



Rationality, Morality, and Collective Action

Author(s): Jon Elster

Reviewed work(s):

Source: *Ethics*, Vol. 96, No. 1 (Oct., 1985), pp. 136-155

Published by: [The University of Chicago Press](http://www.press.uchicago.edu)

Stable URL: <http://www.jstor.org/stable/2381329>

Accessed: 10/09/2012 14:24

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at
<http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



The University of Chicago Press is collaborating with JSTOR to digitize, preserve and extend access to *Ethics*.

<http://www.jstor.org>

Rationality, Morality, and Collective Action*

Jon Elster

Individuals, organizations, and governments often face choice situations with a peculiar structure, that of a collective action problem. To avoid awkward phrases, I shall refer to the decision makers as “individual” agents. Even when in fact they are corporate actors, they are individual in the sense of having one decision to make. The firm sets one price for its product—not as many prices as there are people employed in it.

I want to discuss what kind of explanations can account for the decisions that are made in collective action situations. In particular, I shall consider the importance of rationality and morality in the explanation of these choices. This means that I shall not directly confront the issue of what I believe to be the morally required course of action in such circumstances. Rather, I discuss to what extent people’s choices can be explained by what they believe to be moral. This does not mean that my own moral conceptions are irrelevant for or irretrievable from what I shall have to say.

Here are some examples of the kind of choices I have in mind. Should I cross the picket line or not? Should I leave the beer bottles on the beach or take them back home? Should I take that extra piece of cream cake? Should I vote in a presidential election? Should I fiddle with my income tax returns when there is no risk of detection? Should a union impose a closed-shop policy? Should an OPEC country stick to the cartel policy of restricting production? Should one generation allow the infrastructure handed over to it by the preceding one to run down?

Very roughly speaking, and I shall be speaking more carefully in a moment, these situations share two features. Suppose that many agents are facing similar choices. Then for all of them it is better if all choose one course of action than if all choose the other. Yet for the individual, the most attractive action seems to be the second course, that which leads to the collectively undesirable outcome if taken by all. This situation, of course, is known under the heading of the Prisoner’s Dilemma. I want

* This article was presented first at the Oberlin Colloquium, April 13–15, 1984, and then, in the present version, at the University of Chicago as the George Lurcy Lecture, 1983–84.

Ethics 96 (October 1985): 136–155

© 1985 by The University of Chicago. All rights reserved. 0014-1704/86/9601-0003\$01.00

to argue that collective action problems and their solutions are more complex than one would think given this stark statement of the Dilemma. Yet, in the end, what I shall offer is just another variation on a perennial theme.

The discussion will fall in two parts. First, I shall set the scene by considering a variety of collective action problems, ones with widely differing structures. Then I shall turn to the problem of explaining why individuals choose the one or the other course of action in these situations. More specifically, I shall consider the explanation of collective action—a term that I have not yet defined. By collective action, I mean the choice by all or most individuals of the course of action that, when chosen by all or most individuals, leads to the collectively best outcome. This course of action I shall often refer to as cooperative behavior. Note that I am addressing the question of why individuals decide to cooperate. Hence nothing but individual opportunities, beliefs, and motivations can enter into the explanation of their behavior. In particular, the benefits that cooperation brings to the group cannot by themselves explain it—unless they enter into or otherwise shape individual motivations or beliefs. We should ask not, Why is there cooperation? but, Why do people cooperate? These gestures in the direction of methodological individualism will have to do here.¹

In some of the examples I gave, the individuals were physical persons, but in other cases they were either something more or something less than that. A trade union or a nation-state is a corporate actor interacting with other corporate actors of the same kind. We usually think of trade unions as examples of successful collective action involving physical persons. A trade union is seen as a kind of institutionalized solidarity, preventing individual workers from crossing the picket line. Yet what from one perspective looks like cooperation may, in a wider context, appear as noncooperative behavior. If all unions insist on closed-shop policies, the outcome may be worse for all workers than it would be if no unions did so. This is the central argument in a recent book by Mancur Olson, whose classic work on collective action some twenty years ago set up the problem in the form in which it is still studied. He also applies the same argument to other interest groups, such as colluding groups of firms.²

These cases of nested or hierarchical collective action problems point to a general difficulty. What looks like noncooperative free-rider behavior

1. For arguments to back these assertions, see S. Popkin, *The Rational Peasant* (Berkeley and Los Angeles: University of California Press, 1979); and my *Making Sense of Marx* (Cambridge: Cambridge University Press, 1985).

2. M. Olson, *The Rise and Decline of Nations* (New Haven, Conn.: Yale University Press, 1982). Whereas in his classic work *The Logic of Collective Action* (Cambridge, Mass.: Harvard University Press, 1965) successful collective action was the fact to be explained, in his last book it provides the explanation for another *explanandum*, namely, the relative decline of stable, mature economies. For some doubts about this argument, see my review article "The Contradictions of Modern Societies," *Government and Opposition* 79 (1984): 304–11.

may, in fact, be the very opposite. The worker who crosses the picket line or the firm that refuses to follow the cartel may, of course, do so because they hope to get a free ride on the cooperation of others, but it could also be the case that they do so because they identify themselves with a wider set of interests. Similarly, there is a difference between passive and active abstention from voting, the latter being intended as a signal to the system that it fails to generate a satisfactory range of alternatives. When none of the presidential candidates takes a firm antiwar stand, some voters may think that civic duty requires them to abstain. If blank votes are formally recorded and counted, it is possible to distinguish between the two forms of abstention. Perhaps one could also invent devices to distinguish the conscientious objectors to trade unions from the free riders.³ In the absence of such means of recording active abstention, we always run the risk of misidentifying what we want to explain.

At the opposite end of the spectrum of cases is weakness of will, as in the cream cake example. Weakness of will is a collective action problem within the person, with the "individuals" being successive time slices or "selves."⁴ It is better for all selves if they all show moderation than it is if they all yield to temptation, but for any given self—that is, for the person at any given occasion—there seems to be no reason not to give in. To be sure, the asymmetry of time and the indivisibility of persons make for differences between this case and standard collective action problems, but I nevertheless believe that the assimilation can be quite illuminating.

