https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

UTILITARIAN AGGREGATION

By Russell Hardin

I. Introduction

Utilitarianism is about consequences and their evaluation. It is, therefore, seemingly a natural moral theory for policy decisions, yet it faces notorious difficulties in aggregating from the individual to the collective level. What we generally choose are strategies, not outcomes, and some of the difficulties in aggregating derive from this fact. There are at least three major issues to be resolved. First is the problem of our communicating to each other just what we feel or know about our state of affairs. Second is the metaphysics of comparison of two different people. Third is the set of complications introduced by the indeterminacy of outcomes from our choices of strategies. The first of these issues can be partially resolved in a novel way that David Hume grasped but could not explain, a way that was largely overlooked by thinkers after Hume until recently. We can explain it to some extent and therefore can get beyond the issue. The second of these issues continues to be intractable and in principle may not be resolvable. I will argue that the third issue is not a problem, in the sense that the result or outcome of our interactive choices is often genuinely indeterminate. Hence, it is not a failure of theory that it is indeterminate. I will address these issues in two contexts. First, for conceptual reasons, I will address individual-to-individual comparisons (Section II). Second, to see the import of these issues, I will focus on policies that prosperous nations might adopt on behalf of far less prosperous nations. The specific focus of discussion on actual issues rather than abstract concerns should help us to keep clear concerning what we can and cannot conclude.

In contemporary debates, the framing of issues of justice between rich and poor nations has the quality of persuasive definition. Advocates of poorer nations focus on national responsibilities for per-capita rates of energy consumption. Others—not always avowedly advocates of richer nations, although often they are—focus on the responsibility of nations for their rates of population growth. Population is treated more nearly as a domestic problem by poor nations and as an international problem by wealthy nations. The difference clearly has distributional consequences and is therefore crucial to theories of distributive justice if these are to apply internationally.

This difference reflects, perhaps in reverse, an old distinction in utilitarianism from its days of cardinal, interpersonally additive utility theory.

doi:10.1017/S0265052509090025

© 2008 Social Philosophy & Policy Foundation. Printed in the USA.

Such a utility theory requires that we be able to make such claims as the following: A policy adds, say, 4 utiles to your utility and 3 to mine, for a total increase of 7 utiles. Some utilitarians argued that population increases that involved the addition of people who were at least marginally happy were utilitarian improvements because they increased the total of welfare in the world. Others argued that average utility was the correct focus of utilitarianism. On the second view, a world that is only Iceland is superior to a world that is only Bangladesh, even if the sum of all utility in Bangladesh is massively greater than the sum in tiny Iceland.[1] This raises the metaphysical problem of just what aggregate welfare means and why it would be of any value if we could give it meaning. I will not advance any claims that could resolve this debate, in part because I am inclined to agree with Vilfredo Pareto that such aggregates are metaphysical—indeed, that they border on nonsense.[2] But one can also step around this debate on the ground that it is abstract and seldom relevant to any actual policy choices we might make. For example, if we seek to reform population policy in some nation, the reason will be that *its own population affects that nation's citizens,* not that it affects some abstract consideration of average versus total welfare. A nation could have a fertility policy designed to reduce the rate of population growth, as in many African and Asian nations. But this will be about improving the lives of actual people, not about the average versus total utility debate. It is hard to find a legislative or other policy debate that would take this abstract issue of average versus total welfare into consideration in the drafting of any legislation on population issues.

Moral theory should be made relevant to real policy debates; it should not spin into the stratosphere of hollow thought on matters that are deeply irrelevant to our lives. Even the late Daniel Patrick Moynihan, a Harvard sociologist who had serious interests in moral issues in policymaking and who became a U.S. senator, would find this a silly enterprise and would avoid bringing it up for fear of being dismissed as an irrelevant player in the game of policymaking—as, in fact, he probably already was in many ways, including his failure to sponsor legislation that got adopted into law. He did not, however, generally sponsor silly legislation.

This traditional debate over total versus average welfare is relatively hollow because its background assumptions are very nearly meaningless. As Vilfredo Pareto argues, there can be no relevant cardinal assessment of the welfares of individuals that would allow cardinal interpersonal com-

[1] At this writing, Iceland's population is a bit more than three hundred thousand, and its GDP per capita is about $40,300. Bangladesh's population is about 150 million, and its GDP per capita is about $2,300.

[2] Vilfredo Pareto, *Manual of Political Economy* (1927), trans. A. S. Schwier (New York: A. M. Kelley, 1971), 47–51.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

parisons to be made.[3] What can we say if we suppose that the only comparisons to be made are not additive but only ordinal? It seems plausible to claim that the typical person in stressful poverty is worse off than the typical person in contemporary Iceland. Yet there might be many people in Bangladesh whose lives are quite comparable to the lives of the three hundred thousand people living in Iceland. Indeed, there might be far more than three hundred thousand such people, so that, if the welfare of others in Bangladesh is not negative, we might conclude that aggregate welfare there is greater than in Iceland on a total-welfare view. Would that conclusion matter? For a utilitarian, that is to ask: Does it have consequences?

