective than the first. People can and do change their behaviour if they encounter new information in an unprejudiced way. However, they are also subject to all sorts of cognitive and other biases (for example, confirmation bias, cognitive dissonance reduction, and wishful thinking) that may lead them to discount information that, according to their own goals and values, may result in costly

behaviour changes.¹⁰ Moreover, though direct evidence on this point is scarce, the testimony of history and classical political thought suggest that enduring changes in motivation are difficult to accomplish quickly, intentionally, and on a large scale. Such changes are perhaps possible in the long term, on a small scale, or as unintended consequences of other changes, but may be less likely in the short term, on a large scale, or in a “planned” way. Directly increasing the number of (more or less robustly) virtuous individuals in entire societies, on any reasonable understanding of virtue, seems to be much harder to accomplish than making it easier for the less robustly ‘virtuous’ individuals to act on their unformed prosocial inclinations or enforcing overt compliance with ‘virtuous’ behaviour.

For example, direct ‘indoctrination’ seems not to work to produce virtuous individuals, though it may change some of their beliefs as to what should count as appropriate or inappropriate behaviour. But since virtuous activity is not merely a matter of ‘correct’ belief, requiring as well correct judgement and motivation, such belief changes may not have any impact on the actual proportion of virtuous individuals. For example, even if individuals come to believe that protecting the environment and acting to prevent anthropogenic climate change is an important thing (as polls suggest they do), this does not mean they will be motivated to invest many resources in solving the problem, especially if they perceive the immediate costs to be large and the benefits uncertain and small. Decades of ‘prosocial’ indoctrination in many Soviet countries and China seems to have failed rather spectacularly at increasing the supply of virtue relative to typical commons dilemmas. Furthermore, since beliefs induced by indoctrination are sustained merely by ‘social proof’ (rather than the everyday feedback given by the world for some of our other beliefs) they will at any rate tend to be rather fragile and subject to unexpected and unpredictable changes or intentional manipulation (Kuran 1997).

Even if intentional action by powerful agents (such as the state) can increase the number of relatively virtuous individuals relative to some set of commons dilemmas, this seems to be possible only in small-scale contexts. We find either large-scale cultural change that is not centrally planned (as in the ‘civilising process’ documented by Norbert Elias (1978)) or intentional cultural change that is only sustained through the creation of sharp exclusion boundaries in relatively small groups (as in many of the successful cases of common resource

10 For an overview of some of these biases, see Elster (2007, especially chapters 7, 11–12). For a discussion of how these biases may negatively affect responses to climate change, see Sunstein (2007). It should be noted that *misinformation* may be more destructive of virtuous behaviour than direct attempts to change the internal motivations of prosocial individuals. Many people do want to ‘do the right thing,’ as the phrase goes, and it may be difficult to convince them to act in ways they see as socially harmful unless the costs of not acting in selfish ways are large or they are ignorant of the consequences of their behaviour.

management documented by Gardner et al. (1990)).¹¹ The institutions of Sparta reliably produced ‘courageous’ warriors, and certain isolated communes can perhaps produce environmentally conscious individuals, but these achievements are exceedingly fragile, difficult to replicate, and dependent on the ability of the communities in question to exclude others. Overall, then, we may be justified in thinking that educational interventions have limited effectiveness to shift the balance of virtue in a large-scale society, though they may be more effective in smaller scale societies or over the (very) long term.

We should note that just as incentive schemes may have ‘socialisation’ effects, educational interventions may also have ‘incentive’ effects. The example of the educational activities of others may induce some cooperation in some individuals out of shame, conformism, or other mechanisms. Moreover, through the socialisation mechanism such cooperation, if sustained over time, may make some people internalise these cooperative norms. But since the effect is indirect, it will tend to happen slowly and unreliably, and it may be easily reversed.

