Reason and Action

Charles Fried

Follow this and additional works at: http://scholarship.law.nd.edu/nd_naturallaw_forum

Recommended Citation
http://scholarship.law.nd.edu/nd_naturallaw_forum/115
REASON AND ACTION*

Charles Fried

What I propose is the sketch of a complex theory. It is a way of looking at human actions, a way which I believe removes some perplexities and illuminates others which moral agents face in making decisions, concrete decisions: personal, social, legal. I say this is a sketch because it necessarily involves one deeply in matter of the sharpest philosophical controversy: action and will, reason and action, desire, intention, and so on. At every step there are many philosophical objections to be met, and a score of intricate arguments to be argued. Yet I believe if philosophy is to be relevant to practical men, then it must be possible to present at least the outline of the theory and carry conviction as to its relevance. That, at any rate, is the task I have set myself, and if I can succeed in it in some measure, then this is the best encouragement to the further task of battling out the details of the structure.

Although utilitarianism has been subject to powerful criticism from its inception, and none more powerful than that of Kant, it has entered the common educated consciousness as being at least prima facie the voice of reason applied to questions of morals, of value, and therefore finally of legal and legislative policy. It is, after all, a supremely generous, altruistic, and seemingly objective doctrine. A group of related examples will show this: in law, one of the few modern classics of legal scholarship and analysis, Wechsler and Michael's "A Rationale of the Law of Homicide," which set the way for the codification of a substantial part of the criminal law, explicitly adopts the utilitarian canon.

Applying this canon, the authors disapprove the decision of the Queens

* This paper is an address given at the eleventh annual meeting of the Board of Editors of the NATURAL LAW FORUM, September 30, 1966.

In developing the ideas in this paper I should like to acknowledge the instruction and assistance I have received from Professor John Rawls. My debt to him goes well beyond the very considerable amount I have learned from his published writings. I have profited also from conversations and from the opportunity of reading the notes of his lectures on Kant and Hegel, delivered at Harvard University in 1963.

1 37 COLUMBIA LAW REVIEW 701 (pt. 1), 1261 (pt. 2) (1937).

2 I refer to the American Law Institute's MODEL PENAL CODE, of which Professor Wechsler was the Chief Reporter.
Bench in the celebrated case of *Regina v. Dudley and Stevens,* in which two sailors were convicted of the murder of a weak and dying cabin boy whom they had killed and eaten after many days of drifting at sea in an open boat following a shipwreck. The court in effect affirmed the principle that the deliberate direct killing of an innocent person cannot be other than murder. Wechsler and Michael point out that on many occasions an act which, as a more or less remote consequence, has the effect of bringing about the death of an innocent person, is praised or blamed depending on whether a net saving of life was secured. The sealing off of a portion of a sinking ship, trapping some but saving many more, is such an example. But if the only distinction between such instances and *Dudley and Stevens* is that in the latter the actors directly cause and desire as their means the death of innocent persons, this seems to the authors a circumstance "without . . . import." In both cases it is the final balance sheet that counts.

Another example of the pervasiveness of the utilitarian approach can be found in the theologian Joseph Fletcher's very popular book, *Situation Ethics.* In this book he avows himself a thoroughgoing utilitarian, and states that it is simply incomprehensible to him how any other basis for an altruistic ethic based on love of neighbor can be rational, free from inexplicable arbitrariness, free from what Bentham called "ipse dixitism." For Fletcher the sole rational principle can only be the greatest good of the greatest number, and he is most emphatic in stating that therefore particular decisions can only ultimately be referred to this principle, this "optimific calculus" (he borrows that term from Bentham), and that any other principle for decision, any other conception of right, any other stopping place can only be arbitrary. Fletcher's argument is instructive because he is quite rigorous in drawing out the implications of this position, and he sees that if the single sovereign principle is the greatest happiness principle, then no subsidiary principle can be anything but provisional. It is in the name of this utilitarianism that Fletcher pours scorn on the "pilpulism" of the rabbis who worried whether a particular case fell under a specific prohibition rather than whether or not it was conducive to the greatest happiness of the greatest number. These rabbis entangled themselves in the most tortured complexities in accomplishing with — it seems to Fletcher quite unnecessary — indirectness, ends which could be achieved quite easily and directly.

---

3 14 Q.B.D. 273 (1884).
4 Herbert Wechsler & Jerome Michael, *op. cit. supra* note 1, at 1276.
5 *Joseph Fletcher, Situation Ethics* (1966).
6 *Id.* at 95-97.
7 E.g., *id.* at 31-37, 59, 131-33.
8 *Id.* at 19.
Given these views, one need not guess at what would be Fletcher's reaction to the series of cases recently discussed by David Daube in his book *Collaboration with Tyranny in Rabbinic Law*. The cases there considered deal with situations in which the leaders of a community are asked under threat of dire consequences by an enemy to hand over an individual who will be killed. The rabbis, or some of them, make a distinction between the case where a named individual is demanded and the case where the community itself is to choose the person to be handed over. In the latter case some of the rabbis would see an absolute prohibition against betraying one of their number into the hands of the enemy, and would rather see the whole community destroyed. The analogy might also be drawn to those occupied nations who refused to cooperate with the Nazi persecution of the Jews, though threatened with severe consequences and with little hope of preventing the final result. The relation of these cases to *Dudley and Stevens* is clear. If a net saving of life is said to justify the sacrifice of life by one means, for instance by sealing off the engine room of the ship, why should it not in another case, where moralists would see a direct killing? If it is the end that counts, why do the judges in *Dudley and Stevens* or the rabbis in Professor Daube's account worry about the means? As Fletcher puts it, "Only the ends justify the means, nothing else." And for utilitarianism the end is the greatest happiness of the greatest number.

It will be my thesis, first, that utilitarianism which has found such favor as a general moral theory is such an oversimplification of the human moral situation that its avowed clarifications are purchased not only at the cost of concrete conclusions which are sometimes so objectionable to our intuitive moral sense that we are not at all inclined to take them seriously, but also at the cost of a conceptual confusion about subjects such as means and end, and acts and consequences. Indeed I would say that utilitarianism has failed utterly to account satisfactorily for our intuitive sense of morality as it is expressed in absolute strictures against, for instance, murder, injustice, or cruelty. But the hold of utilitarianism is such that we maintain these intuitive principles, if we do, with a certain embarrassment, as if we were not being fully rational.

