

Mill on Sanctions. *Utilitarianism*, chapter 2.

In chapter 3 Mill says, of any morality, one might ask, “What is its sanction? What are the motives to obey? or, more specifically, what is the source of its obligation?”

Is Mill asking an empirical question? This would be, perhaps, is it as a matter of fact possible to set up a stable utilitarian society, in which the members generally accept the doctrine and conform to it in their conduct? Or is Mill asking a normative question, is it rational for people to regard utilitarian principles as obligatory, as dictating what they morally ought to do? I suppose, the former, at least at the beginning of the chapter.

But by the end of the chapter, Mill seems to be answering a different question. This is, if you care only about your own self-interest, your own happiness (or maybe: you care only about the happiness of yourself and those near and dear to you), would you be better off if you trained yourself to become more impartial, care more about others, and at the limit, to care as much about any other person as about yourself?

In Mill’s usage, the sanction of a moral principle is the motivation one has to obey it. He distinguishes internal and external sanctions. The internal sanction is the voice of conscience (an inner voice that offers opinions when asked, what is morally right and wrong for me to do). This is an internal source of reward and punishment—if you do as your conscience instructs you, you feel good, and if you disobey conscience, you feel bad. At least, this is so, to the extent that you have a conscience.

As Mill observes (page 28), the ultimate internal sanction of any morality as a matter of fact is “a subjective feeling” in each individual’s own mind. What the subjective feeling points toward depends largely on how one has been raised and socialized. This is so whether one takes the internal feeling of duty/morality to be innate or implanted. Mill thinks the sense of duty is implanted, but even if it was innate, social conditioning largely fixes the content it supports.

External sanctions involve one’s desire not to incur the displeasure of others and the punishment they might inflict, together with the desire to incur the pleasure of others and the rewards they might confer. If others who interact with you are disposed to punish you if you disobey their idea of moral duties, and disposed to reward and favor you if you obey, you have an external sanction to obey their idea of morality. Also, if you have sympathy or affection for others, this will also be a sanction favoring your compliance with moral dictates, so far as morality dictates that you constrain your selfish urges and act in ways that benefit others. Also, if you believe in a God who punishes and rewards in the afterlife, this is a further external sanction.

Mill says, in any ongoing society, the external sanctions tend to be marshaled toward whatever the customary, socially dominant morality is. Moreover, people tend to be trained to accept the customary morality, so their conscience will tend to steer people to obey the customary dominant morality. (This is only true to a degree. There are

subgroups in society, with different moral codes, including criminal subgroups.) So there may seem to be a special problem for any proposed morality other than the one that happens to be customary in one's society. The customary society has the weight of sanctions supporting it, while the proposed reformer's substitute hangs unsupported.

Mill points out correctly that nothing in the nature of things prevents a society from adopting utilitarianism as its customary morality. If that happens, then the external and internal sanctions alike that operate on most people will tend to induce them to comply with utilitarianism. Mill thinks that in present-day society, the customary morality is not utilitarian, but rather a jumble of common-sense rules and norms, but these have to some extent been shaped by pressure of utilitarian considerations, so the customary rules to some considerable extent fit with utilitarianism even though they don't officially espouse it.

So Mill says, IF a society adopted utilitarianism, and successfully made it the customary morality, utilitarianism could be stable, could sustain itself as the going morality generally obeyed. Some questions: As a matter of fact, is it at all likely that a society would adopt utilitarianism as its customary moral code? If it made a project of doing this, would the project succeed? If a society was officially committed to utilitarianism, would its morality be stable or would the moral system of that society tend to erode and collapse?

So far as I can see, Mill does not address these questions, except for the line of thought in the final three paragraphs of the chapter.

What Mill says about conscience might seem cynical. Isn't the idea of conscience that people are taught, the idea that they should do what is moral according to their own lights? A conscientious person conforms her conduct to her own view about what is morally right. Being conscientious, she follows whatever moral view strikes her as best supported by reasons. To the extent that is so, the question whether the conscience of people could support utilitarianism becomes the question, could people find that after reflection they would see utilitarianism as the moral view that compared to alternatives is best supported by the balance of relevant reasons. Response: It's possible that a society could become committed to the idea of autonomous conscience—each should do whatever according to her own reasoning she decides is morally right or best. In actual fact, most societies aren't so committed—or they are committed to autonomous conscience only to a degree. For the most part, people are trained into the going customary morality and trained to accept it as best supported by reasons.

At the end of the chapter, Mill muses that if the sense of morality and the disposition to be moral are just artificially instilled in us by our early interaction with our social environment, the sense of morality might be susceptible to weakening to the point of disappearing once reflective adults understand that this is so. Why should I follow my conscience rather than just do what pleases me, if conscience is what Mill says it is and I understand this?

Mill's somewhat speculative reply is that we humans are naturally social and have a strong innate desire to be in harmony with those with whom we interact. As society progresses, this natural desire for social harmony gets attached to a wider and wider sense of community, a wider and wider idea of the "us" with whom each "I" wants to be in harmony. At the limit of progress, our natural desire for sociality becomes attached to entirely impartial utilitarian morality.

Mill: suppose each person for the most part is initially selfish, wants to maximize his own happiness. If our desires changed, so we came to want not just our own happiness but equally the happiness of all persons, we might in fact come to satisfy our selfish desire for our own happiness to a greater extent than would have been the case had we all remained selfish.

To indicate what Mill has in mind, consider the decision problem known as "prisoner's dilemma" (PD) after the story told to illustrate it.