Limits of virtue

This discussion of the effects on the supply of virtue of the different potential solutions to the commons dilemma should be sufficient to show that virtuous agents face real trade-offs when trying to decide which of these options to invest in to resolve a particular commons dilemma. This is so even before considering the effects of these options in the supply of virtue in other parts of society or the very large epistemic constraints under which they make their decisions. In particular, a virtuous agent may be willing to risk a decline in the supply of virtue if a particular commons dilemma seems especially pressing or alternative solutions especially intractable.

Consider again the example of geoengineering. Though the risks of geoengineering are large, and its ‘technical’ nature rewards irresponsible behaviour (after all, it seems not to address the ‘root causes’ of the problem of climate change), it may be that the potential for catastrophe implicit in global climate change and the difficulty of other solutions mean geoengineering proposals should be seriously considered.¹² This is not to say that the virtuous agent will in general look

11 It is not entirely clear *how* sharp exclusion boundaries work to increase the supply of virtue. On one understanding, what matters is consciousness of group membership, which makes individuals sensitive to reasons favouring group welfare over individual welfare; on another, what matters is the fact small groups make it easier to coordinate on expectations about operative norms, which makes individuals sensitive to reasons embodied in these particular norms (Bicchieri 2002).

12 For a technical discussion of the potential for catastrophe implicit in global climate change, see Weitzman (2009). Weitzman uses the standard utilitarian assumptions of economic analysis to argue that geoengineering proposals should be considered in any mix of potential solutions, but my point is that even from the point of view of virtue ethics one cannot escape such calculations entirely.

with favour to technical solutions, given their generally negative effects on the supply of virtue, but only that the virtuous agent may not entirely ignore such considerations, even if they remain sceptical.¹³

Effective technical solutions, incentive schemes, and educational interventions in the context of particular commons dilemmas are themselves *public goods* (Gardner et al. 1990), since benefits from their existence accrue to all users of the common resource. Hence, their production presents secondary collective action problems in addition to the primary collective action problem represented by the commons dilemma. This secondary collective action problem may not be of the same order as the primary one (indeed, to the extent that the primary collective action problem is solvable, it may be only because the secondary one is easier to resolve than the primary one), but it, nevertheless, exists as a problem *even for virtuous agents* who are disposed to cooperate. This is so because virtuous agents need to coordinate on beliefs; even if they all want to invest resources in solving the problem, they may not agree *which* solution will best deal with the problem.

To simplify greatly, imagine that the primary commons dilemma has the structure of an N-person prisoner's dilemma. Three solutions are possible: a technical scheme that increases the payoff to mutual defection, an incentive scheme that increases the payoffs to unilateral cooperation and decreases the payoffs to non-cooperation, and an educational intervention that turns some significant proportion of the players into altruists who will cooperate more or less unconditionally. Suppose now that any of these alternatives may be produced only if all virtuous individuals (who form only a fraction of the population) support them, and that all are equally effective and known to be so. Nevertheless, given their effects on the proportion of virtuous individuals over time and their assessments of the urgency of the primary problem, some virtuous agents prefer the technical solution, some the incentive scheme, and some the educational intervention. But if they do not agree on a *single* solution, no solution will be implemented, and the primary collective action problem (the commons dilemma) will not be solved. In game-theoretic terms, the solution of the prisoner's dilemma game depends here on the prior solution of a 'battle of the sexes'-type game (a pure coordination dilemma). Since pure coordination dilemmas are easier to resolve than prisoner's dilemmas (they sometimes have 'internal' solutions, or focal points, that may guide the choice of strategies), such a commons dilemma stands a good chance of being resolved. However, this

13 Ideally, the best policy for a virtuous agent from the point of view of the supply of virtue might be an incentive scheme with positive feedbacks from the supply of virtue, combined with an educational strategy. But such a policy might not exist, or it may not be known, or it may work only in the long run.

may not always be the case, especially in the face of significant uncertainties about the effectiveness of particular solutions and large disagreements about the urgency of the particular commons problem.