The oversimplification, amounting to fallacy, which I see in utilitarianism,
and which I believe entails these many unacceptable conclusions, I shall seek to explicate by reference to the concept of action. My argument will sketch a theory of action and of the relation of reason to action which seems to me more adequately to account for human moral phenomena.

I shall then proceed to suggest how this concept of action, and of reason as related to action, might be applied to the concrete problems of law and morality which I have already mentioned — Dudley and Stevens, or the handing over of innocent persons to an enemy — and provide at least a framework for analysis and judgment.

II

The version of utilitarianism which I wish to use — ultimately for no other purpose than as a foil for the affirmative doctrine I propose — is the full-blooded hedonistic utilitarianism of Bentham.

I will acknowledge at once that there are few in philosophical circles who maintain this full-blooded view, in part I suppose because of the very great difficulties and incoherencies in that view. Rather it has passed over in a more or less implicit form to others, lawyers, psychologists, political scientists, theologians, where the philosophical critique has not yet quite caught up with it. Bentham survives among philosophers largely in one of two forms. One form is what is known as restricted utilitarianism, according to which it is not individual acts but rules and practices which must be referred to the principle of utility. The few survivors of pure Benthamism reject this revisionism — quite rightly, it seems to me — as being arbitrary, and indeed it is urged only in order that utilitarianism yield results consonant with other principles, notably that of justice, which are not utilitarian at all. In other words the sovereign is — more or less surreptitiously — deposed at least as sovereign by these revisionists. The other form of the felicific calculus popular today is choice theory, which is a sophisticated mathematical tool for maximizing preferences, whatever they may be, in a given domain. With this theory I have no quarrel; it is simply beside the point, as it does not even — as utilitarianism does — purport to come to any substantive conclusions about preferences. Preferences are taken as given. But, of course, a complete substantive moral theory must address itself to the substance of these preferences. This utilitarianism does in its concepts of pleasure and pain.

It is the central tenet of hedonistic utilitarianism that every circumstance, event, or condition is to be judged by its tendency to produce a state of affairs

\[13\] E.g., Roy F. Harrod, Utilitarianism Revised, 45 Mind (n.s.) 137 (1936).

experienced by human beings which is happiness or, more robustly and concretely, pleasure.\textsuperscript{15} This central tenet may be analyzed to display two subsidiary and complementary propositions, both of the highest importance to my thesis:

First, that what is significant about a circumstance, event, or condition is its consequences, and thus as applied to human actions what is significant and solely significant about them is their consequences.\textsuperscript{16}

Second, that the sole consequence of ultimate significance, as Bentham puts it, the “sovereign” criterion, is an end state of pleasure, which is experienced by human persons.

It is against these two tenets — that acts dissolve into their consequences, and that a single passively experienced state is the end of all choice and action — that I shall argue and erect a competing theoretical structure.

III

The first proposition I wish to establish is that the single, determining end of many actions and, I would suggest most significant ones, is not some separate, passively experienced state, but rather that among the ends and often the sole end is the successful completion either of the action itself or of an ordered set of actions of which a particular action is a member.\textsuperscript{17}

The model of the Benthamic pleasure and pain is the itch and its relief. Of course, Bentham enumerates a whole host of other pleasures — for instance, the “pleasures of novelty,” or the pleasures of skill. But, insofar as they are all denominated pleasures, there is the clear implication that the connection between the end state and the action productive of it is such that what counts is the end state, while how it is produced is a matter of technology — what counts is the relief of the itch, and whether this occurs spontaneously, by my scratching, by taking a drug, or by someone else scratching my itch for me, is not relevant to that point.

Now it is evident that even in Bentham’s catalogue there are some items, like the pleasures of skill or curiosity,\textsuperscript{18} which do not fit this model easily. But in Bentham’s treatment the clear implication, and the one which accords


\textsuperscript{17} This point is made by Charles Taylor, The Explanation of Behaviour ch. 10 (1964). See also David Schwayder, The Stratification of Behaviour 159-64 (1965).

\textsuperscript{18} Cf. Bentham, op. cit. supra note 15; Harrod, op. cit. supra note 13, at 33-35.
best with the general direction of his work, is an implication made explicit in the writings of both Hume\textsuperscript{19} and Freud,\textsuperscript{20} that these pleasures are derivative from pleasures such as the pleasures of mastery, which in turn are compounded of pleasures of expectation of gratification to be attained and pains averted. I say this implication accords best with the direction of Bentham's work in that the very focus on the pleasures of this and that suggests a sensationalism as to which the cause of the pleasure is relevant simply as that, the cause of a separate and sovereignly significant entity.

Now I suggest that on the contrary in most significant actions at least an important if not the sole end is the action itself, and that to speak of the pleasure of the action as if it were a separate entity existing outside of the action is to mistake a metaphor for a reality. There are, to be sure, purely passive pleasures, and to them the Benthamic model is well applied. But it should be noted that these are not very important or very pervasive. In general the pleasurable pursuits are activities or actions, and not passively experienced sensations. The pleasures of food, drink, or sex, which are supposed to illustrate most cogently the hedonistic principle, pertain primarily to activities, the structure, appreciation, and context of which make for the pleasures they contain.

As one passes to less physical pleasures—the pleasures of amity, skill, and novelty, to knowledge, art, or love—it becomes more apparent still that the paradigm of the valued entity is not some passively experienced state but a complexly structured action or set of actions. Knowledge, as something enjoyed, for instance, quite clearly is either the active acquisition of knowledge, its appreciation, or at least the rehearsing in the mind of the evidences and interrelations of the knowledge.

Thus in general I would say that before pleasure, is action—in the sense that what we seek, what we desire, what we are satisfied in attaining is only rarely and unimportantly an end state separately identifiable as pleasure or a pleasure; it is rather the successful performance of certain actions or sets of actions. If, then, there be a sovereign master, to use Bentham's term, it is action. Moreover, action is the sovereign principle in another related way: it determines and gives sense to such concepts as emotion, will, desire, satisfaction, and indeed pleasure.