To answer such questions, I wish to work out some of the implications of an ordinal utilitarian analysis of international distributional issues, especially population policy.[4] The central questions are: Can we make ordinal claims for personal actions at the individual level (Section III)? And can we make ordinal claims for policy issues at the collective level (Section IV)? It is perhaps clear that these two questions would admit of seemingly contrary answers. I first address the general problem of utilitarian comparisons between individuals (Section II) and between aggregates (Section IV), then the nature of ordinal transfers (Section VII), and finally the complications of population growth in impoverished nations (Section VIII). The first question might or might not yield insights into this issue of population growth, but it is the policy-level question of aggregation that we eventually wish to answer. That question is clouded by problems of indeterminacy (Section V) and the role of institutions (Section VI).

II. Individual Comparison: Mirroring

If we are to aggregate in any way, we must first be able to assess what others think and value. This raises the hoary debate about other minds. Can I know what you actually think or feel about anything? Do I know that you feel pain or pleasure? If so, how? Hume offers an account of how I might know your feelings and thoughts, an account that was merely descriptive in his time but that has now been experimentally tested, although the contemporary experimenters seem not to know of Hume's interest in the issue and his description of it. For Hume, merely acknowledging that it happens is enough for him to reach conclusions about our

[3] Ibid. See also Russell Hardin, "Efficiency," in Robert E. Goodin and Philip Pettit, eds., *Companion to Contemporary Political Philosophy* (Oxford: Basil Blackwell, 1993), 462–70, at pp. 464–67.

[4] On ordinal utilitarianism, especially at the domestic level, see Russell Hardin, *Morality within the Limits of Reason* (Chicago: University of Chicago Press, 1988), chaps. 3–5.

moral psychology, which is Hume's naturalistic, explanatory concern.[5] Let us consider his account.

Hume quotes Horace as saying: "As men's faces smile with those who smile, so they weep with those who weep."[6] Hume further notes: "[W]e may remark, that the minds of men are mirrors to one another, not only because they reflect each other's emotions, but also because those rays of passions, sentiments and opinions may be often reverberated, and may decay away by insensible degrees."[7] What in twentieth-century philosophy was the problem of other minds—how can we know another's mind?—is assumed away (in limited part) by Hume. To answer this question, there may now be neurophysiological evidence from fMRI (functional magnetic resonance imaging) studies. Sympathy, these studies suggest, is a form of direct, nonverbal communication and the evocation of relevant feelings.

Hume does not explain the phenomenon that he observes and (of more importance) that he experiences.[8] Perhaps we can now explain it.[9] Or, rather, perhaps we can move the explanation one step further back from the bald observation. Recent fMRI studies—studies of the brain's reactions to others' sensations—corroborate Horace's and Hume's observed facts and, in a sense, seem to show the phenomenon of *seeing* another's emotions at work. The fMRI studies do not do much more than Hume already did: they simply establish that mirroring happens, although they are more definitive than Hume's singular testimony. The part of the brain that perceives a smile is evidently the part that engineers a smile of our own, so that Horace's observation may be a biologically hardwired fact of our brains. Smiles evoke smiles. The evolution of this feature of our brains might be explained by the benefits of smiling in gaining the good graces of others, especially when we are too young to survive on our own.[10] Smiling may enable humans to enjoy very long periods of infancy, childhood, and adolescence, so that

[5] See Russell Hardin, *David Hume: Moral and Political Theorist* (Oxford: Oxford University Press, 2007), chap. 1.

[6] Horace, *Ars Poetica,* lines 133–34.

[7] David Hume, *A Treatise of Human Nature* (1739–40), ed. David Fate Norton and Mary J. Norton (Oxford: Oxford University Press, 2000), 2.2.5.21. Hereafter cited as Hume, *Treatise,* followed by book, part, section, and paragraph numbers; see also *Treatise,* 3.3.1.7; and David Hume, *An Enquiry Concerning the Principles of Morals* (1751), ed. Tom L. Beauchamp (Oxford: Oxford University Press, 1998), 5.18. Hereafter cited as Hume, *Enquiry,* followed by section and paragraph numbers. See also Terence Penelhum, "Hume's Moral Psychology," in David Fate Norton, ed., *The Cambridge Companion to Hume* (Cambridge: Cambridge University Press, 1993), 117–47, at p. 143.

[8] There was no need for him to explain; he could observe the phenomenon and could start from there (see Hume, *Enquiry,* 5.17n19).

[9] The following discussion draws on Hardin, *David Hume,* 41–45.

[10] There are recent studies that suggest other connections. Those who yawn when another yawns seem to score higher on empathy tests than those who do not mirror the yawns of others. Henry Fountain, "Tarzan, Cheetah, and the Contagious Yawn," *New York Times,* August 24, 2004, F1.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

we can develop extraordinary abilities that set us apart from other animals. If so, then mirroring might be among the most important of all genetic capacities: It has set humans apart from all other creatures. It might therefore be among the most important twists in human evolution.