Another temporal collective action problem is that faced by governments acting on behalf of a cohort or a generation. Why should we leave anything to our descendants? After all, they never did anything for us. Any kind of gross or net investment is a form of intergenerational cooperation.⁵ Frequently, these temporal externalities are reinforced by geographical ones. Government spending on basic research is an example. To spend the taxpayers' money on something that will mainly benefit other countries and later generations does not sound like a safe way to be reelected. In other cases, these externalities may annul each other. An example is the so-called imperialism of free trade.⁶ The concept was developed for nineteenth-century Britain but has also been applied to

3. The phrase "conscientious objector" immediately points to one possibility, namely, making it obligatory for unorganized workers to pay an amount equivalent to the union fee into a fund used for some socially desirable purpose. (This is incorporated in British legislation.)

4. For discussions of this problematic notion from different points of view, see my "Weakness of the Will and the Free Rider Problem," forthcoming in *Economics and Philosophy*, vol. 1 (1985).

5. A useful discussion is that of E. S. Phelps and R. A. Pollak, "On Second-best National Saving and Game-Equilibrium Growth," *Review of Economic Studies* 35 (1968): 185–99.

6. The phrase has been used with various meanings. The one I have in mind here is that of C. Kindleberger ("The Rise of Free Trade in Europe, 1820–75," *Journal of Economic History* 35 [1975]: 20–55).

the oil-producing countries today. It is sometimes argued that the best cartel policy for OPEC is competition since this will not provide the oil-consuming countries with an incentive to develop alternative energy forms that would drive OPEC out of business. Again, what constitutes cooperation is ambiguous.

Let me now define the problem more carefully. I am considering a set of individuals, each with a single decision to make. The case of repeated choice situations will concern me later. Most of the time I shall assume that the choice is between two courses of action—cooperating and not cooperating. Some real cases are like this, and even when the range of options is larger, so that there can be different degrees and modes of cooperation, the simplification may not matter too much. Also, most of the time I shall assume that the individuals are similar—both in their interest and in their resources. I shall, however, have something to say about the more general case in which individuals are allowed to differ and to have a larger range of options open to them.

Each individual makes one decision and is affected by the decisions of all. The ways in which he is affected define the structure of the interaction. In this preliminary discussion, I consider only one of the many ways in which he is affected—by his consumption of the goods and services produced by the collective action. Using a terminology to be explained later, I restrict myself now to selfish, outcome-oriented benefits. The collective action problem is defined in terms of such benefits. To explain collective action, it may be necessary to consider other benefits.

There is no obviously correct definition of collective action problems. I shall consider a weak and a strong definition, both of which will prove useful later. According to the strong definition, we have a collective action problem if two conditions are satisfied. First, each individual derives greater benefits under conditions of universal cooperation than he does under conditions of universal noncooperation. Second, each derives more benefits if he abstains from cooperation, regardless of what others do.⁷ This definition will do for most, but not all, cases. It will not do for voting and similar cases since here the individual will prefer to cooperate if he is pivotal. He will prefer abstention almost regardless of what others do, except in the special case where others vote in a pattern that makes him pivotal.

A weaker definition is provided by retaining the first condition that went into the strong definition while replacing the second by two others: cooperation is individually unstable and individually inaccessible. By individual instability I mean that each individual has an incentive to defect from a situation of universal cooperation, by individual inaccessibility

7. This is the definition used by T. C. Schelling, *Micromotives and Macrobehavior* (New York: W. W. Norton & Co., 1978), chap. 7, and underlying the "Schelling diagrams" (see n. 11 below). Schelling also imposes two additional conditions that need not concern us here. See also M. Taylor, *Anarchy and Cooperation* (Chichester: Wiley, 1976), pp. 43–44, 62, n. 7.

that he has no incentive to take the first step away from a situation of universal noncooperation. It may not be immediately clear how this weaker definition allows us to see voting as a collective action problem. I have in mind the situation that arises when the supporters of a given party know that they will beat the opposition by a solid margin if they all turn out, so that any individual can safely abstain—and if no one else turns out there is certainly no point in his doing so.

There are also interaction structures in which the cooperative behavior is either individually unstable or individually inaccessible but not both. Let me give two examples. Consider a lake around which there are adjoining peasant plots. These plots are threatened by erosion as a result of deforestation, which was originally undertaken to get more land for cultivation. On each plot, the erosion can be stopped if and only if trees are planted on it and on both adjoining plots. Cooperation then is individually inaccessible but stable. Conversely, assume that erosion will occur if and only if trees are cut down on the individual plot and on both the adjoining ones. Here, abstaining from cutting one's trees is individually accessible but unstable behavior.

Of these cases, the first is less serious than the standard collective action problem in that only information is needed to generate cooperation. I shall refer to this case as the information problem.⁸ The converse case is, if anything, more perverse than the standard problem since even with perfect information there is no clear guidance for action. If others cooperate, the individual should not, and if they do not, he should. Assuming them to be no less rational than he is, he cannot figure out what they will do and hence what he should do. I shall refer to this as the indeterminacy problem.⁹

The last piece of stage machinery that must be introduced relates to the technology of collective action, that is, the relation among individual participation, costs, and benefits. Assume that cooperation produces a continuously variable good or a continuously variable probability of the provision of a discontinuous good. The first case is illustrated by littering, the second by lobbying for a tax benefit. Again, voting falls in neither category; hence what follows does not apply to all forms of collective action. If the outcome of collective action is continuously variable, the amount in which it is provided depends on the number of cooperators.

8. It is usually referred to as an "Assurance Game," a term introduced by A. Sen ("Isolation, Assurance and the Social Rate of Discount," *Quarterly Journal of Economics* 80 [1967]: 112–24). The present example differs from many other illustrations of the Assurance Game in that it does not turn on externalities in the utility functions—only on production externalities. Altruism may turn a Prisoner's Dilemma (in terms of the selfish payoff structure) into an Assurance Game (in terms of "inclusive benefits"). See also n. 25 below.