Desire is desire for an end to be attained; and if I am right about my criticism of the pleasure-pain model, then it follows that what is desired is the successful performance of various actions, be they acts of eating, of sex,

\textsuperscript{19} David Hume, A Treatise of Human Nature bk. 2, pt. 3, § 10 (1739).
\textsuperscript{20} For a good discussion see White, Ego and Reality in Psychoanalytic Theory (Psychological Issues Monograph 11, 1963).
of understanding, of friendship, or of worship. This desire may precede action, but it can only be recognized as a desire for this or that, and what the desire is cannot be identified without the concept of the object desired. Desire is a palpable disposition to action. So also emotion may be seen as the appreciation of a relation—between an external object and our interests, our interests being those things we desire to do. And pleasure itself, as a generic term, is to be understood as the successful performance of what it is we do. This was Aristotle’s definition.  

But these are complex issues requiring careful elaboration. For present purposes I wish only to suggest their relation to my principal assertion, that the central concept is that of action, that what we desire, need, take pleasure and satisfaction in is primarily action. This centrality of action should, indeed, be put still more forcefully: we are what we do. The same view which holds that our ends, the things we value, are passively experienced end states, also would see the unity and essence of the individual in this ultimately passive recipient of sensations, it would locate the center of our individuality in the flow of perceptions and sensations. It is my assertion that we are primarily what we do, and secondarily what we are inclined to do, wish to do, and are able to do.  

Two consequences of this action-oriented view are, first, the changed

21 Aristotle, Nicomachean Ethics bk. 7, 1154a. It should be evident that my account draws on Aristotle’s both directly and through his modern followers such as Anscombe and Hampshire. For recent discussions see Brice Noel Fleming, On Intention, 73 Philosophical Review 301 (1964); Anthony Kenny, Action, Emotion and Will chs. 1-6 (1963); Moreland Perkins, Emotion and Feeling, 75 Philosophical Review 139 (1966); David Schweyder, The Stratification of Behaviour 165-71 (1965); Charles Taylor, The Explanation of Behaviour chs. 2, 3, 10 (1964).  

In some recent discussions, particularly those of Kenny and Taylor, much is made of the argument that desires, emotions, pleasures, intentions, etc. are logically related to their objects, whereas those against whom this argument is made are said to posit a causal relation. This contrast seems so sharp that it is comforting to make points in terms of it. Thus the thesis of this paper might have been stated in terms of the distinction between ends which are contingently or causally related to the acts which attain them and those ends which are logically so related. Indeed this is a temptation to which I succumbed in an earlier draft, when it was pointed out to me that, first, this seemingly crisp opposition may be a good deal less clear than it seems, and, second, that anything I want to say can be said without recourse to what is at best a dubious distinction. Briefly, there does indeed seem to be a question whether the relation between an antecedent and a consequent which can be specified by reference to laws or theories is less logical or more contingent than a relation between elements which is stated to be logical or to depend on the definition of the terms, especially when the terms are as complex and open-ended as those in question here. For discussions which show sensitivity to this issue, see Richard Brandt and Jaegwon Kim, Wants as Explanations of Actions, 60 Journal of Philosophy 425 (1963); Donald Davidson, Actions, Reasons, and Causes, 60 Journal of Philosophy 685 (1963).  


23 Although I would not attribute this view to them, I refer the reader at this point to Stuart Hampshire, Thought and Action (1959); Brian O’Shaughnessy, Observation and the Will, 60 Journal of Philosophy 367 (1963).
emphasis this view gives to questions of evaluation, and second, the changed role that this view assigns to reason and rationality in relation to human action.

IV

It should be fairly evident how it is that the view I put forward complicates the problem of evaluation: if there is some single end state which is the touchstone of all value, then the question of evaluation is reduced to a question which has no further reference to values, a technical question as it were, a question of means only—how to maximize, either for one person or in society, the persistence and intensity of this end state. This, of course, is just how Bentham viewed the matter. To the extent that the ends of action, however, can be conceived of only as the successful completion of the actions themselves, then no single quantitative measure will be available. As there are many sorts of actions, so there will be many ends, and since these ends will not be commensurable by the single standard of the felicific calculus, how are these various ends to be evaluated against each other, how are they to be rationalized? This problem is difficult, indeed central, for it is the attraction of utilitarianism that it appears to have an answer which is at least in principle clear. Indeed the final picture I shall present, though I hope far more true to life, will be fragmentary, partial, and developing. I would attack this central problem by dealing first with the problem raised by an action-oriented view of the relation of reason to action.

V

The relation of reason to action in the classical utilitarian view is an easy one: as Hume put it, reason is and can only be the slave of the passions—and for “passions” one can easily substitute pleasure and those things inclining to it.24 Again, reason has a wholly technological significance—this is the meaning of the slave metaphor; one might as well say, after all, that business judgment is the slave of profit. We have a single value, and the maximization of that value is then an inquiry into means alone, not into ends.

Now in the action-oriented view I propose here the role of reason is quite a different one. There is no single end of all action, and therefore reason cannot have so simple a role. But what is that role to be? Here recur with me to my central concept of action—it is a notion of activity, of movement, mental or physical, internal or external, and action is or may be directed at its

24 For a most illuminating discussion and criticism of this dictum see J. D. Mabbott, Reason and Desire, 26 Philosophy 113 (1953).
own successful performance. Some clear examples are, on the intellectual level, the acquisition or exercise of knowledge; on the aesthetic, the composition, performance, or comprehension of a piece of music; on the physical level, the sexual act. More complex examples are acts of justice, acts of kindness, love, or friendship, acts of revenge or gratitude. In each case the end of the activity, of the changes, gestures, behavior which are included in the act, is at least the successful performance of the act itself.