It is psychological mirroring that leads me to like or dislike something that is done to you, by letting me sense what you enjoy or suffer. Contemporary neurophysiological findings seem to strengthen Hume's claims for the moral psychology of mirroring, although the mechanism is not yet clear. Those readers who have had difficulty accepting this part of Hume's argument or even its implicit resolution of the problem of other minds might soon find it easy to accept. Hume appears to be right on the psychology here. The only question that might remain for some is that of his general claims for what John Rawls calls Hume's account of morality psychologized.[11] Do we have moral reactions (reactions of approbation or disapprobation) to the feelings we get from mirroring? Those would be moral reactions on behalf of another. That is to say, the important and very difficult trick Hume needs to complete his explanatory theory is to evoke *my* sentiments—that is, a moral judgment—in response to actions that affect *your* interests.

From the fMRI data, it appears possible that these two phenomena—sympathy and moral sentiments—are at least partially run together in our brains.[12] Hence, Hume's theory is complete, but in a way that he apparently did not see. The knowledge and the feeling, the sympathy and the sentiments, may come in a single package. There is no mediating interpretation that our brains have to make. A nearly brand new baby smiles back at our smile. It is implausible to suppose that the baby is interpreting our kindness or good will in its first days of life; it is reacting from an apparently hardwired capacity. Empathy seems to "mirror" another person's emotional responses in one's own brain.[13] Happily, "mirror" is Hume's word and also the terminology of contemporary neurophysiological science.[14] One might suppose that mirroring is a two-step process. Our perceptions of (say) a smile stimulate thoughts, which guide our behavioral response: smiling back. Studies of brain activity, as measured by fMRI brain scans, suggest that the whole reaction is immediate in a single step, not mediated by thought. The part of my brain that recognizes a smile also forces or stimulates my own smile and my own feeling of

[11] John Rawls, *Lectures on the History of Moral Philosophy* (Cambridge, MA: Harvard University Press, 2000), 21; see also Hardin, *David Hume,* 32–33.

[12] See also Páll Árdal, *Passion and Value in Hume's Treatise* (Edinburgh: Edinburgh University Press, 1966), 47n.

[13] The German psychologist Theodore Lipps coined the German term for empathy in 1903, and he described the phenomenon of mirroring. See Bruce Bower, "Repeat after Me: Imitation Is the Sincerest Form of Perception," *Science News* 163 (May 24, 2003): 330–32.

[14] Hume, *Treatise,* 2.2.5.21.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

pleasure.[15] Seeing your smile triggers mine. Again, *we are to a degree hardwired to each other.* This might sound like an astonishing claim, except that we can give an evolutionary explanation of the development of such a capacity, even though we may not yet have a neurological account of how it works.

As a youth in his twenties, Hume seems to have grasped the nature of this phenomenon to a sufficient degree as to make it the foundation of his moral psychology. He does not attempt an explanation of the phenomenon, but merely starts from it to explain morality as a matter of fellow feeling. In fact, of course, he had no way to prove his assertion of the nature of this psychological trick other than to elicit our agreement that we too have the experience he describes. The technology of fMRI now seems to give us some entrée to the phenomenon.

In the fMRI studies, sympathy appears to be a form of direct, nonverbal communication and the evocation of relevant feelings. In another context, Hume dismissively says of the possibility of an innate sense of rules of property, "We may as well expect to discover, in the body, new senses, which had before escaped the observation of all mankind."[16] In actual fact, he may well have discovered, along with some others, including Horace and recent psychologists, what we might come to call a sense: the sense of sympathy. It is a sense that may have been much more acute in Hume than in most people but that is clearly evident in large numbers of people, including newborn babies, and apparently in other species as well.[17] It may be as hardwired as the sense of taste or smell.

Hume says that our sympathy for those on a ship sinking offshore will be greatly heightened if they are close enough for us to see their faces and their frightened responses. He does not explain this fully but only says that contiguity makes their suffering clearer to us.[18] The fMRI studies suggest that the issue is not that we have to see their expressions in order to understand their emotions; our reason is adequate for such understanding. The issue is that *we have to see their expressions in order to trigger the mirroring of our own similar emotions.* This is a phenomenon that is not mediated by thought or reason, and perhaps it cannot be replaced by thought or reason when the actual visions are not available.

Suppose we accept this entire account of our moral sentiments and of their apparent mirroring. If they are merely a fact of our psychology, should they determine our morality? Yes, in Hume's functional way. That is, our sentiments about others evoke responses from us that are responses

[15] See various contributions to A. N. Meltzhoff and W. Prinz, eds., *The Imitative Mind* (Cambridge: Cambridge University Press, 2002). See also Greg Miller, "Reflecting on Another's Mind," *Science* (May 13, 2005): 945–47.

[16] Hume, *Enquiry,* 3.40.

[17] Chimpanzees and Macaque monkeys, even in infancy at three days, apparently mirror emotions of others. See Bruce Bower, "Copycat Monkeys: Macaque Babies Ape Adults' Facial Feats," *Science News* 170 (September 9, 2006): 163.