9. The two-person instance of this game is usually referred to as the game of "Chicken." For discussions, see A. Rapoport, *Two-Person Game Theory* (Ann Arbor: University of Michigan Press, 1966), p. 137 ff.; and M. Taylor and H. Ward, "Chickens, Whales and Lumpy Goods: Alternative Models of Public Goods Provision," *Political Studies* 30 (1982): 350–70.

In the simplest case, each new cooperator adds the same amount as the previous one. Most frequently, less is added as more cooperate. Sometimes, however, an additional cooperator will produce a larger increment than his predecessor did; at least this may hold over some range of numbers. Collective action, in other words, may show constant, decreasing, and increasing marginal productivity.¹⁰

Similarly, the costs of cooperation may vary with the number of cooperators. An example of increasing cost is calling in to support public radio stations: as the lines become more congested, it takes more time to get through. Decreasing costs are illustrated by revolutionary collective action: when a given police force has to spread itself more thinly over a larger number of revolutionaries, the risk for each of the latter is reduced. The constant-cost case is illustrated by littering.

By combining these various cost and benefit structures, we can generate a large variety of cases. One important application is to the problem of defining the average benefit—the average being taken over cooperators and noncooperators—as a function of the number of cooperators. In the simple case of constant marginal productivity and constant marginal cost, the average benefit increases steadily with the number of cooperators. It is easy, however, to find examples of the average benefit first rising and then declining; or first declining (i.e., becoming negative) and then rising; or first declining, then rising, and then declining again.¹¹ This will turn out to have important implications for morality as a motivation for cooperation.

I now turn to the question of explaining collective action. I believe there is no more important problem in the social sciences, and none that is more difficult. Understanding why people cooperate and trust one another may be the first step toward bringing about more cooperation and trust. Actually, understanding collective action may also have the effect of rendering it less likely, as I shall explain in a moment, but by and large I believe the net effect of more insight will be positive. But if we are afraid that it might not, we can find comfort in the fact that not much insight looks likely for the time being. It is not that we do not know a great deal about the motivations that enter into collective action. Rather, our problem lies in understanding the way in which these motivations interact with others that lead the individual in the opposite direction. In particular, we do not seem to have any robust understanding of the relation between the two homunculi to be found within each of us—*homo economicus* and *homo sociologicus*.

10. For extensive discussion of these cases, see P. Oliver, G. Marwell, and R. Teixeira, "Interdependence, Group Heterogeneity, and the Production of Collective Action: A Theory of the Critical Mass, Part 1," forthcoming in *American Journal of Sociology*.

11. Fig. 1 (taken from Schelling, p. 220) illustrates these possibilities. The unbroken lines show the payoffs to each cooperator (R) and noncooperator (L) as functions of the number of cooperators. The dotted lines show the average benefit to all members—cooperators and noncooperators—as a function of the number of cooperators.

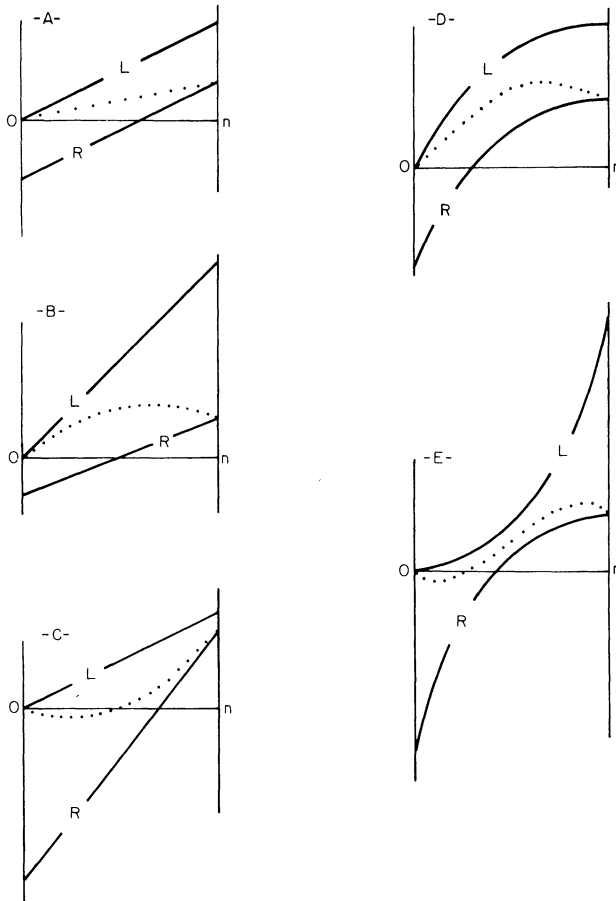


FIG. 1.—Taken from Thomas C. Schelling, "Hockey Helmets, Concealed Weapons, and Daylight Saving: A Study of Binary Choices with Externalities," *Journal of Conflict Resolution* 17, no. 3 (September 1973): 388. Copyright © 1973 by Sage Publications, Inc. Reprinted by permission of Sage Publications, Inc.

As a preliminary, let me introduce the notion of duty, in a broadly Kantian sense. The relevance of duty to collective action is captured in the phrase, But what if everyone did that? What if everyone left their beer bottles on the beach, stayed home on voting day, or fiddled with their tax returns? Duty enjoins us to do what we can rationally will that everyone should do. If causality is, in Hume's phrase, the cement of the physical universe, duty may be the cement of society. If people did not find the question, But what if everyone did that? a persuasive one, society would be in constant threat of disintegration. The puzzle is why the question is so persuasive. Acting according to duty can be pointless and even harmful, as we shall see in a moment. The sensible course would

appear to be to do one's duty only when it can be expected to promote the moral goal. Yet even from an instrumental point of view, noninstrumental morality may be indispensable. I return to this later.

I shall first discuss the view that people sometimes cooperate because they are irrational and then go on to consider various rational choice explanations. I shall not stop to define what I mean by rationality, except to state that I have in mind some very minimal conditions of intellectual consistency.¹² No one could possibly violate these conditions across the board and yet be recognizable as a human being with beliefs and goals. But there may be local breakdowns of rationality when we fail to act on what we believe or fail to form beliefs that are justified by the evidence available to us.