A more realistic depiction of the secondary collective action problem involved in producing a solution to a commons dilemma would recognise that sometimes partial solutions to commons dilemmas can be partially effective (that is, that unanimity may not be necessary, and full coordination may not be required). However, in general the production of solutions to commons dilemmas will be more effective the easier this secondary collective action problem is to resolve. On the other hand, different solutions to a commons dilemma may present different secondary collective action problems, a consideration that should affect the potential calculus of virtuous individuals wishing to resolve it. For example, one attraction of geoengineering schemes as a solution to global climate change is that they seem to present a far easier to resolve secondary collective action problem than the production of global incentive scheme to regulate carbon, even if this additional ease is purchased at the cost of creating *other* collective action problems elsewhere, in particular problems of governance (Schelling 1996; Barrett 2008; Schneider 2008; Victor 2008).¹⁴

There are several ready-made ways to resolve these secondary collective action problems to address the primary commons dilemma. In the context of a democratic state, one would expect that when the supply of virtue for some commons dilemma is low (and hence incentive schemes or technical solutions impossible to implement effectively), the virtuous should focus on educational interventions to increase their own number. They should do this until a threshold is reached that allows them to use the usual coordination opportunities provided by ordinary democratic politics to impose a particular incentive scheme that 'solves' the problem. This scheme would then be implemented through the action of a state (which itself may be conceived as an 'already solved' collective action problem for some range of issues).

This is the pattern found in many successful 'social movement' campaigns (Tilly 2006). A long 'education' campaign increases the number of people willing to sacrifice something for the sake of a social benefit, until a point is reached at which the 'virtuous' are numerous enough to impose their preferences on the rest of the population through legal regulation (consider here, for example, the 19th century campaigns against slavery or the 20th century campaigns against smoking). Such a pattern of collective action is more difficult where substantial disagreement exists among the virtuous regarding the appropriate forms of

14 It is also not entirely clear that the secondary collective action problems presented by geoengineering schemes really *are* easier to resolve than the secondary collective action problems presented by global incentive schemes.

education (since different forms of education may undercut one another) or where significant uncertainty exists about the problem (which may give rise to contradictory information about it) or where others deliberately attempt to use misinformation to protect their interests (which diminishes the effectiveness of educational interventions). Moreover, depending on how ‘robust’ such education is and on the specific characteristics of the incentive scheme eventually imposed through legal means, such a scheme may then prove self-reinforcing (through socialisation mechanisms) or it may be undermined in the short or long term, necessitating further campaigns.

Such ready-made solutions to the secondary collective action problems facing the virtuous are not, however, easily available in many large-scale commons dilemmas. The art of the ‘virtuous’ activist (indeed, the art of politics, if one wants to be grandiose about it) is instead an art of ‘bootstrapping’ – finding the right ‘games’ where collective action can be easily organised (for example, where the number of the virtuous is already large or can be easily increased) and using collective action there to affect collective action in other games and at other scales. This is usually not easy, given the complexity of modern societies; and the virtuous should not count on being able to resolve these problems by insisting on long-term educational schemes for the large-scale transformation of character.

References

- Annas, J. 2005. ‘Comments on John Doris’s *Lack of Character*.’ *Philosophy and Phenomenological Research* 71(3): 636–42.
- Appiah, A. 2008. *Experiments in Ethics*. Cambridge, MA: Harvard University Press.
- Arpaly, N. 2005. ‘Comments on *Lack of Character* by John Doris.’ *Philosophy and Phenomenological Research* 71(3): 643–7.
- Barrett, S. 2008. ‘The incredible economics of geoengineering.’ *Environmental and Resource Economics* 39(1): 45–54.
- Barry, J. 1999. *Rethinking Green Politics: Nature, virtue, and progress*. Thousand Oaks, CA: Sage.
- Barry, J. 2006. ‘Resistance is fertile: From environmental to sustainability citizenship.’ In A. Dobson and D. Bell (eds). *Environmental Citizenship*. Cambridge, MA: MIT Press.

