Introduction

In 1987 and 1988 we organized two conferences, one at the University of California (Davis) and one at the University of Chicago, to study the problem of interpersonal comparisons of well-being.[1] We enlisted philosophers, economists, political scientists, and psychologists in the hope that an interdisciplinary approach would help to make some progress on a notoriously intractable issue. The chapters in this book do not yield a complete solution to the problem. We believe, however, that they make us understand better what would count as a solution. Also, they illustrate a number of approaches to the practical task of comparing the well-being – welfare, utility, standard of living – of different people.[2]

We may distinguish among four main problems that arise in this context. First, what do we mean by well-being? Second, is a comparison of the well-being of different people at all meaningful? Third, can such comparisons actually be carried out, precisely or at least approximately, routinely or at least under ideal conditions? Last, how does the purpose for which the comparison is carried out affect the answer to the first three questions? The topic of the chapters in this book is, "How to construct a notion of well-being that is (1) interpersonally comparable, and (2) adequate for purposes of distributive justice." Whenever the contributors object to a particular concept of well-being, it is rarely on the grounds that comparisons of well-being in this sense are devoid of meaning, but rather that the concept itself is inappropriate or that it is inappropriately operationalized, given the

[1] The proceedings of an earlier conference on related topics, Foundations of Social Choice Theory, were published in the same series in 1986. In that book, Allan Gibbard and Donald Davidson discuss the problem of interpersonal comparisons of utility in papers that are widely cited in this book.

[2] Here and later, we use "well-being" as the generic term for the comparandum. As will become abundantly clear later, the specific meaning given to the term can differ widely.

normative purposes it is to serve. Consequently, we do not focus here on the issue of meaningfulness of comparison.[3]

This introduction begins in Section I by considering briefly some other purposes that might motivate the attempt to compare the well-being of different individuals. In Section II we offer a survey of interpretations of well-being. In Section III, we discuss various methods that have been proposed for actually carrying out the comparison of the well-being of different people.

I. Interpersonal comparisons for other purposes

There are at least two purposes other than to implement distributive justice for which interpersonal comparisons of well-being are necessary. First, contrary to a widespread opinion, positive or explanatory theory sometimes has to rely on such comparisons. Second, they might be needed for evaluative purposes that are unrelated to the concerns of distributive justice.

Positive theory never requires the observer or theorist to make interpersonal comparisons of well-being. To explain behavior he may, however, have to take account of the fact that the agents themselves constantly make such comparisons. An obvious example is the importance of invidious comparison in much of human life.[4] A person sufficiently concerned with relative well-being may be willing to cut off his nose to spite his face – a behavior that is hard to explain using conventional assumptions about motivation. As argued by Kahneman and Varey in their chapter, well-being itself may be determined in part through such comparisons. We then have to ask whether this concept of well-being is relevant for purposes of distributive justice – that is, whether loss of well-being due to invidious comparison is a ground for redistribution. John Rawls, in particular, argues that "excusable envy" is relevant for normative purposes.[5]

The remarks in the last paragraph illustrate the elusiveness of the concept of well-being or, rather, one of the many ways in which it is elusive. Assume that *A* is stagnating and *B* is flourishing, but that *A* is unaware of *B*'s existence. Given a specific definition of well-being

[3] For some remarks on that issue, see Gibbard (1986), p. 183–4.
[4] For a brief survey, see Elster (forthcoming).
[5] Rawls (1971), p. 534.

and a way of measuring it, we might conclude that A's level of well-being is 4 and that of B is 7. Assume now that A becomes painfully aware of B's existence and his higher level of well-being. According to one conception, A's level of well-being remains the same, although he now suffers pangs of envy that he was previously spared. According to another conception, the awareness of B's prosperity has reduced A's well-being from 4 to, say, 3. But then A must be expected to carry out another comparison – between B's level and the level to which he, A, is reduced after the first comparison. Presumably, the process will converge to some well-defined level of well-being for A. One might wonder whether this logically impeccable construction is phenomenologically plausible. Sometimes, it probably is. The envious man can feel even more miserable when he notices his envy. More often, perhaps, it makes better descriptive sense to say that the level of well-being is unaffected, because the pangs of envy do not really belong to the same category. This approach is related to a suggestion made in Scanlon's chapter to the effect that experiential states may be a (multidimensional) vector of heterogeneous components rather than a (one-dimensional) scalar.[6]