[18] Hume, *Treatise,* 3.3.2.5. See also ibid., 2.1.11.6 and 8.

to the utility, pleasure, or pain of those others. What typically brings pleasure to others is their own benefit, which is *good for them.* We cannot go further to say it is good per se unless we go so far as to say that something like utilitarianism is the right moral theory. Hume does not make this claim, but in his analysis of the motivating force of mirrored reactions, he does imply that he and we are psychologically utilitarian. One of the things we can tell about other people through mirroring is how something affects their welfare, pleasure, or pain. This fact is important if we are psychologically utilitarian—and mirroring virtually makes us be, as though evolution has produced utilitarianism as our moral response. Psychological utilitarianism connects observation to judgment. These facts do not make utilitarianism the true moral theory; they merely characterize our psychology as moralized through mirroring. This psychology gives us a science of moral beliefs and approbations; it cannot additionally justify those approbations or make them right.

Mirroring is a major discovery for Hume despite the fact that seemingly all people experience it, so that it might well have been a matter of widespread common knowledge. It apparently remains unconscious and inarticulable to most people even while it often regulates their emotions and behavior. Hume is sufficiently perceptive that, once he has noticed the phenomenon, he finds mirroring to be a fundamental part of the psychology of sympathy and, therefore, a fundamental part of distinctively moral psychology. Mirroring makes Hume's theory psychologically richer than any of the then-contemporary moral sense and sentiments theories, which were inherently psychological in their foundations. Their proponents were generally content to stop their inquiries at the point of asserting that we just do know right from wrong, that reason can determine these, or that god has given us such knowledge. Hume empirically observed—and supposed we could all observe—the now seemingly joint phenomenon of mirroring and sympathy.

III. Ordinal Values

Throughout the nineteenth century, economic and utilitarian value theories were essentially the same. Indeed, the major utilitarians wrote on economics and, in some cases, contributed to the field. John Stuart Mill wrote a major, standard textbook that was issued in several editions.[19] And the great mathematical economist F. Y. Edgeworth wrote on mathematical utilitarianism.[20] This connection, which strengthened both economics and moral theory, was continued through the turn of the twentieth

[19] John Stuart Mill, *Principles of Political Economy* (1848), in J. M. Robson, ed., *Collected Works of John Stuart Mill,* 7th ed., vols. 2 and 3 (Toronto: University of Toronto Press, 1965).

[20] F. Y. Edgeworth, *Mathematical Psychics: An Essay on the Application of Mathematics to the Moral Sciences* (London: C. Kegan Paul, 1881).

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

century by Henry Sidgwick, who also wrote a major textbook on economics but whose main work on utilitarianism was among the most important books in its tradition.[21] The connection was broken under the influence of G. E. Moore, who took ethics into the realm of the hyperabstract,[22] and whose baleful influence cut off further developments in the application of economic value theory in ethics. The result, through most of the twentieth century, was the grievous weakening of utilitarianism as its value theory was left far behind by developments in economics that were almost entirely neglected by moral philosophers.

Many of the difficulties in utilitarian value theory were implicitly addressed by economists, but without much impact on developments in ethics, which was left to stagnate in a dead value theory. One result was to make utilitarianism seem too abstract and irrelevant for serious analysis of policy issues or of personal morality. A kind of gamesmanship in inventing supposedly clever twists and seeming paradoxes in value theory took over in the absence of serious concerns. In particular, Moore's value theory carried us back to a vision of value as inherent in the objects that we value—the objective theory of value. If no one wants X, that makes no difference to our assessment of the value of X. This general vision is related to the more narrowly focused objective labor theory of value, in which the value of X is determined by how much effort it took to produce it.

This is the Salieri theory of value. In the possibly false tale of a conflict between Mozart and Antonio Salieri, the latter worked very hard on writing operas, and supposed therefore that they must be highly valuable, even if everyone would prefer to listen to the operas that Mozart dashed off seemingly without effort. In an era in which the emperor decided whose work would be performed, Salieri was superficially very successful. Sadly for Salieri, however, he reputedly knew better, and he knew in particular that Mozart was a better composer than he was. In the movie *Amadeus* (1984), based on Peter Shaffer's play, Salieri calls himself the patron saint of mediocrity. In the world of utilitarian value theory, economists were under the sway of Mozart and philosophers were followers of Salieri. No wonder utilitarianism lost its formerly appreciative audience.

IV. Aggregation of Welfare

Beyond mirroring to communicate individual sentiments, etc., there is a further trick that might still be difficult, and which is the core of our concern here: evoking an emotional response of a similar kind to *the*

[21] Henry Sidgwick, *The Methods of Ethics*, 7th ed. (London: Macmillan, 1907); Sidgwick, *The Principles of Political Economy* (London: Macmillan, 1883).

[22] G. E. Moore, *Principia Ethica* (Cambridge: Cambridge University Press, 1903).

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

interests of society. An individual's identification with the interests of society must be very weak psychologically, and it is grounded in reason more than in sympathy. The neurophysiological fMRI studies probably cannot address such an abstract phenomenon as responding psychologically to the interests of society. Seeing the interests of society forwarded or abused is physically not comparable to the visual cue of a smile or a frown. There seems likely to be no individual-level analog of seeing society and its pains and pleasures. We might note that we mirror similar feelings in every member of some social group, and we might be told that every one of us does likewise. In this way, we might sum from individual sympathy to societal sympathy. But to do even this for a very large society would be impossible merely because we could not have mirroring experiences with enough people to make it work.