I shall consider only what I take to be the most important kind of irrationality behind collective action. It arises in two cases, which have been studied in some detail by psychiatrists and psychologists. George Ainslie has shown that self-control, as a form of intrapersonal collective action, can arise by what looks like a form of magical thinking.¹³ In this case, however, it is not clear that the behavior is irrational. George Quattrone and Amos Tversky have shown that the same reasoning arises in interpersonal cases of collective action in which it is demonstrably irrational.¹⁴ Part of the interest of this finding is that the irrational attitude seems to be closely related to the notion of a moral duty.

Consider first self-control as a solution to weakness of will. The latter can, for the present purposes, be described as a preference for smaller earlier rewards over greater, delayed rewards. When I consider this choice at some distance, that is, well in advance of the time at which the first reward will be available, I may tell myself that I shall choose the latter one. But when the choice approaches, I find myself unable to stick to my decision. If this is a choice that is to be made only once, there may not be much I can do about it. But if I know that I shall have to make many similar choices in the future, I can overcome my weakness by bunching them together—by making the first choice a predictor of the later ones. The reasoning might go as follows. If I accept a drink, a cigarette, or a cream cake now, there is no great harm, and I can always refuse on the next occasion. On the other hand, if I accept now, it is not plausible to think that I shall not also accept on later occasions essentially similar to the present one. Hence I should refuse now, so as not to knock over all the dominoes further down the road.

I leave open the question whether this form of thinking is a form of irrationality. It is certainly often very successful. Sometimes, indeed,

12. I attempt to state these conditions more fully in chap. 1 of my *Sour Grapes* (Cambridge: Cambridge University Press, 1983).

13. G. Ainslie, "Beyond Microeconomics," in *The Multiple Self*, ed. J. Elster (Cambridge: Cambridge University Press, 1985).

14. G. Quattrone and A. Tversky, "Self-Deception and the Voters' Illusion," in Elster, ed.

it can be too successful. The strategy of never allowing a single exception easily leads to a rigid, compulsive character—the person becomes a slave to the rules he has set up to curb his impulsiveness.¹⁵ It may well be the case that the average benefit to the person is maximized if he follows the rules most of but not all the time. But how is he to know when he can give himself a break? The interpersonal analogy is clear.¹⁶ To sustain democracy, a voter turnout of 100 percent is not really necessary, but how could one decide who is to vote and who is allowed to stay home?

Consider next interpersonal magic. It has been shown that—when placed before a hypothetical choice of whether or not to vote—people can be induced to act in a way that can be rationalized as follows. I am a fairly typical member of my political reference group. If I vote, it is pretty likely that others will vote as well. Being like me, they will tend to act like me. Hence I shall indeed vote, to bring it about that others vote as well. The last step in this reasoning is a slide from unexceptional diagnostic thinking to an unjustified causal argument. The slide is well-known from many other contexts, the best known being the psychology of Calvinism and the philosophical conundrum known as Newcomb's problem.¹⁷ A statement from an eighteenth-century Baptist leaflet is a perfect illustration: "Every soul that comes to Christ to be saved . . . is to be encouraged. . . . The coming soul need not fear that he is not elected, for none but such would be willing to come."¹⁸

In collective action, this form of irrational inference may be connected with the What if everyone did that? question. That question invites us to compare two states of the world: the one in which everyone votes and the one in which no one votes. If these two were really the only options, then it would have to be the case that by deciding to vote rather than to

15. It is part and parcel of Ainslie's argument that compulsive and impulsive behavior are causally related to one another—the former being an overly successful way of coping with the problem posed by the latter. Similarly, Donald Davidson (*Essays on Actions and Events* [Oxford: Oxford University Press, 1980], p. 30) argues that there is a conceptual link between these two modes of behavior in that they can both be subsumed under the heading of weakness of will. Although in the text I use weakness of will synonymously with impulsiveness since this is what we usually have in mind when referring to weak-willed persons, the more general notion is that of "acting against one's own better judgment," which also includes compulsive behavior of certain kinds. This also corresponds to the important Freudian insight that the ego (the will) is threatened by both the impulsive id and the rigid superego. I discuss these matters in the introduction to Elster, ed.

16. The analogy also has its limits, however. The damage to the person of a slavish adherence to rules goes much beyond that of forsaking specific pleasures. The lack of "toleration of ambiguity" and the inability to decide cases on their merits tend to become general character traits with debilitating effects on the general capacity for enjoyment.

17. For the original statement of this paradox, see R. Nozick, "Newcomb's Problem and Two Principles of Choice," in *Essays in Honor of Carl Hempel*, ed. N. Rescher (Dordrecht: D. Reidel Publishing Co., 1969), 440–72. For a demonstration of the relevance of the problem for collective action, see D. Lewis, "Newcomb's Problem Is a Prisoner's Dilemma," *Philosophy & Public Affairs* 8 (1979): 235–40.

18. Cited from E. P. Thompson, *The Making of the English Working Class* (Harmondsworth: Penguin Books, 1969), p. 38.

abstain I could also bring it about that all others make the same choice. Of course, there are many other possibilities, such as the state of the world in which I do not vote but some others do, but this is not allowed by the terms of the question. To ask, What if everyone did that? is not exactly like asking someone if he has stopped beating his wife, but it shares with the latter a certain insinuating quality. To the extent that the persuasive force of the question derives from an invalid inference, irrationality rather than duty may be the cement of society, a socially beneficial illusion, like Voltaire's God. It may not be a good thing if social scientists spend too much of their time discussing such connections among rationality, morality, and collective action.

Let me pass on to firmer ground and attempt to survey rational choice explanations of cooperative behavior. I shall classify the motivations that may enter into such explanations along two dimensions. First, we may distinguish between selfish and nonselfish benefits. Here I intend selfishness to be construed narrowly, so that the pleasure I derive from my neighbor's pleasure does not count as a selfish benefit. By selfishness, I have in mind direct, first-order benefits from my own consumption or my own activities. Second, I shall distinguish between outcome-oriented and process-oriented benefits. The former derive from the outcome of collective action, the latter from participation in the action itself. By crossing these distinctions, we get four types of motivation for participation in collective action. Of these, three form familiar categories, but one may be less recognizable. This is the notion that behavior may be motivated by nonselfish, process-oriented benefits. Although I feel extremely unsure of my terrain here, I shall propose that this is where the elusive category of social norms—as distinct from moral norms—may have a place.