Interpersonal comparisons also have explanatory relevance in bargaining theory. Within the framework of Nash bargaining theory, outcomes are predicted given the set of feasible utility pairs and the particular utility pair corresponding to the disagreement outcome.[7] One of the Nash axioms (linear invariance) can be interpreted as precluding the interpersonal comparison of utilities. Therefore a solution like "split the utility difference" cannot even be defined within the framework, as the difference of utilities between persons is not well-defined in Nash's model. Nevertheless, it is clear that interpersonal comparisons are made by bargainers, and that they do affect the outcome. Consider, for example, the following bargaining situation. 1 and 2 are bargaining over 100 chips. The number of chips a person has determines his probability of winning a lottery: Assuming that they agree on an efficient division $(x, 100 - x)$, 1 has a chance of $x\%$ of winning a bicycle in a lottery L_1 and 2 has $(100 - x)\%$ chance of winning a bicycle in another lottery L_2. For specificity, let us assume

[6] Cf. also Sen (1980–81).
[7] A simple exposition, sufficient for the present purposes, is Luce and Raiffa (1957), Ch. 6. See, however, Roemer (1990) for a criticism of this approach.

that they agree that each should have 50 chips. Now assume that 2's lottery is changed to L_2', with the prize now a Rolls-Royce. According to Nash's linear invariance axiom, the outcome of bargaining in terms of the distribution of chips should be the same as before, a fifty-fifty division. In actual bargaining this would never happen. Because 1 perceives that a Rolls-Royce is worth much more to 2 than a bicycle is to him (1), 1 can credibly threaten to break off the negotiations unless he gets, say, 95 of the chips. We may quote Ehud Kalai on this topic:

> The argument just presented suggests that in bargaining situations, interpersonal comparisons of utility may take place. For example, this can be detected in the statement "Player 2 stands to gain more than player 1". Many game theorists like to disregard theories involving these types of interpersonal comparisons on the grounds that they cannot be done using von Neumann–Morgenstern utilities with their arbitrary scales. However, if one believes that interpersonal comparisons do take place in bargaining situations, then it would be a mistake to ignore them because they inconvenience us when put together with individual utility theory.[8]

To repeat, the theorist need not himself carry out the task of comparing the utilities. His concern is to model the comparison made by the parties. These comparisons might well be inaccurate, based, for instance, on the kind of distortions discussed by Kahneman and Varey in their chapter.

Interpersonal comparisons may also be made for evaluative purposes that are unrelated to distributive justice. It is sometimes said, for example, that evil persons are never happy. Someone might want to verify this proposition for, let us assume, religious purposes. If it could be proved, it might make converts out of skeptics by providing evidence of divine justice. For the purpose of this comparison, the relevant notion of well-being includes the pleasures that evil persons might get from torturing others. The potential convert will not be impressed if told that evil persons are unhappy if we subtract the

[8] Kalai (1985), p. 89, slightly modified to fit our example. But see Roemer (1990) for a resolution of the Rolls-Royce 'paradox' without abandoning von Neuman – Morgenstern utilities.

pleasures they derive from satisfying their evil desires. For purposes of distributive justice, however, we do not want to count satisfaction or nonsatisfaction of evil desires as grounds for redistribution (see section II).

II. The comparandum

The chapters in this book discuss a number of concepts of well-being that may lend themselves to interpersonal comparisons. Along one dimension, they may be divided into subjective mental states (hedonic satisfaction), degree of objective satisfaction of subjective desires, and objective states. Along another dimension, they may be differentiated through the principles by which states of pleasure or desire-satisfaction are admitted or discarded as components of well-being. We shall offer a selective survey of some of these concepts.

Within subjective mental states, several subdivisions can be made. First, there is Scitovsky's distinction between pleasure and comfort, explained and discussed by Kahneman and Varey. This is an old distinction. Leibniz, replying to Locke, insisted that rather than pain ("uneasiness") being the opposite of pleasure, discomfort and small pains are a core component of pleasure:

> "Je trouve que l'inquiétude est essentielle à la félicité des créatures, laquelle ne consiste jamais dans une parfaite possession qui les rendrait insensibles et comme stupides, mais dans un progrès continuel et non interrompu vers de plus grands biens".[9]

Pleasure is produced by the continual overcoming of the pain and discomfort attached to unsatisfied desires. For purposes of distributive justice, comfort is more important than pleasure. It is more important to transfer resources to people who experience constant discomfort than to help those who, because of the lack of occasional discomfort, are unable to get much pleasure out of life. The latter are just planning their lives badly, and have no claim on our assistance.