Moreover, the objection to interpersonal comparisons (raised by Pareto, for example) is that any such notion is metaphysical.[23] There is simply no meaning in the claim that your welfare is comparable to mine or that your welfare can be summed with mine to get a measure of total welfare. It is clear that we regularly do make such comparisons and that some of them are compelling. For example, I would readily grant that your crushed leg is worse than my itchy finger. Nonetheless, summing welfare across all members of some group might still seem metaphysically drunken. We can probably go further with Hume and suppose that we actually can and do share the feelings of some others, those whose emotions we can and do mirror. Hence, Pareto may be partially but not entirely right in dismissing such sharing of another's feelings.

After Moore, there were two major insights in economic value theory that should have become central to debates in moral value theory, especially utilitarian value theory. These were a set of insights due to Pareto.[24] Pareto supposed (1) that we can only do ordinal value theory, and (2) that there can be no interpersonal values. I can say that I prefer X to Y, but I cannot say by how much I prefer it. And no one can say that I am better or worse off than you are. These claims place two massive constraints on aggregation. First, we can say that some group (G) of people is better off under one policy dispensation than under another only if every member of G is better off, or if some member is better off while no member is worse off. That implies that we often cannot say that one policy is either worse or better than another.

The second of the Paretian constraints is that we cannot maximize over possible outcomes. We can only say that one outcome is ordinally, not cardinally, superior to another, because we cannot add welfares to get a largest sum of welfare for a group. This means that, instead of a maximum value to a group, we can only speak of the mutual advantage from

[23] Pareto, *Manual of Political Economy,* 47–51.
[24] Ibid.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

any enhancement of welfare for the group. This is a useful concept because mutual advantage is essentially a group-level equivalent of individual-level self-interest.[25] But it is not a moral principle, and it is indeterminate, often grossly indeterminate.

Now recall the account of mirroring, in which there is a further trick that might block aggregation: evoking an emotional response of a similar kind in response not to the interests of an individual but to *the interests of society*. Identification with the interests of society must be very weak psychologically, and it is grounded in reason more than in sympathy. The neurophysiological studies probably cannot address such an abstract phenomenon as responding psychologically to the interests of society. As noted, seeing the interests of society forwarded or abused is not comparable to the visual cue of a smile or frown.

The standard measure of aggregate welfare in a nation is gross domestic product, or GDP. For average welfare, the standard measure is GDP per capita. GDP is not easily measured across varied economic conditions, such as, say, the conditions of the United States and Kenya or even the conditions of the United States in 2009 and in 1949. But when the differences are substantial, we can arguably take the comparisons they yield as at least ordinally correct. Hence, the welfare of the United States is greater, both in aggregate and on average, than that of Kenya. When these two comparisons (aggregate and average) match, as in this case, we can commonly conclude that the debate between advocates of aggregate utility and average utility is not an obstacle to utilitarian agreement.

GDP is not strictly a utilitarian measure, because it is not a measure of utility or welfare. Rather, it is a measure of some part of the resources and consumptions of the relevant nation. Hence, it is at best a proxy for welfare. It is not a very accurate proxy, because it does not capture all that might matter in a welfare assessment. It is especially poor for comparing welfare under substantially different economic conditions, as when comparing different nations at different stages of economic development or even the same nation at radically different points in time. But in these cases, we can often suppose that the differences are so great that GDP gives a correct ordinal ranking.

With ordinal interpersonal comparisons, the only aggregate comparisons we can make are analogous to those of set theory, in which we lay our comparison sets side by side to compare them one-to-one. When we lay the integers and the even numbers side by side, we discover that they are equally numerous, because there is exactly one even number corresponding to each integer, whether odd or even. Similarly, we could make a one-to-one comparison of the peoples of Iceland and Bangladesh. If there are 300,000 people in Bangladesh who are at least as well off as the

[25] See Russell Hardin, *Indeterminacy and Society* (Princeton, NJ: Princeton University Press, 2003), chap. 3.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

300,000 people of Iceland, then there is at least as much welfare in Bangladesh as in Iceland. If there are not 300,000 people in Bangladesh who are as well off as the 300,000 people of Iceland, then we cannot say whether there is as much welfare in Bangladesh as in Iceland.

Such an ordinal comparison might yield compelling results in some cases. For example, on a one-to-one ordinal comparison of their present states with their states several decades ago, the welfare of Congo and Uganda may have fallen over the past several decades for the best-off people in each nation (up to the number of people who lived there several decades ago). Because the populations of these countries have increased, however, we cannot conclude from these incomplete one-to-one comparisons that overall cardinal welfare has fallen. What we plausibly could conclude is that different policies or regimes in those nations during those decades would likely have produced changes that entailed greater welfare on the one-to-one comparisons across time.

One might argue that we still need to be able to make such comparisons as that between Iceland and Bangladesh on the ground that there is a chance we will face a real policy choice that involves analogues of such comparisons. Outside the playful realm of extremely lunatic examples, it seems unlikely that this is true. Lunatic examples seem to give many philosophers and others pleasure, so they should not be ruled out of order. But they have no policy or action implications beyond the intrinsic joys of playing with them.