I begin with the methodologically central class—rational, selfish, outcome-oriented behavior. My reasons for saying that it has a methodological primacy over other forms of behavior are the following. First, in the study of human beings there is a general presumption for rationality over irrationality, as I have already stated.¹⁹ Second, selfishness is logically prior to nonselfishness. The pleasures of altruism logically presuppose the pleasures of egosim, but not vice versa. There can be no pleasure of giving unless one believes that the recipient is pleased by the gift—and not pleased only by the act of giving. Many forms of moral reasoning, such as utilitarianism, also involve second-order motivations, which presuppose first-order benefits without being presupposed by them. Lastly, most process benefits are similarly parasitic on outcome benefits. If the process of participation is to provide an occasion for self-realization, it must be devoted to some outcome beyond the process itself.

The claims about selfishness and outcome-oriented motivation are, as I said, methodological, not substantive. They do not imply anything about what motivations are most frequently found or which benefits most

19. The general argument for this view is stated by Davidson.

valuable. Logically, it could be the case that only one person in society derived any pleasure from consumption, while all others got theirs from contemplating him and each other. Also the benefits from participation could be "more important" than the outcome benefit of collective action in the sense of providing more subjective satisfaction, but the latter is "more important" in the sense of being logically primary.

It follows from my definition of a collective action problem that rational, selfish, outcome-oriented actors will never choose to cooperate. Hence it would appear that we can pass on immediately to the other motivations. But there is an alternative: relax some of the highly restrictive assumptions. Some ways of doing this amount to changing the terms of the problem out of recognition, and I shall not discuss them here.²⁰ More central is the idea that it may be rational—in the sense under discussion—to cooperate when the individuals know that they shall face similar collective action problems in the future. Note that this proposal does not apply to intergenerational collective action problems. Hence for such problems, there is a strong presumption that selfish, outcome-oriented motivation will never lead to collective action. But when the individuals are to make their decisions at the same point in time and know that at later times they shall have to confront each other again, they may decide to cooperate in hope of reciprocation, fear of retaliation, or both.²¹

This will not always work. It is worthwhile spelling out some of the conditions needed to make cooperation appear attractive to individuals motivated solely by the benefits they personally derive from the outcome of cooperation. First they must be able to detect noncooperative behavior and to recognize the noncooperators when they meet them again since otherwise they would not know when retaliation or reciprocation is appropriate. Second, they must have some reason to think that others are in fact sensitive to these reactions. In particular, they must believe that

20. One line of argument is by assuming that one or more of the individuals have such a large interest in the outcome of the collective action that he will provide it single-handedly. Thus one shipowner might have such a large fleet that it pays him to construct a lighthouse even if he thereby makes it possible for others to be free riders. Note that, if there are several such individuals, the question of deciding which of them shall provide the public good becomes an indeterminacy problem. Another suggestion is that collective action is achieved through side payments, i.e., by selective benefits and selective punishments to cooperators and noncooperators, respectively. As observed by N. Froelich and J. A. Oppenheimer ("I Get By with a Little Help from My Friends," *World Politics* 23 [1970]: 104–20), this "solution" to the collective action problem presupposes that it is already solved. Instead of the original collective action problem (e.g., why should the individual not cross the picket line?) we get a new one (why should anyone beat up the persons who cross the picket line?).

21. Compare Taylor; and R. Axelrod, *The Evolution of Cooperation* (New York: Basic Books, 1984). These writers assume that the retaliatory strategy simply is the strategy of noncooperation. The chances of voluntary cooperation on the part of rational, selfish, outcome-oriented individuals may be enhanced if stronger forms of retaliation are available, as shown by L. Telser in an unpublished work on "Self-Enforced Agreements" (1984, typescript).

others are, like themselves, able to give some weight to future benefits. This is in fact a third, most important condition. If an individual attaches very great importance to immediate small rewards at the expense of larger delayed rewards, he may not be motivated by threats and promises concerning the future. Hence we see that the solution to the intrapersonal, intertemporal collective action problem may be a condition for the solution to the interpersonal problem. Rational, selfish, outcome-oriented persons will be able to cooperate only if each of them is able to impose the cooperative strategy on his successive selves. This also applies to international relations. Governments will abstain from protectionism only if their time horizon extends somewhat beyond the next election.

I want to consider, briefly, selfish, process-oriented motivations for cooperation. Could not participation in collective action be viewed as a benefit rather than as a cost for the individual? In particular, might it not be the case that participation in collective action offers a source of self-realization when the individual does not find it in his work? This way of turning the collective action problem on its head might appear quite liberating. An analogy could be the change in development economics that took place when it was realized that, at some stages of economic development, the alternative of consumption versus investment was a misleading one. Increasing consumption levels was also the best investment in the future via the impact on health and productivity. Similarly, as Albert Hirschman has put it, "The benefit of collective action for an individual [may not be] the difference between the hoped-for result and the effort furnished by him or her, but the *sum* of these two magnitudes."²² In both cases, we escape an unpleasant dilemma by turning the means into ends in themselves.

In the case being considered here, I believe there is little to be gained from this move. Even if we grant the assumption that participation in collective action is in itself a source of satisfaction, it does not follow that it is also a source of motivation. In my opinion, the benefits from participation belong in the class of states that are essentially by-products.²³ They arise as supervenient rewards from activities undertaken for some independently defined goal. One cannot get them if one engages in collective action solely to get them. This, of course, is not to deny that some individuals might believe this shortcut to be possible and join collective action on such grounds. To the extent that the efficacy of collective action depends on the sheer number of participants, such free riders on the movement—to use that term in a nonstandard sense—could actually

22. A. Hirschman, *Shifting Involvements* (Princeton, N.J.: Princeton University Press, 1982), p. 82. My objection to this idea can be stated in the same algebraic language. Denote the benefit of collective action for an individual by z , the expected value of the outcome of collective action by x , and the expected value of the process by y . To Hirschman's statement that $z = y + x$, I would add that y is itself a function $f(x)$ of x , constrained by $f(0) = 0$. At least this holds for rational actors.