Second, we may distinguish between personal and other-directed components of subjective well-being. The latter include the sadistic

[9] Leibniz (1875–90), vol. 5, p. 175; see also p. 152.

and spiteful pleasures we derive from the pain of others,[10] as well as the altruistic pleasures we derive from their happiness. There is general agreement that if the purpose of interpersonal comparison is to prepare the ground for transfers of resources, the former should not count. If I am miserable because I am not allowed to indulge in my favorite activity of torturing others, I have no claim to compensation. Conversely, people should not lose claims on compensation if they happen to get a lot of pleasure from the pleasure of others. Amartya Sen's distinction between (what he calls) well-being and standard of living is relevant here.[11] The former includes, as the latter does not, other-directed pleasures.

There are strong reasons for not using subjective satisfaction, even if shorn of other-regarding pleasures, as the standard of well-being. The most important, perhaps, stems from the "happy slave" or "sour grapes" problem. The fact that people adapt to their current state, and the fact that some adapt more easily than others, make it difficult to build a theory of distributive justice on a purely subjective criterion of well-being, as emphasized by Scanlon and Kahneman–Varey in their chapters. The issues are complex. Tocqueville, for instance, asks:

> Should I call it a blessing of God, or a last malediction of his anger, this disposition of the soul that makes men insensible to extreme misery and often gives them a sort of depraved taste for the cause of their afflictions? Plunged in this abyss of wretchedness, the Negro hardly notices his ill fortune; he was reduced to slavery by violence, and the habit of servitude has given him the thoughts and ambitions of a slave; he admires his tyrants even more than he hates them and finds his joy and pride in a servile imitation of his oppressors.[12]

From one point of view, a slave is better off if he adjusts to his circumstances – the adjustment is a "blessing of God." This cannot,

[10] The distinction between sadism and spite is that in the former case I derive pleasure from the pain I impose on others, whereas in the latter case, pain of any kind satisfies me equally. Acting out of spite, I may cause another to feel pain, but my pleasure derived from this pain is not larger than the pleasure derived from any equal-size pain caused by other factors. For this distinction, see also Hirschleifer (1987).

[11] Sen (1987), p. 27ff.; also Feinberg (1984), p. 74.

[12] Tocqueville (1969), p. 317.

of course, be relevant for evaluating justice. From another point of view, the adjustment makes him even worse off – It is "a last malediction of his anger." Marxists, especially of the Frankfurt School, argue that alienation is even more severe when (and because) it is not subjectively experienced.[13] The underlying idea is that having autonomously formed preferences is itself a component of well-being, regardless both of the actual content of those preferences and the extent to which they are satisfied. If autonomous preferences are those we would not mind having even if we knew how they had been formed,[14] an infinite regress is lurking. The trade-off between autonomy and (subjective or objective) preference satisfaction can itself be determined only by an appeal to preferences, the autonomy of which might then be questioned.

Another standard objection to the purely subjective concept of well-being is that it fails to take account of the fact that we are inclined to say that peoples' lives go well for them when their strongest desires are satisfied, even if they are not aware of the fact. Allan Gibbard writes:

> It seems clear enough . . . that people want things other than feelings of well-being: sometimes they want revenge, sometimes posthumous fame, sometimes the fidelity of friends or spouse, sometimes the well-being of others. A jealous husband may even prefer a 'fool's hell' in which his suspicions rage but his wife is in fact faithful, to a 'fool's paradise' in which his suspicions are allayed but in fact he is unknowingly cuckolded.[15]

In this book, Scanlon makes roughly the same point. There is, however, a difficulty in this view. Consider the following states:

A. Jack's wife is faithful to him. He wants her to be. He believes she is not.

B. Jack's wife is not faithful to him. He wants her to be. He believes she is.

13 See, notably, Geuss (1979).
14 This is roughly the concept of autonomy proposed in Elster (1983), Ch. I.
15 Gibbard (1986), p. 169.

Now, in state A, Jack's desire is satisfied; in B it is not. In state A, Jack is unhappy; in B he is happy. Now, assuming that both preference-satisfaction and happiness are valued goods, there might be choices and trade-offs between them. Following Gibbard, we may ask Jack whether he would prefer A or B to be the true state. Assume he would prefer B, but that A is the state that actually obtains. His preference for B over A is not satisfied. Then we have to weigh that fact against the other aspects of A. Perhaps, were Jack asked to choose between the state in which he prefers B but A obtains and the state in which he prefers A and A obtains, he would prefer the latter. But then, suppose that B is the state that actually obtains, and that Jack has the preference just stated. We can go on forever; the argument seems incompletable. We should not say (and Gibbard and Scanlon are careful not to say) that A is better just because Jack's preferences are satisfied in that state, whereas in B they are not. That would be to give too much weight to objective satisfaction, and too little to hedonic well-being. But if the weights assigned to preference-satisfaction and well-being are themselves a matter of preference, we again run into an infinite regress.