Some writers seem to suppose that testing theories against such examples is a necessary part of deciding on the coherence and even the truth of those theories. Such examples might test coherence in some sense, but the test might not be a relevant one because it might bear not at all on the coherence of the theory in real-world application. Such examples surely cannot test the "truth" of any theory by seeing whether it is applicable—in current jargon—in all possible worlds. The notion of "possible" here is odd, because some of the possible worlds are those in which the laws of physics are radically (and selectively) violated. This is, of course, also the fundamental trick of most science fiction, in which plot lines often depend on the capricious mixing of violations of the laws of physics along with adherence to them most of the time. The choice in each moment is more a matter of moving the plot along or of achieving hokey effects than of working out the nature of a possible world.

Since it has taken uncounted thinkers millennia to work out a halfway coherent account of our one actual world, it would be an extraordinary surprise if a playful philosopher—a rank amateur in physics, and probably in logic as well—somewhere could work out the slightest bit of an alternative world with any hope of keeping it coherent. Indeed, in any moral theory that is consequentialist, all one needs is the very limited talents of a fourth-rate science fiction writer to contrive a test that twists causal effects to make efforts to do good produce bad results. Alas, such

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

talents are in such abundant supply that much of critical moral theory is a wasteland. Not even that much talent is needed to test any theory against one's so-called intuitions.

V. Indeterminacy

For many theorists, ordinal theory has a major drawback: it is indeterminate. If the best account of the real world in other ways, such as causal relations, is not determinate, however, then it is odd to complain of our moral theory that it is not determinate. It cannot be determinate if it is to be relevant. Just how indeterminate our moral world is is suggested by game theory, which gives a descriptive account of our choices and their implications. The main insight is, of course, that an individual typically cannot determine the outcome of any interaction of social significance. I can choose a strategy, and I can do my best to make my strategy choice anticipate the strategy choices of all others with whom I am interacting in the context of my choice. The range of choices available to me in social contexts is often staggering and staggeringly complex. Theorists commonly seek equilibriums, but there may be none. There is commonly not even a best choice for any of us to make.

Ordinal utilitarians, faced with the indeterminacies of our world of social interaction, must generally be modest in their recommendations for both personal actions and public policies. It is a mistake to insist on Benthamite cardinal and interpersonally comparable welfare, because its assumptions do not fit the indeterminate nature of our world and our lives.

If we put this constraint together with the indeterminacies of our world of choices, we can see massive limits to making ostensibly good choices in the sense of making choices that entail good outcomes for ourselves (in the context of self-interest) and for all (in utilitarian contexts). Often, we can make very good choices. This is particularly true in contexts of coordination and especially contexts of conventionalized coordination, as in such simple interactions as driving on the highways. In the case of driving, and in many other similarly repetitive interactions, we can come to be extremely confident of how all others are going to behave, and we can therefore be very confident of what is the best choice for ourselves. There are even some interactions in which there is only one plausible coordination choice. Such interactions are called games of harmony. In such games, we do not even need to have an iterated interaction. We know in even a single-shot interaction that one strategy choice is better than any other, and we can be confident that others will see this fact as well as we do, so that we all make the right choice of strategy. But this is a happy case unaffected by indeterminacy. Although it might be hard to establish numbers of all interactions or even of all important interactions, it seems very unlikely that harmony is a very common prospect. Most prospects are

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

indeterminate and conflictual, so much so that we cannot stipulate what is the best strategy choice whether for self-interest or for utilitarian intentions.

VI. Institutions

An obvious fact of our lives is that we commonly require help from social institutions to accomplish the things we wish to do. As a general background, for example, we need institutions to maintain social order so that we may go about our own lives without concern to watch our backs at all times. One of the things that our institutions can do for us is to override indeterminacies to give us greater predictability in our actions. In much of the writing on institutions, one has the sense that the authors suppose that we can simply put an institution in place and then expect it to fulfill its purpose or mandate as though there were an overall designer of the institution, a designer whose plans for the institution will be carried out as designed. Any significant social institution is sure to violate any such hope. Our institutions eventually reach a state at which we would say they are grown rather than designed in fundamentally important respects. Hence, our institutions evolve over time so that they are examples of unintended consequences. Much of this evolution is also indeterminate and relatively unpredictable.

Hume presents a nascent legal philosophy that starts from this vision of the system of justice as itself an unintended "design." This system evolves to meet changing conditions, such as technological and demographic changes. Hume's legal theory was displaced by one of the most antithetical possibilities: Jeremy Bentham's rigidly designed system of positive law. Against the view of Thomas Hobbes—that there must be a draconian sovereign who is above the law and has absolute power—Hume supposes that conventions can control our governors enough to obviate the need for such a freewheeling sovereign.