Now what is the role of reason in this concept of action—is it that of a slave to an externally posited end? It might be, of course. Few if any of the actions referred to above can simply be performed at will—circumstances must be right, skills must be present, and reason, in Hume's sense, can certainly subserve to bring about such skills and circumstances. Moreover, many if not all of such acts may be performed for some ulterior reason and not for their own sake, and reason may be involved in the choice and guidance of the act as a means to an end. But reason, in my view, has a further and altogether different role, and although utilitarians might acknowledge this role, it plays no part in the development of their general moral theory. They do not build on it. It is an account of this role that I proceed to develop.

Take an activity like the intellectual activity of solving a problem: there are, to be sure, movements, gestures, behavior which constitute this activity. These may be internal or mental, but whether one speaks aloud or to oneself these gestures must be there. But surely no one would say that one performs the action of solving a problem just by going through these gestures. Jonathan Bennett in his excellent recent book Rationality tells the story of

the horse which was believed by some to be able to add and subtract numbers, because it "answered" questions . . . by tapping the ground the right number of times with its foot. It was then discovered on each occasion, just as the series of taps reached the right number, the horse's owner could be observed to hold his breath in honest anxiety. . . .²⁵

Now the horse in this story was not counting, although he performed all the gestures appropriate to counting, gestures without which—in some form, mental or overt—there could be no counting. The horse did not count, as Bennett so well points out, because the gestures were not motivated by reason, by an understanding of the principles, of the argument, if you will, by which these gestures lead to the solution of problems, by which they count as counting. So also one can neither play chess nor win at chess—an instance of an activity—merely by making correct chess moves, although, to be sure, one must make those gestures to play. One must also understand the rules

²⁵ JONATHAN F. BENNETT, RATIONALITY 40 (1964).
of chess, the point of the game and the point of games; and these understandings, these principles or arguments, must motivate the gestures, lie behind them, before one can be said to play chess. Consider an example from art. One can hear the notes in a piece of music, but one cannot be said to hear them as music unless he comprehends their relation to each other, their argument as it were. Finally, and most pertinently to the present purposes, imagine an act of justice: for instance the restoration of unjustly acquired property. One can take another's property and then — because one no longer needs it or through inadvertence, or perhaps as a gift — return it to the owner. But these are the gestures without the argument; the sense of justice involves an understanding of the arguments of justice, and to make an act of restoration is not only to make the appropriate gestures but also to make them in response to the principles of justice, as a result of following the argument of justice. It might be said that for an act of restoration to be such an act, it is as necessary to follow, however inchoately, the arguments of justice as it is necessary for an act of mathematical calculation to follow the canons of arithmetic, however inchoately.

Thus a rational action is an action which is performed not only according to a principle or program which an outside observer can articulate, but an action which is programmed or organized self-consciously in the sense that the agent is in possession at least inchoately of the argument according to which he organizes his behavior. To develop fully a theory of what it is that differentiates rational action from instinctive actions — such as walking or eating — would take us too far afield. I have used terms such as canons, principles, score, argument to refer to the element to which the agent self-consciously adverts in organizing his behavior into any particular instance of rational action. These terms are, I know, at best suggestive metaphors, and although I will not develop the full theory I would like to clarify the meaning of this element of rational action.

In rational action the agent's behavior is organized by reference to a plan which has as its content the organization of the behavior of an agent who refers to it; that is his reference, not that of an outside observer. Moreover, it is a further condition of the action's being a rational action that this plan exhibit a minimal degree of coherence. That is why I have found the term "argument" an apt one. Thus we will not have an instance of rational action if there is a plan that states that gesture a is to be followed by gesture b, and if furthermore the agent can be said to know of this plan and to have con-

27 Compare the discussion in Schwayder, op. cit. supra note 26, at 84-115 with that of Bennett, op. cit. supra note 25, at 86-93.
sulted it as he performed the specified gestures. The further element, which is that of coherence, requires that there be a specifiable relation between the parts of the plan, and that it be possible to articulate and defend that relation. Now that relation may be in the minimal case purely logical, so that the solution of a problem in logic will consist of a number of steps, each of which is related to the other by the canons of logic alone. Or in the playing of some very simple game, the individual moves may unfold by the application of the rules of the game to the situation, these rules operating as might the axioms of a proof. But the relation to the moves may be more complex, as in a dance where there are preexisting standards of grace, and where other dancers may interject different elements which dictate the appropriate response, given the previous gestures and the accepted canons. Or the relation may be more complex still, as in scientific inquiry, where the relations between the steps are the canons of valid scientific truth, and the movements—which may be purely mental movements—are dictated by the combination of those canons and the emerging evidence. Thus, what is needed for rational action is that the action be scored by an articulable plan, the steps of which exhibit a certain coherent relationship to each other.

I would enter two caveats at this point. First, neither the premises nor the laws of logic by which the plan of a rational action is a coherent plan need be openly articulated by the agent. They are immanent in the sequence of the plan. They come to the fore to the extent that incoherences, when they become apparent, bring about a revision of the plan of action in order to make it more coherent. And here too there need be no explicit reference to the canons of logic or coherence; it suffices that the agent understand and be able to make such a reference. Second, to the extent that premises for a coherent plan may be chosen at random, the variety of rational activities is infinite—this is the basis of play and of some forms of art. Moreover, instinctual activities and purely instrumental activities may be taken over by reason, and each be invested with a self-contained coherence—hence the art of dance, of love, of cooking. Thus we see reason seeking to involve more and more of our activities, a constant turning of the instinctual and the instrumental to that which has in part a self-contained and a self-conscious coherence.

I have sought to show thus far that there are actions which need be directed at no end other than their own successful performance. These acts almost invariably consist of an integrated, ordered series of gestures, and it is as much the maintenance of the order as the execution of the gestures

---

which is required for the successful performance. Of these there is a further subclass of acts which demand for their successful performance not only the execution of an ordered series of gestures but also a self-conscious reference to their own principle, the self-conscious reference to an argument. This I call the class of rational actions. Reason may discover and describe the principle of order in all such internally motivated actions, but only of the subclass of rational actions is it true that the agent himself must have some awareness—explicit or inchoate—of the principles of his act.