23. I discuss these states in chap. 2 of my *Sour Grapes*.

help it. If, however, dedication and long-term planning are needed, those who are into collective action just to realize themselves will not contribute much to the success and may well detract from it. Successful collective action often requires the ability to wait—to delay action rather than to seize on any occasion to act. The self-defeating character of activism, or left-wing opportunism, is a well-known theme in the history of social movements.²⁴ Activism may appear to be a highly motivated form of cooperation, but in a temporal perspective it actually represents the noncooperative strategy.

Turning now to the class of nonselfish, outcome-oriented benefits, I believe these divide naturally into altruism and morality. Altruism, as I shall use the term, denotes a purely psychological inclination, not a moral attitude.²⁵ I may just find myself deriving pleasure from other people's pleasure, even when there is no obligation or indeed any occasion to act for the sake of promoting their pleasure. To be sure, for altruism to have any bite, it must imply that there are some occasions on which I would increase someone else's first-order pleasure rather than my own. But there is no need to assume that such behavior always derives from a feeling of moral obligation. If the members of a fan club make an effort, at their own expense, to surprise and please their idol, it is hard to believe that there is any moral feeling involved. In other cases, the dividing line between altruism and morality may be more difficult to draw. These, frequently, are also the cases in which the analytically useful distinction between moral norms and social norms is hard to apply. I want to postpone the discussion of these mixed and messy cases so as to keep the analysis unmuddled as long as possible.

Morality, for the present purposes, refers to any kind of impersonal evaluation.²⁶ A moral agent, on this neutral definition, could be motivated

24. Compare esp. M. Meisner's fascinating study *Li Ta-chao and the Origins of Chinese Communism* (Cambridge, Mass.: Harvard University Press, 1967). Meisner shows the influence of Bergson's philosophy, with its emphasis on the sacredness of the now, on Chinese activism and populism.

25. The importance of altruism in explaining collective action is discussed by Taylor, chap. 4; G. Marwell, "Altruism and the Collective Action Problem," in *Cooperation and Helping Behavior* (New York: Academic Press, 1982); and esp. H. Margolis, *Selfishness, Altruism and Rationality* (Cambridge: Cambridge University Press, 1982). Margolis's conception of altruism is interesting in that it also includes the concept of fairness. Whereas both Taylor and Marwell postulate that the individual in his utility function incorporates the welfare of others weighted by some constant factor, Margolis makes the more realistic assumption that the weights themselves are variable. More specifically, the more the individual has already spent on altruistic purposes, the higher relative weight he gives to his own welfare compared to that of others. I believe this captures better than any other formulation the subtle interplay of moral reasoning and psychological inclination that is observed in these cases. Unfortunately, Margolis's applications of this idea to specific problems are not very convincing.

26. Important discussions of the relation between morality and collective action are D. Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), pt. 1; and D. H. Regan, *Utilitarianism and Cooperation* (Oxford: Oxford University Press, 1980).

by duty or by the desire to maximize average or perhaps minimum welfare. He could be animated by the desire to ensure the protection of certain rights, by a wish to protect the environment, or by the goal of promoting the growth of knowledge. Although morality may require me to act in a way that favors particular persons, such as my immediate family or even myself, this can only occur when it is the best way of promoting the impersonal moral end. Note, however, the danger of self-indulgence that lurks here. If a utilitarian refused to cooperate on the grounds that the psychological cost to him of doing so would be very high, we might be inclined to accept the argument if he suffered from pathological shyness but not if he simply found the participation extremely boring.

Among the forms of moral motivation I shall consider only the utilitarian goal of promoting average welfare and the Kantian principle of action according to duty, that is, the Categorical Imperative. I shall argue that utilitarians have strong grounds for being critical of Kantianism—but also for appreciating the presence of a few Kantians among them. The latter conclusion is reinforced if we make the realistic assumption that in actual practice most utilitarians are more like altruists. They do not consider the welfare of everybody, nor do they weigh everybody's welfare equally.

First, however, consider the collective action problem in a community of pure utilitarians. If average welfare increases with each new cooperator, the moral choice is easy. If, however, the average utility curve has some decreasing parts, it is more complicated. In the simplest case, the decreasing stretch comes at the beginning, with the maximum average welfare being produced when all cooperate.²⁷ In the terminology introduced earlier, this is an information problem, not a collective action problem. Being an information problem, it requires information for its solution. The utilitarian agents must know that they are all utilitarian and that all know this. If the individual believes that no one else will cooperate, he has a moral obligation to abstain, too. In a moment I shall also consider somewhat better reasons for abstention in such cases.

A more complex case arises when the curve has a decreasing portion toward the end.²⁸ In that case, each has a moral obligation to cooperate if no one else can be expected to and an obligation to abstain if everyone else can be expected to cooperate. This means that we are dealing with an indeterminacy problem. Even with full information, it is not clear what the obligation of the agent is. A frequently proposed solution is this. Suppose that the number of cooperators that maximizes the average welfare is n percent of the total. Then each agent should decide by using a lottery device that leads him to cooperate with probability n percent. The law of large numbers will ensure that the desired outcome is produced

27. This is case *C* in fig. 1.

28. This is illustrated by cases *B* and *D* in fig. 1.

if everybody decides on this basis. Apart from the lack of realism, this solution is not even satisfactory in the ideal case. In the erosion example of indeterminacy, what is needed to produce the best outcome is not only that a certain proportion of the peasants abstain from cutting down their trees but also that the abstainers are spatially distributed in a certain way. This cannot be ensured by individual randomization. Of course, a collective randomization will give us what we need, but then the problem of collective action is assumed away rather than solved.

Finally, the average welfare curve may have decreasing slope at the beginning *and* at the end.²⁹ This would actually satisfy the weak definition of a collective action problem, although not the strong definition. Since the end point of the average welfare curve must be higher than the point where it begins, it must have some increasing stretches on which cooperation is morally preferred. Conceivably this stretch could be very short, to the point of making cooperation morally required only if the individual expects to be pivotal. Again, the abstractly optimal solution that would be chosen by fully informed moral agents is that of randomization, assuming that only the number of cooperators is the decisive factor.