Welfarism is often dismissed as a bad moral theory. Note, however, that if welfarism is a bad moral theory, then it is also a bad pragmatic theory in the sense that it would not be a good guide to personal choice that is self-interested. But to suppose that it is a bad pragmatic theory is so utterly implausible that one cannot sensibly hold that it is a bad moral theory. At the very least, it must be a significant part of any credible moral theory that is broadly applicable, especially one that is applicable to policy choices. This conclusion is particularly true for any effort to design institutions for regulating and managing social life with all of our necessarily indeterminate strategic interactions with each other. This is most conspicuously true, perhaps, in the context of designing institutions for our social life. The institutions of justice, for example, have a purpose which is essential and whose service to us is broadly beneficial. Without good institutions of justice, society would be far less orderly and far less likely to enable us to have good lives.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

VII. Transfers

We might not be able to say, on a strictly ordinal assessment, that there would be an increase in welfare overall if the wealthy nations transferred substantial sums to the poorest nations. There might be causal problems in the way of such an assessment. For example, we might rightly suppose that transferring funds to a nation with a rapacious dictator such as the late Mobutu Sese Seko (president of Congo/Zaire from 1965 to 1997) would have had little effect on the welfare of the people of that nation. He would have invested much of the booty abroad on his own account and the rest in enhancing the capacity of his nation's police to suppress popular movements against his regime. Or funds transferred to a nation that then (or soon thereafter) has a viciously fundamentalist regime might be used to expand the regime's fundamentalist grip and to encourage higher rates of birth.

But suppose there are no such causal worries or, rather, to be more realistic, suppose such worries are less severe, so that we might suppose the transfer of funds from the United States to, say, Kenya would not be entirely ineffective but would have substantial positive effects, even though some significant fraction of the funds would be wasted—just as there would be waste in any domestic welfare agency or charitable organization. It would be odd to suppose that foreign aid programs could avoid similar inefficiencies. Further suppose what must be true of many plausible aid programs, namely, that without the aid, conditions in some nations would deteriorate over time, as has been happening in much of Africa for the past several decades.

What ordinal comparisons might we then be able to make? We could likely not say that the overall welfare of the donor and the recipient nations with the transfer is greater than that without the transfer, or vice versa. But if the economy of the wealthy donor nation is growing in per-capita terms, we might well be able to say that the overall welfare after some years of the transfer is greater than that before the transfer. Indeed, the per-capita welfares of the citizens of both the donor and the recipient nations might increase on the proxy GDP per capita measure. Foreign aid programs that transferred less than the amount of growth of GDP per capita might sound trivial, but they could radically exceed the actual programs of the past half-century.

Unfortunately, we would not be able to say that the state of affairs after a few years with the transfers is ordinally better than that without the transfers; those two states would be noncomparable. Indeed, the status without transfers could also be noncomparable to the original status quo ante, because the poorer nation might be worse off after a few years than it was in the status quo ante. This is the recent history of much of Africa. The only way there could be an ordinal improvement in that case would be with transfers adequate to improve the conditions in the poor nation

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

into the future by, for example, creating productive enterprises and institutions that would enable the aid recipients to secure their own welfare. Someone who supposed that the real issue is cardinal addition of welfares might contend or implicitly suppose, therefore, that the (ordinal) conclusion that the transfer would increase welfare is (cardinally) wrong. But this conclusion is inherently meaningless if there is no cardinal additive measure. In the ordinal comparison of the two possible policies—transfers or no transfers—with the status quo before the period in which the resources for the transfer are created within the donor nation, only the policy of making the transfers is ordinally superior to that status quo ante. Transfers would only go from the ordinally better off to the worse off, and hence only from the relatively prosperous to the poor. We can therefore morally defend a transfer policy, though we cannot similarly defend the policy of no transfer.

This conclusion is similar to the ordinal utilitarian defense of promise-keeping. Suppose that I promise to do x in return for your doing y and that you have now done y. A standard, somewhat silly, refutation of utilitarianism is that it cannot support a claim that I ought now to keep my promise (while, of course, we all know that promise-keeping is morally required). The criticism is that once you have done y, our promise is irrelevant to the judgment of the further utility of my doing x, so that I should consider *de novo* whether to do it.[26] This criticism is far too fast. When we initially promise, all we may know is that we are both better off with the exchange of x for y than in the status quo ante. We do not have a very fine-grained measure of the utility to each of us of the full range of possibilities. These include the status quo, my performing while you do not, your performing while I do not, and our both performing. For typical contexts of promise-keeping, we may be able to rank only the first and the last of these, and we make our promise to one another only if we both are better off after joint performance than in the status quo. But this means that, once you have performed, I have no utilitarian ground on which to renege. I can make our interaction utilitarian only by performing as promised.[27]

A correlate of the claim that aid programs to nations in which conditions are deteriorating can be utilitarian on a strictly ordinal account is that aid to nations in which conditions are improving need *not* be utilitarian. For example, on the proxy measure of per-capita GDP, for three decades South Korean welfare was increasing even without aid from abroad. An aid program from, say, the United States to South Korea would therefore not have been uniquely ordinally superior to no aid program. Both nations would have been better off after no aid and also

[26] D. H. Hodgson, *Consequences of Utilitarianism* (Oxford: Oxford University Press, 1967), 38–42.

[27] For further discussion, see Hardin, *Morality within the Limits of Reason*, 61–63.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

after aid. On these factual assumptions, an international analog of Rawlsian distributive justice might have commended transfers from the United States to South Korea. A strictly ordinal utilitarianism could not have done so, but could only have been indeterminate in its policy implications.