- Bell, D. 2005. 'Liberal environmental citizenship.' *Environmental Politics* 14: 179–94.
- Bicchieri, C. 2002. 'Covenants without swords: Group identity, norms, and communication in social dilemmas.' *Rationality and Society* 14(2): 192–228.
- Bicchieri, C. 2008. 'The fragility of fairness: An experimental investigation on the conditional status of pro-social Norms.' *Nous (Philosophical Issues 18 Interdisciplinary Core Philosophy)* 18: 227–46.
- Bowles, S. 2008. 'Policies designed for self-interested citizens may undermine "the moral sentiments": Evidence from economic experiments.' *Science* 320(5883): 1605–9.
- Camerer, C. F., and E. Fehr. 2006. 'When does "economic man" dominate social behaviour?' *Science* 311(5757): 47–52.
- Camerer, C. F., and R. M. Hogarth. 1999. 'The effects of financial incentives in experiments: A review and capital–labor–production framework.' *Journal of Risk and Uncertainty* 19(1): 7–42.
- Chuah, S., R. Hoffmann, M. Jones, et al. 2009. 'An economic anatomy of culture: Attitudes and behaviour in inter- and intra-national ultimatum game experiments.' *Journal of Economic Psychology* 30(5): 732–44.
- Dawes, R. M. 1980. 'Social dilemmas.' *Annual Review of Psychology* 31(1): 169–93.
- Dobson, A., and D. Bell. 2006. *Environmental Citizenship*. London: MIT Press.
- Doris, J. M. 2002. *Lack of Character: Personality and moral behaviour*. New York: Cambridge University Press.
- Driver, J. 2001. *Uneasy Virtue*. Cambridge, UK: Cambridge University Press.
- Dyson, F. 2008. 'The question of global warming.' *New York Review of Books* 55(10).
- Elias, N. 1978. *The Civilizing Process*. 1st American edn. New York: Urizen Books.
- Elster, J. 2007. *Explaining Social Behaviour: More nuts and bolts for the social sciences*. Cambridge: Cambridge University Press.
- Fischbacher, U, S. Gächter, and E. Fehr. 2001. 'Are people conditionally cooperative? Evidence from a public goods experiment.' *Economics Letters* 71(3): 397–404.

- Frey, B. S. 1994. 'How intrinsic motivation is crowded out and in.' *Rationality and Society* 6(3): 334–52.
- Gardiner, S. M. 2001. 'The real tragedy of the commons.' *Philosophy and Public Affairs* 30(4): 387–416.
- Gardiner, S. M. 2004. 'Ethics and global climate change.' *Ethics* 114(3): 555–600.
- Gardner, R., E. Ostrom, and J. M. Walker. 1990. 'The nature of common-pool resource problems.' *Rationality and Society* 2(3): 335–58.
- Goleman, D. 2009. *Ecological Intelligence: How knowing the hidden impacts of what we buy can change everything*. New York: Broadway Books.
- Hailwood, S. 2005. 'Environmental citizenship as reasonable citizenship.' *Environmental Politics* 14: 195–210.
- Hardin, G. 1968. 'The tragedy of the commons.' *Science* 162(3859): 1243–8.
- Harman, G. 2009. 'Skepticism about character traits.' *Journal of Ethics* 13(2): 235–42.
- Havel, V. 1992. *Open Letters: Selected writings, 1965–1990*. 1st Vintage Books edn. Edited by P. Wilson. New York: Vintage Books.
- Henrich, J., R. Boyd, S. Bowles, et al. 2001. 'In search of homo economicus: Behavioural experiments in 15 small-scale societies.' Paper read at 113th Annual Meeting of the American Economic Association, New Orleans, Louisiana, 5–7 January.
- Humphreys, D. 2009. 'Environmental and ecological citizenship in civil society.' *International Spectator* 44: 171–83.
- Hursthouse, R. 2001. *On Virtue Ethics*. Oxford: Oxford University Press.
- Kamtekar, R. 2004. 'Situationism and virtue ethics on the content of our character.' *Ethics* 114(3): 458–91.
- Keith, D. W. 2000. 'Geoengineering the climate: History and prospect.' *Annual Review of Energy and the Environment* 25(1): 245–84.
- Kuran, T. 1997. *Private Truths, Public Lies: The social consequences of preference falsification*. Cambridge, MA: Harvard University Press.
- Kurzban, R., and D. Houser. 2005. 'Experiments investigating cooperative types in humans: A complement to evolutionary theory and simulations.' *Proceedings of the National Academy of Sciences of the United States of America* 102(5): 1803–7.