For this whole class, the analysis of means and ends, an analysis which utilitarians would impose quite universally, cannot offer the sole basis for analysis. The act itself is not simply and invariably a means to some separately identifiable and attainable end, and the constituting gestures are not related to the constituted act as means to end, but as a part to a whole. This may be easy to miss as such actions are of course extended in time and are constituted by a determinate commencement and termination, but to see the commencement and continuation as means and the termination as an end—which is what hedonistic utilitarianism and certain psychological theories derived from it suggest—is as absurd as to see the opening bars of a sonata as the means to the final chord as end, or to see winning as the end and playing the game as a means to that end.

But it is in the subclass of rational actions that we can see the special inappositeness of the universal application of the means-end analysis. Reason, calculation, planning, rationality in general, are not at all the means to some external end. Rather, rationality, as we have seen, is the constituting form of the action, the score which lends coherence to the performance. How far then is reason from being the slave of the passions! How far, to put it in our terms, is reason from being the slave of our dispositions to act! It is, indeed, the form of some of our most persistent and human actions, and thus of our most persistent if not urgent dispositions.

VI

This brings us to the problem of evaluation, which, as I have indicated, is considerably complicated. All I feel able to provide is something far less compelling than the techniques which the felicific calculus would provide. To those, indeed, who demand of a system of moral evaluation sharp and simple imperatives—an instinct which utilitarianism sustains—I can offer no help at all.

My argument depends on the proposition that we are what we do and what we can do. This primacy of activity in defining what we are I shall not argue further at this stage. I shall argue, first, that our capacity for
rational action, as it has been defined, is a necessary condition for the recognition of an agent as a human person; second, that there is a subclass of rational action, which I shall call moral action, which is a further necessary condition for this recognition. This assertion will be supported by my arguing, third, that material moral principles such as justice are applications of the capacity for moral action in certain specified circumstances. Fourth, there is a systematic connection between the principles of justice and such attitudes as respect and trust, and emotions such as love and friendship. Fifth, an agent who did not have these attitudes and did not feel these emotions would not be recognizable as a human agent. And sixth, whether one accepted such principles as those of justice could only be shown in rational, moral actions scored by those principles, so that evaluation becomes a matter of determining the principles on which one acts, and whether those principles are the principles of agents we would recognize as human. There are certain problems connected with this line of argument which I shall not consider: whether other animals are capable of rational or moral action, and the issue about which Professor Hillary Putnam has written so wittily, whether this capacity should be attributed to robots.29

Rational action has been defined as action which is scored by an articulable plan, the steps of which exhibit a certain coherent relationship to each other. This capacity for formulating coherent plans of action and for acting according to them is a necessary condition for an agent's being recognized as a human agent, since if an agent lacked this capacity he would be unable to identify his own action as actions: he could formulate neither to himself nor to others the rationale behind his discrete gestures and movements. He would have no sense of their unity as actions—which is not to say an outsider could not hypothesize such a unity as an external explanatory hypothesis about the behavior. So also such an agent would lack a sense of the rational unity inherent in other agents' actions: having no conception of his own actions he could have none of those of others. I think it is sufficient to assert that an agent which lacked any conception either of the coherence of his own behavior or of the coherence of the behavior of human agents would be so strange to us that we would decline to consider him human.30

It is a further necessary condition for the recognition of an agent as a human person that the agent must demonstrate a capacity for a special subclass of rational action which I shall call moral action. This is the capacity

30 Although I hesitate to attribute this argument to him, I was at least started thinking along these lines by Stanley Cavell, The Claim to Rationality ch. 8 (unpublished doctoral dissertation in Harvard College Library, June 1961).
to score one's actions not only by a principle of reason but by a principle of reason which takes as a premise or system of premises the perception of other human persons as having similar capacities, desires, ends, and needs to one's own. This premise is at the foundation of much of what we call morality.\textsuperscript{31} In the scheme of my argument I take moral action to be a further development of rational action, a more complete and pervasive expression of that capacity. It is a more complete expression of rationality because the perception of essential similarity or equality in entering into the argument or score of moral actions demands of that argument that the perceived equality of persons not be overridden by arbitrary inequalities.\textsuperscript{32} Now it might be objected that if the agent in plotting the fullest satisfaction of some complex desire at the expense of a fellow human being takes into account his fellow's similarity to himself and thereby, say, succeeds in trapping him where he might otherwise have failed, then here too the perception of the similarity of persons has entered into the argument of his action. But we would certainly not call his action moral. The answer to this objection is that the fact of similarity or equality has not, in the asserted counterexample, become a premise of the argument, since the agent, though recognizing the equality, still takes as his overriding premise a proposition which is inconsistent with this, that the satisfaction of his desires is to be accorded a different weight from those of his victim. Yet there is no premise which justifies such a preference. The closest would be the principle that each agent is justified in preferring his own desires. But the agent does not assert that in trapping his victim. It is only in principles which give equal value to all equal agents that reason is fully instantiated.

This perception of oneself as an agent equal to other agents gives rise to the constraints of morality, which are expressed by our recognition of the equal rights of others. This perception of equality, if it is to enter into the argument of our actions, can only do so in the form of the concept of equal right and the principles which derive from it. Roughly this is the concept that in acting we must only do so on a principle which leaves an equal scope to others to act.\textsuperscript{33} Otherwise we would all be doomed — whenever there is a danger that in acting we must impinge on others — to inaction or a devouring egocentricity. He who would interact with others and score that interaction in part by the perception of their common humanity must derive from that common humanity principles which allow the fullest equal scope for the action of all human agents in the interaction. As in the general case of

\textsuperscript{31} This account is, of course, essentially Kantian.
\textsuperscript{32} Compare R. M. Hare, \textit{Freedom and Reason} (1963).
rational action, there need be in moral action no explicit reference to the perception of equality, of common humanity, nor of the principle of equal rights which expresses that perception in action; it is only necessary that the sense and coherence of the action performed depend on an argument which assumes such common and equal humanity.