VIII. Population

Rapid population growth seems to be deleterious to longer-term prospects of improved welfare for many impoverished nations. As an extreme example, in the decades before its recent bloodbath and subsequent wave of emigration, Rwanda had annual rates of natural increase in population of more than 3 percent, doubling its population in about twenty years. Rwandans lived mostly as subsistence farmers (more than 90 percent) on too little land, and population growth reduced average prosperity. Consider some possible responses to such growth. The traditional utilitarian argument of those who plumped for higher aggregate utility seems to imply that individuals have a moral duty to procreate in order to increase total welfare in a future world. This is a strangely abstract conclusion that is unrelated to the moral urge of utilitarianism, which is concern for the welfare of actual people. That moral urge is the same as the face-to-face concern with particular persons. It is not something that ignores the face-to-face concern and substitutes some abstract principle or entity for it. The vision that abstracts from real people has, among other things, given rise in our time to the perverse debate over respect for persons and the vacuous claim that utilitarianism is morally inferior to certain other moral theories because it violates respect for persons.

What is the ordinal utilitarian recommendation on population? If the next generation is worse off (individual by individual) than the current generation, there can be no utilitarian claim for increasing population. If some policy would lead to a next generation that is better off (individual by individual) than the current generation, then that policy would pass a utilitarian test. There might be more than one such policy, and the results of the several policies might not be comparable in the utilitarian sense that one of them could be judged to entail greater welfare than any other. But the fine-tuning of policies is, in any case, out of the question because our social science is inadequate to assess finer differences in the causal consequences of various policies. Hence, we will likely be able, at best, to judge the general tendency of any policy. It seems plausible that the general tendency of policies to employ people more fully, which typically will entail reducing their fertility, will make them better off and will make the next generation better off than the current generation on an individual-by-individual comparison.

The implication of the preceding paragraph is that a utilitarian policy might entail reduction in fertility rates. Would a policy merely of reduction of fertility produce greater welfare? This question might be harder to

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

answer. For a family living on tightly limited fixed resources, such as a small subsistence farm, more children are likely to mean lower levels of welfare for the family members individually. Fertility control may, however, trade off with other policies for enhancing welfare in impoverished nations. Absent other such supports, however, fertility control may commonly be beneficial. An autarkic state might have a utilitarian duty to reduce fertility, as India and China have attempted to do in recent decades.

IX. Conclusion

Ordinal evaluation can have seemingly anomalous implications. For example, an inegalitarian Bangladesh can be compared to Iceland if its three hundred thousand best-off citizens are better off than the citizens of Iceland, while a Bangladesh with the same resources overall, but more equally distributed, might not be comparable one way or the other to Iceland. Hence, (1) the inegalitarian Bangladesh might be found to have greater welfare than Iceland, while (2) the egalitarian Bangladesh might be found to be, on a Rawlsian or other assessment, more just than the inegalitarian Bangladesh, while, nonetheless, (3) the egalitarian Bangladesh could not be compared to Iceland either on a Rawlsian or on an ordinal utilitarian assessment. The first of these pairs permits a welfarist comparison; the second permits a Rawlsian justice comparison; and the third permits neither comparison. Although such implications may sound strange on first hearing, they are likely to be commonplace in an ordinal world. If such implications sound strange, we may find it difficult to redirect our understanding to think in ordinal terms. But that is what we must do if we wish to make consequentialist sense of our real world. The alternative—to dismiss welfarism as a moral and policy-guiding theory because of its indeterminacy—is a bad move, not least because all the other plausible candidate theories are also indeterminate. A determinate theory cannot track the world to which it is to be applied.

Note an interesting fact about actual policymaking under institutions that might have been adopted under ordinal constraints. We might generally create institutions and then let the institutions as created make and execute policies. We could reach unanimous agreement on an institutional structure—at least in principle, although this is perhaps unlikely in actual practice. The structure might itself allow less than unanimous choices of actual policies. Indeed, it could even allow violation of the constraints of our value theory. Thus, for example, even while recognizing that interpersonal comparisons are not meaningful, we might nevertheless use them. We might even allow cost-benefit analysis despite our theorists' best arguments against them. We could do that in order to get policies adopted and to keep our society on the move. It is mutually advantageous to have government that uses rules or procedures that are not themselves mutual-advantage devices. All of the metaphysical and nor-

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press

mative objections we might have to cost-benefit analysis, for example, need not block our pragmatic preference for using it in actual practice.[28]

An inherent implication of an ordinal utilitarianism, as opposed to a Benthamite cardinal utilitarianism, is that we must be more modest in our claims about the good we might do in the world. Often, our theory yields ambiguous implications or, essentially, no implications at all for what policy we should adopt. This is a conclusion that all moral theorists might well take to heart. If moral debate allowed greater play for ambiguity and indeterminacy, it would have a less assertoric and dismissive quality. It would also be more relevant to the indeterminate, ambiguous world in which we live. That is, of course, the world to which we want our moral theory to apply.

Politics, New York University

[28] Ibid., 53 and chaps. 4 and 8.

https://doi.org/10.1017/S0265052509090025 Published online by Cambridge University Press