The hard problems arise when there is imperfect information, as is almost always true in real cases. The utilitarian might not know whether others share his motivation or whether they know that it is shared. According to the usual canons of rationality, one should then try to form an estimate about the probable number of cooperators and choose to cooperate if that number is on an increasing part of the average utility curve. The estimate could be grounded in knowledge about the number of persons that have cooperated on earlier occasions or in a theory about why and when people cooperate. I am going to suggest, for two distinct reasons, that this line of argument is unsatisfactory and that Kantian motivation has a more central place.

First, when the utilitarian forms his estimate, he will be greatly helped by knowing that there are a good many Kantians in the population, that is, unconditional cooperators. If there were no people who could be counted on to cooperate in all circumstances, the probability estimates would have a very shaky foundation. Indeed, I would argue that there would not be any basis for the estimate. In that case, the rational utilitarian should act on the assumption that the worst will happen, that is, that his vote will indeed be needed. He should, in other words, behave like a Kantian. Or again, either the utilitarian is parasitic on the presence of Kantians or he has to behave like one himself.

The latter conclusion is reinforced by the problem of self-indulgence that I referred to earlier. Since the costs that might make it rational for the utilitarian not to cooperate, given his estimate about what others will do, derive only from his personal discomfort, he ought to regard them with some suspicion. One way in which the weakness of the moral mo-

29. This is case *E* in fig. 1.

tivation might show up is in exaggerating the costs of cooperation so as to make abstention morally acceptable. Knowing this, the moral agent should be prepared to discount the costs of cooperation. Again, for most practical purposes, this would turn him into a Kantian.

These are pragmatic arguments for acting like a Kantian, even if one's underlying motivation is a different one. Or they serve as arguments for welcoming the presence of some genuine Kantians—to make non-Kantian choices possible. On the other hand, there is one important consideration that points against Kantian behavior. This arises in cases where the average benefit curve has a declining stretch at the beginning because unilateral cooperation actually hurts the noncooperators as well as the cooperators. Two examples will have to suffice. If one country disarms unilaterally, this could create a power vacuum into which other states might be tempted to move, thus unleashing a general war. From the history of social movements, many cases are known in which individual acts of heroism or sacrifice give the authorities an excuse to crack down on the whole group—bystanders as well as activists. In such cases, unconditional cooperation is positively harmful, not just pointless or superfluous.

These opposing considerations may account for what I believe to be our ambiguous attitude toward Kantian behavior in collective action. On the one hand, society needs people who act morally without considering the consequences when misperception of the consequences is easy or likely. On the other hand, blindness for consequences may be disastrous. The best sometimes is the mortal enemy of the good. Logically, it might be possible to be an unconditional cooperator in situations in which the cost of unilateral cooperation is incurred only by the cooperator and to be more sensitive to consequences when the cost also falls on other people. Psychologically, I believe this is most difficult to accomplish. The motivating force of duty does not suffer too many exceptions. Be this as it may, I feel confident that the presence of Kantian motivation often enters importantly in the explanation of collective action. The more sophisticated utilitarian motivation may also play some role, but probably a rather small one. Whenever we find people acting instrumentally for the sake of other people's welfare, it will usually turn out that they are acting on altruistic rather than impersonal utilitarian grounds.

Altruism differs from morality in that other people are not considered impersonally. I do not care about everybody else's welfare, and I do not care equally about those for whom I do care. For an altruist, the benefit of collective action is not average utility but some weighted sum of his own and other people's welfare, with the weights assigned to the welfare of others typically being smaller.³⁰ I shall refer to this as the inclusive benefit of collective action to the person, as opposed to the exclusive or

30. This is compatible with the weights being themselves variable, as explained in n. 25 above.

selfish benefit. In the typical case, the population will have individuals with varying degrees of altruism, as well as some selfish agents, some moral agents, et cetera.

If the individuals have nonselfish motivations of different strength, collective action may arise by a snowball effect.³¹ Assume, namely, that each new cooperator brings successively larger increases in the average benefit. If some individual is highly altruistic or moral, the inclusive benefit to him of the first step may be sufficient to induce him to cooperate. Since the second step will create a larger increase in average benefit, less motivated individuals may join, and so on, until either all have joined or a point is reached at which nobody's inclusive benefit of the next step is positive. With full information, the process can even be anticipated and the choices made simultaneously rather than sequentially. Indeed, anticipation may lower the threshold for participation since an individual can now take account both of the immediate consequences of his action and of the consequences of the actions of those whom his action will motivate.

When the marginal productivity of cooperation is decreasing rather than increasing, the following problem arises. On the one hand, the highly altruistic individuals will probably be the first to join since their motivation is so high. On the other hand, they really ought to save their contribution until a part of the average-benefit curve is reached at which less motivated people cannot be expected to cooperate. Here simultaneous decisions made with full information may produce less cooperation than sequential decisions with foresighted altruists. In the first case, the less altruistic would know that the more altruistic will cooperate in any case, and hence they would not find their cooperation worthwhile. The point can also be stated in terms of activism. Activists can be useful in starting up a movement, provided there are others who are willing to take over when the marginal productivity falls below the level at which activists want to participate.

My final category is that of nonselfish, process-oriented motivation. Actually, what I shall be talking about are social norms, and I need not spend time discussing whether they fit into the box to which I have assigned them. While social norms are frequently very important in generating or sustaining collective action, I do not think their mode of operation is well understood. Part of the problem is the difficulty of predicting when norms will override other considerations. This is not simply a question of how much it takes to bribe people to violate their norms.³² Other, subtler factors also come into play.

31. This paragraph and the following one draw heavily on Oliver, Marwell, and Teixeira, with one modification. Whereas they assume that individuals differ in their selfish interest in the collective good (as a large shipowner has a greater interest in a lighthouse than a smaller one does), I assume that they differ in the strength of their altruism. Formally, this amounts to the same thing.