The capacity and disposition for moral action, that is, the capacity to score one's actions on the assumption of common equal humanity, I assert to be a necessary condition of recognizing an agent as human.\textsuperscript{34} That this is so can be seen by imagining an agent who lacks this capacity. Let us suppose he is capable of rational action, that he sees that other agents are capable of rational and of moral actions (as I have defined this term) and that they have instinctual ends similar to his own, but this moral monster (for that is what he would be) cannot understand at all how the fact of those capacities and ends in others imposes any constraints on his actions except perhaps as providing him with information so as better to exploit others. We would then imagine a rational agent for whom the humanity of others never provided a premise in the arguments of his actions — except as a basis for exploitation. Such an agent would have no conception of the rights of others, for — as I have argued — that conception derives from the adoption of common humanity as a premise for one's actions. And having no conception of the rights of others he could have no conception of his own rights. But an agent incapable of any conception of the rights of others or of his own rights would be so strange to us, and the forms of his interactions with us would be so alien and terrifying, that once again we could not recognize such an agent as human. Thus the capacity for moral action is a necessary condition for the recognition of humanity in an agent.\textsuperscript{35}

\section*{VII}

We have strayed a long way from utilitarianism, from specific moral judgments, from issues of means and ends and consequences in moral and legal judgments. What I have presented has been a sketch of a theory of action, of the subclass rational action, and sub-subclass of moral actions. I have also sought to indicate the relation between reason and action and how that relation yields a basis for evaluation in terms of the capacities for action which are necessary to the recognition of a person as a human being like ourselves. I now propose to return to the more concrete issues by applying

\textsuperscript{34} What follows draws heavily on Rawls, \textit{The Sense of Justice}, 72 \textit{Philosophical Review} 281 (1963), and Jean Piaget, \textit{The Moral Judgment of the Child} (Gabain trans., 1932).

this theoretical framework to problems raised by specific acts of justice, and particularly the problems raised at the beginning of this paper -- the direct killing of an innocent person, i.e., the Dudley and Stevens case; and the betrayal by a community under duress of innocent persons into the hands of an enemy, i.e., the case pondered by the ancient rabbis and raised again during the Nazi persecutions.

In each of the examples there are two choices open. One of these choices (refusing to render up the hostage person or hostage peoples or refusing to kill and eat one of your fellows) is likely to lead to a net loss of life, yet it is also the choice which seems the path of the higher moral virtue, which corresponds to our intuitive sense of justice in the face of the very greatest temptations. The utilitarian would quite probably urge the other alternative, the nonheroic one, which would effect a net saving of life. Indeed the utilitarian — with one qualification which I shall mention presently — would consider the course of conduct leading to a net saving of life the only rational and therefore the only moral course, and would have to condemn the one which our intuitive sense might incline us to praise, if not to demand.

In what follows I will try to draw out, by reference to the argument already made, a little bit better what it is that leads to our intuitive judgment that the heroic course is the better course in these situations; more specifically, that intuition would assign the killing or surrender of an unwilling and innocent victim or hostage to the category of injustice.

In relying on our intuitive sense of the right course in these moral dilemmas and on the common notion of justice I have not justified that intuitive sense or vindicated the claims of the common notion of justice against those of utilitarianism. The utilitarian, after all, offers a critique of common sense morality and of the overriding claims of justice. To meet this critique I must recur to the concept of moral action. Instead of measuring the choices by the felicific calculus I would ask which alternative in these dilemmas is an instance of moral action, which would be the choice of the agent who scored his action by the premise of the common humanity of the persons involved.

The answer in terms of the view I put forward is necessarily complicated, and I shall certainly not attempt a formal derivation here. Indeed in order even to sketch out how the concept of rational, moral action leads to a solution in accordance with the intuitive sense of justice it is necessary to give now a fuller account of the concept of moral action. Moral action is all rational action which takes as its premise the perception of common humanity. In speaking of rational action I spoke of the intrusiveness of reason, of the tendency of rational agents to score more and

36 See, e.g., John Stuart Mill, Utilitarianism ch. 5.
more of their actions by principles of reason, to lend coherence and self-containment to the instinctual and the instrumental. This same intrusiveness of reason is present in the area of moral action. For I would not have the premise of common humanity viewed simply as a negative, as a constraint imposed upon action. After all, were the principle simply a negative one it would be satisfied by inaction with respect to others, by passivity or suicide. But moral action is an instance of activity, of assertion, of self-definition through action. The correct emphasis must bring out the creative and the inventive in moral action, in that the forms and premises of moral action too can take over more or less of the instrumental and the instinctive, can invent new forms of actions scored according to the rational principles of moral action, and can invade, more or less completely, the amoral, that is, the instrumental and instinctive.

This intrusiveness and creativity of reason can be seen along the whole spectrum of sociability. Interaction with other persons is both a necessary instrument to the fulfillment of many of our most basic instinctual needs — food, shelter, defense — and a recognized part of our common animal instinctual apparatus, whether in the form of the herd instinct, or in the more exclusive forms of the maternal and the sexual instincts. In moral action, reason has invested these forms of interaction with a score or argument based upon the premise of common humanity. Thus in our forms of social organization we display acceptance of this premise by limiting even the common defense against hunger, the elements, and enemies to forms which are compatible with the premise of common humanity. Thus these pursuits, which might have been referred entirely to ends outside of themselves, become incorporated in actions which are scored by a principle of reason which recognizes not only the common needs of all but also the common humanity of all participants in the action. The concept of justice is primarily concerned with the forms which this recognition of common humanity takes, viewed as a limitation upon the pursuit of other ends. As such it is something of an abstraction, emphasizing a negative aspect of the matter. But if we keep in mind the total context this will not mislead us.

This principle of interaction in social enterprises is, furthermore, intimately connected with a family of other concepts: persons working together according to the premise of common humanity display a sense of justice or a sense of fairness in their dealings with one another. Further, as they display this sense of justice they will also display attitudes of respect and trust. And these attitudes of respect and trust are the indispensable conditions of feelings

37 See Kant, op. cit. supra note 33.
38 See Piaget, op. cit. supra note 34.
of love and friendship. On the basis of these feelings, which incorporate the rational principle of moral action and the instinctive needs for closeness to fellow beings, more elaborate and self-contained social forms still evolve, in which the very sharing of activity is itself an end, this sharing being always on the basis of mutual respect. Most instances of higher civilization are examples of this: drama, music, the fellowship of scholars in intellectual pursuits. But in fact this aspect of moral activity reaches back into the more instrumental activities, so that not only is there justice but also friendship in such complex forms as the army and the factory.