An objection that can be made both to the hedonic and the desire-satisfaction conception of well-being is that they give too much weight to aspiration levels. Egalitarian concerns might lead us to transfer resources to individuals at low levels of "personal success," to use Ronald Dworkin's phrase for degree of preference-satisfaction, even if the level is low only if measured against their excessively high ambitions.[16] In his chapter, Scanlon makes a similar point when he remarks that preference-satisfaction has no moral force unless backed by an argument about what there is reason to prefer. This is in a way the converse of the "happy slave" problem. Just as we do not want to punish those who adjust to circumstances by excessively low aspiration levels, we do not want to reward those who for some reason form unrealistically high aspirations. As Dworkin remarks, the notion of ambitions being "excessively" high already presupposes a theory of just distribution, and hence cannot without circularity or regress be used as a premise of this theory.

Instead of looking to the person as the final arbiter of what is good for him, we might appeal to more impersonal standards. In this book,

[16] Dworkin (1981), pp. 204ff.

both Griffin and Scanlon move in that direction. Griffin argues that we cannot assess a person's well-being as the satisfaction of private, possibly idiosyncratic desires. Instead, we must assess the desires and hence his well-being from the point of view of a world of shared values. Scanlon, moving even further away from the subjective conceptions of well-being, argues for a conception related to Rawls's notion of primary goods – goods that any person would need in order to carry out his life plan. A difficulty with Rawls's conception, from the point of view of interpersonal comparisons, is that it is a vector rather than a scalar. If one bundle of primary goods is better than another on some dimensions and worse on others, there is no way of telling which is larger and better. The best we can achieve is a partial ordering of bundles.

Amartya Sen has argued against the primary-goods conception of well-being on the grounds that it embodies a "fetishist handicap."[17] One should be concerned not with goods as such, but with what goods do for human beings and what human beings do with them. Sen advocates a different nonsubjectivist conception of the good, based on the individual's freedom and capability to act.[18] The proposal might, however, have some unacceptable consequences. Often, what matters to the handicapped is not to reach the same end-states as the nonhandicapped: It may also matter to reach them in the same way. They may prefer to go to work by train like everyone else rather than being reimbursed for taxi expenses, even if the special modifications needed for the former solution cost more than taxi subsidies and the nonhandicapped would much rather go by taxi than by train. They may prefer to work for an income rather than receiving an outright pension, even if the extra equipment needed to enable them to work costs more than the value of what they produce and the work is inherently boring and monotonous. These preferences would appear to be condoned by Sen's proposal to focus on "basic capabilities" as the target of allocation. They might, however, be thought to be somewhat irrational.[19] For purposes of resource transfers, end-state comparisons may be more relevant.

17 Sen (1979); see also Cohen (1989) for illuminating discussions of the issue.
18 Sen (1985, 1987).
19 Elster (1988) argues that work created solely for the purpose of promoting self-respect cannot be rationally counted on to achieve even that end.

We must distinguish between objective conceptions of well-being and objective proxies for subjective well-being. Even if one adopts a wholly subjective notion of well-being as pleasurable mental states, any attempt to measure and compare well-being thus defined would have to rely on observable features of the individuals involved. On the basis of introspection and reports from others, one might decide that the correlation between a suitable bundle of observable traits and possessions is high enough to use the bundle as a proxy for well-being. To be sure, such correlations are never perfect. A striking instance in which the subjective value of a life could not be predicted from external circumstances is found in a play planned by Jean-Paul Sartre:

> Colette Audri, with whom Sartre once discussed this play, tells us that the play was to be called *The Wager* (after Pascal's wager), and would concern a child who is not wanted by his father. The mother, however, does not let herself be pressed into abortion, although a horrible life has been prophesied for the child: severe trials and reverses, poverty, and finally death at the stake. The child is born, grows up, and everything takes place as prophesied. 'In fact he changes nothing material in his existence,' Sartre says, 'and his life ends, as foretold, at the stake. But thanks to his personal contribution, his choice and his understanding of freedom, he transforms this horrible life into a magnificent life.'[20]

For practical purposes of redistribution, it would be impossible to take account of such deviations. (On the grounds that nobody should be punished for a great fortitude of character we might not even want to do so. But that is another argument.) We now discuss matters of implementation more systematically.

III. The comparison

Let us assume that there is a *fact of the matter* in an interpersonal comparison of well-being, considered as hedonic enjoyment or satisfaction of desires. It does not follow that we could ever discover it.

20 Føllesdal (1981), p. 403–4.