The story is that two arrested prisoners are accused of serious crime, but the state's evidence is weak. The prosecuting state's attorney goes to each prisoner and makes the same proposal to each: If you squeal on your buddy and he does not squeal on you, you get off scot-free and he gets a long sentence, 15 years in prison. (On the other hand, if he squeals on you and you do not squeal on him, he gets off scot-free and you get 15 years in prison. If neither of you squeals on the other, you both get convicted of a minor offense, each of you gets 2 years in prison. If both of you give evidence against the other (squeal), you both get convicted of the larger offense and each of you gets five years in prison. Each prisoner knows each was made the same offer. Each know that both care only about his own sentence, and each wants to serve as little time as possible in prison. The next day, the prisoners are led to separate courtrooms and face this decision problem. The numbers represent years in prison. In the payoff matrix below the first prisoner's payoffs are the first number in the box, and the second prisoner's payoffs are the second number. What decision each makes affects the payoffs for both as follows: Each has the choice of cooperating (not squealing) or defecting (squealing, ratting on one's fellow prisoner).

		Second prisoner	
		Cooperates	Defects
First Prisoner	Cooperates	2, 2	15, 0
	Defects	0, 15	5, 5

Where

1. Each chooses independently of the decision of the other. One's own choice does not causally affect the probability that the other will make on or another choice.
2. Everything that each player cares about is registered in the payoff matrix, in his own payoff.
3. Each player knows the payoff structure and knows the other knows it as well.
4. Each player is assumed to be rational, to act rationally to maximize the satisfaction of one's own interest (to make one's own payoff as good as possible).
5. The decision problem occurs only once.

More schematically:

		Second player	
		Cooperates	Defects
First player	Cooperates	R, R	S, T
	Defects	T, S	P, P

Where

P is the Punishment payoff for mutual noncooperation. Neither cooperates.

R is the reward payoff for mutual cooperation. Both cooperate.

S is the Sucker's payoff. You cooperate and the other does not.

T is the Temptation payoff. You defect and the other cooperates.

And

$$T > R > P > S$$

And

$$R > (T \text{ plus } S) / \text{two}$$

In this decision problem, what is it rational for each to do?

Answer: One might think, I need to know how probable it is, that the other will make one or another choice. But this seems wrong. Whatever the other does, I am better off defecting. If the other cooperates, I do better defecting. If the other defects, I do better defecting. Therefore, I should defect.

Since the same reasoning applies to each, the predictable result is that fully rational individuals placed in this decision problem will choose in such a way that they end up with an outcome that is worse for each than another outcome that they together could have reached.

This may seem like an artificial puzzle, but many think the PD casts a long shadow on social life. Many human interactions are not single-play, but repeated over time, and in repeated PD, the analysis of what it is rational to do is delicate (what one does in one play may alter the behavior of others in future play, and thus one's future payoffs). Also, I have given a simple two-person illustration, but PD games can involve many persons (think of littering at the beach). But a lot of social life goes into preventing the occurrence of single-play PD and the predictable arrival at the noncooperative outcome. If the police will punish you for failure to cooperate, the game is no longer PD. If you care about what happens to the other people who are playing the game, that caring registers in your payoff matrix, so the game then may be transformed into something other than PD.

Back to Mill. Mill is abstractly right, as the PD game illustrates. If all of social life consisted of single play PD, and we all came to cease being selfish (caring only about our own jail sentence) and instead to care for all persons equally who might be affected by our actions or omissions, each of us might be better off, attain more happiness for herself, than would have happened if we had all instead been selfish, each caring only about his own happiness.

However, there are limits to Mill's optimistic speculation. For one thing, in a world with some groups very well off and many large groups very badly off, predictably, if better offs came to be impartially motivated, they would become worse off, not better off. Becoming motivated to cooperate with "us" (fellow members of our small community) and not with "them" (outsiders) might be the better strategy in self-interested terms.

For another thing, there may be limits to the malleability of human motivation. Maybe we are biologically evolved to care mainly about ourselves and a narrow circle of those near and dear to us, and socialization cannot greatly alter that biologically set psychological disposition. Maybe. If so, there are biological limits to the growth of human altruism and impartiality.

Third, suppose to some extent we can take actions that make a difference to our own character (and those of our children). Which dispositions would a self-interested person (who cares now only about his own long-run happiness) choose to instill in himself? What dispositions would a parent who wants her own child to do as well as she can for herself seek to instill in her child? The answer depends on the social environment you expect to face over the course of your life. With what sorts of persons are you likely to interact over the course of your life? This is a complicated issue. But you can easily see that whereas training yourself to be purely egoistic might not be your best egoistic choice (suppose others will shun you if they can and exclude you from mutually profitable cooperative schemes), in many social environments it would not be best in self-interested terms to dispose yourself to be purely impartial and to care for others just as you care for yourself. Suppose for example the world contains people who are exploitable and cheatable, and others who will cooperate with you only if you reciprocate. Then you should adopt a variable disposition, to exploit the exploitable and cheat the cheatable and cooperate with the conditional cooperators. Recall that we are considering what type of

character an egoist should choose, to the extent she can form her own character. “You should” here just means doing this would be the best egoistic strategy. We are supposing also, somewhat vaguely, that character is to some extent alterable at first, but eventually becomes pretty fixed and unalterable.

So Mill’s inspiring rhetorically powerful speculation and expression of hope for humanity at the end of chapter 3 may be a utopian fantasy. Or maybe not. The empirical questions here are tough ones.