What is true of action in the sphere of complex, organized social activity is true also in the context of more or less personal interactions. Friendship does not necessarily require a full social context — it may take place on a desert island or on a lifeboat in midocean. But it does require the concepts of trust and of respect, and so the acceptance of the premise of common humanity in the activities which the concept of friendship scores. So also reason and morality can take over the instinctual sexual relation, incorporating it into larger morally defined contexts of action and investing it with forms, feelings, and entailments expressive of the arguments of love, friendship, and respect, which in turn derive from the premise of common humanity.39

Thus morality is essentially a type of rational action, and it is an action which is performed in a variety of contexts in which it gives rise to the concepts of justice, fairness, respect, trust, friendship, and love.

Applying this to our examples, it should be clearer what is the rational basis of the intuitive judgments that the killing of the innocent cabin boy or the turning over of an innocent hostage to an enemy are acts violative of justice and therefore violative of morality. For morality is exemplified and exhibited in actions which have the perception of common humanity as their premise. The concrete cases, the lifeboat as well as the threatened community, present complex contexts in which the situation and its possibilities for action are defined by the forms which morality assumes in those contexts. Not only do these cases occur in contexts where rational and moral acts of friendship and love are possible, but where justice too requires the pursuit of our interests only insofar as the constraints of the premise of common humanity permit.40 In each case the question will not be merely what act will effect a net saving of life, but what act in that context is an expression of the respect for common humanity, as love, friendship, mutual respect, and loyalty are expressions. And if the dilemma is put in those terms, the argument for the

39 See Max Scheler, The Nature of Sympathy ch. 7 (Heath trans., 1954).
heroic course as the better course becomes far more plausible. For defending against the common enemy without turning over the hostages, or supporting each other in the ordeal of the lifeboat without killing an innocent and unwilling fellow seems at least prima facie the best expression of justice, and therefore of friendship, love and respect and, more generally, of morality.

If this account makes the heroic course also at least plausible as the better course, the question might still be asked whether I have given a reason for following it, in the sense of having shown that there is an obligation to follow it. To this question the only answer I can venture is that if we would be moral men then we must act according to the principles of morality; and if we would show a sense of justice, love, friendship, trust and respect, then we must score our actions by morality in the appropriate contexts. And the sanction is at least this: if we failed utterly to express these feelings and attitudes, to act according to these principles — in short, if we failed utterly to be moral — we would not be recognizable as men and could not expect to be treated as such.

VIII

I have suggested — and no more than suggested — how the view I put forward of action as the touchstone of morality might provide a solution to concrete moral problems which is consistent with our intuitive moral sense and opposed to the solution which utilitarianism would demand. I would like to consider an objection which might be raised on behalf of utilitarianism which would deny that my general view leads to different conclusions in these specific cases.

Might it not be said after all that the perception of the importance of an organized community as the scene of human activity of the highest order of rationality justifies the sacrifice of the single individual, that the victim himself should approve the sacrifice? And further, for the shipwreck case, might it not be said that since life is necessary to sustain rational activity or any other kind of activity, the utilitarian response to that case too is the correct response?

This very plausible argument commits a subtle fallacy. For the community to persist, its members must, of course, survive, and the conditions of their cooperation must continue to obtain; but the community, as an instance of rational activity, is present in that activity and not simply in the persons and circumstances that make it possible. Thus, in my example, the act of justice would be the most intense possible expression of the principle of community. In a real sense, then, the community would be more severely jeopardized by yielding to the urgings of the enemy and committing an act
of injustice. And this is but a further instance of the general point that what one is — whether as a person or a collectivity — depends on what one does.

The fallacy is instructive because it points out a further contrast to utilitarianism, in which survival and therefore the factors of life and death take on a precise but intuitively bizarre significance. Since value consists for the utilitarian in the maximization of the single end state, pleasure, individual survival is relevant only insofar as the individual is a possible locus for the experience of that end state, and his survival is justified at any particular moment only insofar as the discounted expectation of pleasure the individual will both cause and experience exceeds the pain. Hence Wechsler and Michael's confident solution to the problem of Regina v. Dudley and Stevens set out at the beginning of this paper. But we all suspect that this calculus is both incoherent and wrong. Why we reject it may be clearer now from the example I have given. What we survive as is more important than whether we survive in the simple sense of physical persistence; and thus too, since we must all die, it has always seemed important to those who have lived to the fullest extent of their human capacities that somehow this feature of life be significantly incorporated into a rationally principled action or series of actions. To recur to the musical metaphor, the fact that the music must stop is in any finished composition drawn into the very argument and structure of the piece that is played — it is not just an arbitrary external circumstance to be postponed or hastened, depending on how things are going.

Lest it be thought that I am referring to occasional heroics or even melodramatics, I should say that the point I am making here, the contrast I am drawing to utilitarianism, seems to me to be the basis for the moral and legal judgment that killing to save life is murder or not depending not simply on the final balance sheet of lives saved and lives lost, but also on how the act was done, how the lives were saved. In summary form, if lives are saved unjustly then this may be murder, no matter what the score. This was the judgment of the court in Dudley and Stevens, a judgment which Wechsler and Michael reject in the name of utilitarianism. Moreover, if there is no injustice, then let the final balance sheet be never so lethal, there is no act of murder.