32. This is the approach in D. North, *Structure and Change in Economic History* (New York: W. W. Norton & Co., 1981), chap. 5, esp. p. 47. I am not denying that the concep-

Norms initially arise through the expectations of other people, together with their expression of approval and disapproval. The more deeply held norms form part of one's self-image. There are things one would not want to do because one simply is not the kind of person who does such things. Violation of social norms provokes feelings of embarrassment and shame, whereas violation of moral norms rather tends to produce guilt feelings. Also, moral norms are outcome oriented in a way that social norms are not. Although it is often hard to tell whether a norm is a moral or a social one, some cases are clear enough. Consider a man who mows his own lawn. He would not mow his neighbor's lawn for twenty dollars, yet he would pay his neighbor's son eight dollars (but not more) to mow his. This would appear irrational, given the economist's view that opportunity costs ought to be treated like any other cost. The explanation of the anomaly may indeed be that people treat out-of-pocket expenses and opportunity costs differently.³³ But I am more inclined to believe in a different explanation. The man would not mow his neighbor's lawn because he is not the kind of person who mows other people's lawns for money.³⁴ This is a purely social norm, with no moral connotations. Similarly, the person who refuses to take a drink out of fear that this will have a domino effect on his future choices is motivated by his self-image. Taking a drink now will make it impossible to think of himself as the kind of person who is able to resist temptation.

Social norms are Janus faced. They present themselves as absolute, yet they may be quite extraordinarily corruptible. On the face of it, they would seem to be constraints on action, but in their actual operation, they are more like preferences. Or if they are constraints, they can be made to crumble almost instantaneously by a suitable reframing of the choice situation. The man who would not mow his neighbor's lawn for money might be willing to do so if his neighbor gave twenty dollars to charity, thereby saving himself a similar sum that he had intended to give.³⁵ The situation is materially equivalent to the first one, but reframing leads to a different outcome. Similarly, one may be able to take a drink without harmful effects on one's self-image, if the situation is presented as a genuine exception.

The importance of social norms for collective action is hard to assess directly because in their operation they may be hard to distinguish from morality and altruism. A clear-cut, although hypothetical, example is the following. I intend to vote out of civic duty. If I abstain, I will enable

tualization of norms as psychological costs may be useful for some purposes, but I believe it fails to capture the essence of the phenomenon.

33. This is the explanation suggested by R. Thaler, "Towards a Positive Theory of Consumer Behavior," *Journal of Economic Behavior and Organization* 1 (1980): 39–60.

34. I owe this suggestion to Amos Tversky.

35. This would be consistent with the theory of framing, as developed in A. Tversky and D. Kahneman, "The Framing of Decisions and the Rationality of Choice," *Science* 211 (1981): 543–58.

two other people to vote instead. Yet I prefer to vote myself since to me the important thing is that I do my duty. Also, more centrally perhaps, there are cases in which people cooperate because they do not want to be seen not to do so. I make sure that nobody is looking my way before I drop my empty cigarette pack on the sidewalk. Note that this need not turn on my fear of sanctions and disapproval. Disapproval may be important in establishing norms but does not have to be present as a threat in each and every case of norm-governed behavior. I may abstain from littering even if the onlookers are passengers in a train passing by. Even if littering is compatible with my self-image, littering openly and brazenly may not be.

Indeed, I believe that, if people cooperate out of fear of social sanctions, we are no longer dealing with a collective action problem as I have defined it. Clearly, this is true if other people can withhold material benefits from noncooperators. This simply means that free riding ceases to be a temptation. And it would be arbitrary, I think, to conclude differently if the benefits withheld are other people's company and friendship. And it is not similarly arbitrary to draw a dividing line when what I lose by not cooperating is my own self-image. My self-image is not a benefit: it is what defines what counts as a benefit. This, however, is treacherous ground where it is best to shun any pretension to false precision. I do not really think we have an adequate conceptual scheme to deal with these matters. If we had, we might also be able to say something about when norms will stand fast against pressure and when they will break down. Just as some materials are immensely resistant to stress and yet crack by one well-directed blow, even the most fundamental norms may lose their hold on behavior almost overnight. In the case of norms of cooperation, this is facilitated because of the ambiguity of what constitutes cooperative behavior in any given case. Abstention from voting that is forbidden by the norms of duty may turn into an act of civic duty if redefined as active abstention in protest against the system. By switching allegiance from a smaller to a larger group, cooperation and noncooperation may take on new meanings. But to know that such phenomena occur is not to know when they do. Until we have a firmer understanding of such gestalt changes in normative behavior, the study of collective action may not be able to go much beyond "thick description."

Let me summarize some of the main points I have been trying to develop. First, the importance of mixed motivations in collective action should be clear. It is not only that different forms of collective action are held up by different motivations. A given case of collective action will also in most cases have participants who are motivated by quite different concerns. The presence of unconditional cooperators—acting for the sake of duty or for the pleasure of participation—may be a necessary condition for the emergence of conditional cooperators. These, in turn, may bring the level of participation up to the point at which new people join because they would be ashamed of being free riders. The permutations are almost endless.

Second, I have been concerned with bringing out the relation between the problem of collective action and weakness of the will. Indeed, weakness of will is a collective action problem within the person. Moreover, some of the strategies that individuals employ to achieve self-control are also used to induce cooperative behavior in interpersonal collective action problems. Also, in both cases, second-order collective action problems arise about who shall cooperate when there is no point in everybody doing so.³⁶ There is no reason why I should always refuse a drink, nor is it necessary for a viable democracy that everyone should vote. In addition to these conceptual parallels, there is an important causal connection between self-control and the interpersonal collective action problem in that myopia in the individual may block cooperation between individuals.

Third, the notion of what constitutes cooperative behavior in any given case is ambiguous. To ask for cooperation and collective action may, in a broader perspective, be a collectively self-defeating appeal to parochial interests. The pacifist who is asked, Who would fight the enemy if everybody acted like you? may justly reply, If everybody acted like me, there would be no enemy to fight.

36. For perceptive remarks on this problem, see P. Oliver, "Rewards and Punishments as Selective Incentives for Collective Action," *American Journal of Sociology* 85 (1980): 1356–75.