- Ludwig, D., R. Hilborn, and C. Walters. 1993. 'Uncertainty, resource exploitation, and conservation: Lessons from history.' *Science* 260(5104): 17–36.
- Mason, A. 2009. 'Environmental obligations and the limits of transnational citizenship.' *Political Studies* 57: 280–97.
- Olson, M. 1971. *The Logic of Collective Action: Public goods and the theory of groups*. Rev. edn. New York: Schocken Books.
- Orr, D. W. 2004. *Earth in Mind: On education, environment, and the human Prospect*. 10th anniversary edn. Washington, DC: Island Press.
- Ostrom, E. 1990. *Governing the Commons: The evolution of institutions for collective action*. Cambridge: Cambridge University Press.
- Ostrom, E. 2000. 'Collective action and the evolution of social norms.' *Journal of Economic Perspectives* 14(3): 137–58.
- Ostrom, E., J. Burger, C. B. Field, et al. 1999. 'Revisiting the commons: Local lessons, global challenges.' *Science* 284(5412): 278–82.
- Ostrom, E., J. Walker, and R. Gardner. 1992. 'Covenants with and without a sword: Self-governance is possible.' *American Political Science Review* 86(2): 404–17.
- Rasch, P. J., S. Tilmes, R. P. Turco, et al. 2008. 'An overview of geoengineering of climate using stratospheric sulphate aerosols.' *Philosophical Transactions. Series A, Mathematical, Physical, and Engineering Sciences* 366(1882): 4007–37.
- Roth, A. E., V. Prasnikar, M. Okunofujiwara, et al. 1991. 'Bargaining and market behaviour in Jerusalem, Ljubljana, Pittsburgh, and Tokyo: An experimental study.' *American Economic Review* 81(5): 1068–95.
- Sandler, R. L. 2007. *Character and Environment: A virtue-oriented approach to environmental ethics*. New York: Columbia University Press.
- Schelling, T. C. 1996. 'The economic diplomacy of geoengineering.' *Climatic Change* 33(3): 303–7.
- Schneider, S. H. 2008. 'Geoengineering: Could we or should we make it work.' *Philosophical Transactions of the Royal Society* 18(08).
- Searchinger, T., R. Heimlich, R. A. Houghton, et al. 2008. 'Use of US croplands for biofuels increases greenhouse gases through emissions from land-use change.' *Science* 319(5867): 1238–40.

- Solomon, R. C. 2005. 'What's character got to do with it?' *Philosophy and Phenomenological Research* 71(3): 648–55.
- Sunstein, C. R. 2007. 'On the divergent American reactions to terrorism and climate change.' *Columbia Law Review* 107(2): 503–58.
- Tilly, C. 2006. *Regimes and Repertoires*. Chicago: University of Chicago Press.
- Valencia Sáiz, A. 2005. 'Globalisation, cosmopolitanism and ecological citizenship.' *Environmental Politics* 14: 163–78.
- Victor, D. G. 2008. 'On the regulation of geoengineering.' *Oxford Review of Economic Policy* 24(2): 322–36.
- Warren, M. 1992. 'Democratic theory and self-transformation.' *American Political Science Review* 86(1): 8–23.
- Weitzman, M. L. 2009. 'On modeling and interpreting the economics of catastrophic climate change.' *Review of Economics and Statistics* 91: 1–19.