In the light of these remarks perhaps we can see the sense behind the seemingly hairsplitting distinctions the rabbinical doctors suggest between the case where an enemy state demands an individual and where a robber band makes the demand, for in the former case, as Professor Daube points out, it is possible that the named individual has committed some crime for which

---

41 See G. E. Moore, Principia Ethica 156-57 (1922).
the enemy state is entitled to punish him,\footnote{Daube, \textit{op. cit. supra} note 9, at 21.} and where there is no larger community of states which determines how fugitives from due punishment are to be captured a bellicose threat is not such an outrage. The same argument relates to the demand for a named individual as opposed to the demand for an unnamed hostage who is to be chosen by the threatened community. For even where it is quite clear that the victim is to be rendered up for arbitrary murder there is a deeper perversion of the machinery of the state (which is the exemplification of the bonds of community of its members) in the case where not only does the community push the victim out of doors, as it were, abandoning him to a fate which it did not create, but participates in the killing in the active way of actually using its processes to choose a victim. The distinction, indeed, is closely related to the distinction between bringing about a result by action and bringing it about by a failure to act. This too is a distinction which utilitarians reject,\footnote{Bentham, \textit{op. cit. supra} note 15, at 70-72; Wechsler and Michael, \textit{supra} note 1, at 724-35, 751 n. 175.} but which our intuitive moral sense tells us is of the highest importance. Where the community chooses the victim its involvement in the crime is more active, the crime becomes in part its action. If we feel the distinction between the two cases is inadequate, may I suggest that this is not because, as the utilitarians might argue, there is the same loss of life in both cases, but rather because, though the community's participation in the crime is less active in the former case where a named person is involved, it still seems active enough.\footnote{Professor Robert Nozick has suggested a further distinction: that between the case where the city casts one out of doors, though he be a stranger to the community, and that where it simply refuses to grant asylum. Again the distinction is between action and inaction, and it is echoed in the constitutional doctrine which allows Congress and the Executive the greatest latitude in deciding whom to admit to the country or whom to admit to citizenship, while imposing much more stringent limitations on deportation and denaturalization.} It is interesting to note, incidentally, that even the most rigorous doctors would not have objected to an attempt to persuade the victim to give himself up.

In the case of \textit{Dudley and Stevens} too the question is not simply the one of numerical survival but of the quality of the act by which survival is purchased. What are the premises of the argument, what are the principles of an act which has as its direct and immediate object the death of an unwilling victim with the end in view that others may live? (I leave out of account the gruesome way in which the death enabled the others to survive.) Certainly they are not compatible with the arguments of mutual respect and friendship, since the killers say to their victim by their act that his unwillingness to die at this time and at their hands must be subordinated to their wish to live, that his life was worth less than theirs, that his right to determine how he shall
live and for what he shall die must be subordinated to their determination
to live longer, and for the reason alone that they are part of a majority who
would survive. Now I will not go into the formal demonstration of why a
principle such as this is incompatible with a morality based on mutual respect
for the moral personality of all persons, and thus why it cannot accord with
justice. I will rest content with the assertion that respect, friendship, trust,
loyalty or love could not be exhibited in the actions of persons who adopted
such a principle for their actions. And what would be the conclusion if the
defendants had allowed their victim to die naturally, or if they had all agreed
to some procedure where lots would be thrown, or if the victim had volun-
teeered for the sacrifice? If one feels, as I do, that the issue at least becomes
harder in these cases, I suggest that this is because, though the net saving of
life is the same, the quality and principle of the act may be quite different.

Before leaving these examples I should deal with a utilitarian objection
which might be called the subjectivization defense. The utilitarian might
object that if I seem to make my case against him by the use of examples
involving the breach of principles such as mutual trust and respect it is be-
because I take too simple a view of his position, for the utilitarian is willing to
take everything into account, even the effect on the agent and on the whole
future of the human race of outraging sentiments of loyalty and mutual re-
spect. If these sentiments are socially useful, and if in effecting a net saving
of life they are to some extent undermined, why that too must be taken into
account and may well tip the scales. But this seemingly plausible response is
no answer at all, for these sentiments of loyalty, respect, love, and trust are not
simply sentiments but, as I have been arguing, principles of action which are
quite incompatible with the utilitarian point of view. Of course the utilitarian
is free to treat these principles simply as data about the people whom he would
manipulate, but in so doing he must deny their validity and so undermine
them at least as his principles. He cannot, in short, be a person who re-
spects, trusts, and is loyal to his fellow men unless he views these principles
as principles, and then he cannot just add them into his utilitarian calculus
simply as factors.

IX

The last point I would make is a caveat. Utilitarianism is a comprehensive
philosophy. In principle it holds all the answers, and whatever difficulties
there may be are difficulties about the facts and their consequences. If such

44 This move seems to be made by Chopra, The Consequences of Human Actions, 65
PROCEEDINGS OF THE ARISTOTELIAN SOCIETY 147 (1965). See also Paul A. Freund,
Rationality in Judicial Decisions, in NOMOS VII—RATIONAL DECISION 109, 114-15 (Carl J.
a view is stiflingly rationalistic, how much more so must my proposal seem, given the more intimate role assigned to reason in it. But this is a mistake. I have spoken of reason organizing, providing the principle for actions. Also I have spoken of cases where the external facts are such that some action is almost inevitable, and any action involves a crucial choice among principles. But I do not think that this is the usual condition of our lives. Often inaction will involve no particular principle of reason, and furthermore the situations in which the most crucial principles become relevant are not all situations into which we have deliberately thrust ourselves; we may simply have drifted into them or have them thrust upon us. Thus the system I have in mind is indeed a good bit more fragmentary than that which utilitarianism offers. It is a picture of life in which circumstances or our own doing may throw up certain crucial occasions for action when what we are is severely tested; it is also a picture where we invent numerous occasions for action, scoring the actions according to principles that may be more or less coherent with each other from occasion to occasion. And finally it is a picture of a life where we invest, showing more or less ingenuity and creativity, the necessary, the instrumental, and the instinctive actions of our life with the forms and elaborations of reason. It is a picture made up of work, play, art, much idleness and tedium, and a few moral crises. It is in short a rather fragmentary and discontinuous picture, presenting none of the smooth relentlessness of the constantly maximizing utilitarian man's life.

To be sure, some of us tend to organize larger and larger chunks of our lives into coherent wholes. Some of us tend to make into integrated actions, ordered by a single internal principle, large segments of life and conduct, which for others remain discrete and fragmented. And there is a tendency in reason to organize more and more, to take the fragmented and incoherent and make it part of ever larger wholes. But this is a matter of more or less, and I have not yet met the man who has succeeded in incorporating into one all-inclusive, coherent action the various elements, the various discrete actions of his whole life.45

---

45 Compare the discussion in J. D. Mabbott, Reason and Desire, 28 PHILOSOPHY 113 (1953).