



1 Your Intellectual Journey

1. Could you tell us about your journey from economics to philosophy?

I always had an interest in philosophy. When I was an undergraduate at Cambridge, I was turned away from the subject by the philosophy tutor I went to consult about it. But as a graduate student in economics at MIT, I was allowed to take some courses at Harvard, and I indulged myself by taking some philosophy. I found John Rawls's course on liberalism dull, but Stanley Cavell's course on Wittgenstein was gripping. It was his course that drew me into philosophy. Cavell was charismatic and a mesmerizing performer. I didn't understand much of the course, but I wrote it all down and I spent many hours puzzling over Wittgenstein's *Investigations*.

My first academic job was in the Economics Department at Birkbeck College in London University. This department was just being created under the leadership of Bertie Hines. Bertie had persuaded the college that all the teaching staff needed to be in post for a full academic year preparing their courses, before any students arrived. So this very fortunately gave me a year during which I was able to take a master's degree in philosophy. I became a student at Bedford College where I was taught mostly by David Wiggins.

Sadly, a master's degree was not a sufficient qualification for getting a job teaching philosophy. On the other hand, since I was well qualified with a Ph.D. in economics,

R. Bradley

London School of Economics and Political Science, London, UK

e-mail: r.bradley@lse.ac.uk

M. Fleurbaey (✉)

CNRS, Paris, France

e-mail: marc.fleurbaey@gmail.com

Paris School of Economics, Paris, France

I could easily get jobs and funding in that subject. I stayed in Birkbeck for a few years and then moved to the Economics Department at the University of Bristol. All the time I was writing about the foundations of welfare economics, which was basically philosophical. This is a place where economics and philosophy meet. So I was building up some recognition in philosophy.

While at Bristol, I was lucky enough to be invited by Derek Parfit to spend a year as a visitor at All Souls College in Oxford. He was engaged in writing *Reasons and Persons*. I learnt a great deal of philosophy during that year, and I owe a very big debt to Parfit. A few years later, I visited Princeton University, where I taught in the Philosophy Department—another fine learning experience. The graduate seminar I gave there became my book *Weighing Goods*, which I think became in due course my entry ticket to a job in philosophy.

A year or two afterwards, I moved half-time into the Philosophy Department in Bristol. The big shift came when the University of St Andrews offered me a full-time professorship in philosophy. This was the doing of John Skorupski, who is another person to whom I owe a great debt of gratitude. I had been an economist for thirty years, and I was relieved to abandon the subject at last. Still, for many years, I continued to feel an interloper in philosophy since I am not properly educated in the subject. By now—another twenty-three years on—I feel a bit more confident.

2. What do you think have been your most important ideas?

I don't remember having ideas much. What I remember is working things out. Analytical philosophy is like that. You try to get to the truth by working your way through difficulties and puzzles. I have tried to work out some parts of the structure of good, and of rationality, and of normativity. So the job is problem-solving rather than thinking up important ideas.

True, successfully working things out usually requires you to come up with a sequence of relatively small ideas that make things fall into place. For example, I remember realizing, one snowy afternoon in Uppsala, that requirements of rationality often have a wide scope covering a whole conditional statement rather than a narrow scope covering just the consequent. For example, rationality requires of you that, if you believe you ought to do something, you intend to do it. By contrast, it is not necessarily the case that, if you believe you ought to do something, rationality requires you to intend to do it. The first of these statements does not imply the second. Even if you believe you ought to do something—so the antecedent condition is satisfied—it does not follow from the first statement that rationality requires you to do it. That is all to the good. It does not seem plausible that merely coming to believe (perhaps falsely) that you ought to do something should put you under a real rational requirement to do it. This discovery of wide scope helps to solve various problems about the structure of rationality and normativity. It became associated with me, but I was not the first to discover it; Jonathan Dancy was ahead of me.

A more recent, smaller discovery (which I think genuinely was mine) is that the best account of the logic of requirements is built on something called a 'neighbourhood semantics'. This is a technical matter, but it also helps a lot of things to fall

into place. Another example was my realization that fairness requires, not the maximization of anything, but proportionality in the satisfaction of claims. This provides a good explanation of the fairness of lotteries, and it also gave me my account of the value of equality, which I presented in my book *Weighing Goods*.

Working things out has brought me to some rather extensive philosophical beliefs, which are sometimes out of line with popular thinking. I would call these standpoints rather than ideas. For example, I think political philosophy has been on the wrong track for some decades since it became obsessed with justice and started ignoring goodness. I think the philosophy of normativity and rationality has been on the wrong track since it became obsessed with reasons at about the same time. As a particular example, I think the very popular idea that rationality consists in responding correctly to reasons is badly mistaken.

Hardly anyone has noticed my cleverest intellectual achievement. I proved a version of Harsanyi's Theorem within Bolker–Jeffrey decision theory. To a mathematician, this would have been easy, but to me it was hard.

3. You started your research on taxation. Would you like to recall how this started your intellectual path?

It's true that one of my first academic jobs was to work on Jim Mirrlees's theory on optimal income tax, as his research assistant. All I did was programme the numerical solution of his equation. Later I discovered by chance that, if you constrain the tax function to be linear and apply maximin rather than utilitarianism as Mirrlees did, the optimal tax rate of income tax comes out after only a few lines of algebra. It is $(2 - \sqrt{2})$, which is to say 58.6%. I thought this made a nice parody of Mirrlees's paper, which contained many pages of extremely fancy mathematics. I hope he didn't mind my publishing it.

However, this didn't start my intellectual path. I don't think anything developed from it. Later, I was interested in welfare economics, but in its foundations rather than its applications.

My Ph.D. at MIT was on general equilibrium theory, and afterwards, I got a job at Birkbeck College in London among a group of Marxist economists. The result was that my first book was a textbook on Ricardian general equilibrium theory. However, that didn't start my real intellectual path either. That line of work petered out after the book.

My real intellectual path was philosophical from the start. At MIT, I planned to write a Ph.D. thesis on the philosopher William Godwin. I spent half a year on that subject. But although everyone was very friendly about it, I got the impression that it was not considered a suitable topic for the MIT Economics Department. All that came of that half-year's work is that when I moved to London, Amartya Sen read it and was encouraging. But at MIT, I gave it up and did general equilibrium theory instead.

When I took my MA in philosophy at Birkbeck, my thesis was about the philosophical foundations of welfare economics. During that year, I also wrote a paper about the value of human life in economics. The value of life is a topic where philosophy

and economics are very tightly connected. It suited me very well, since I worked in an Economics Department but was interested in philosophy. I have pursued this question ever since. Since the economic theory of the value of life is wrapped around with risk and uncertainty, I learnt about decision theory. Since the value of life has important practical applications to medicine and public health, I found myself involved in those subjects. And then by accident, I became involved in another application of this same topic, which is the ethics of climate change.

4. You have written on applied measures such as QALYs. Do you think there are now good empirical measures of wellbeing, or should we still invest in developing new measures?

No I don't. Those who measure wellbeing empirically generally take it for granted that a person's wellbeing is a subjective matter—a matter of how good she feels her life to be or how well she thinks it is going. But this is a big presumption and should not be taken for granted. An alternative view is that some components of wellbeing are objective and independent of what you think about them or feel about them. For example, if someone is well fed and healthy, we might plausibly think that she is to an extent well-off, even if she does not recognize or appreciate her own good fortune. If a person has a serious disability, she is to an extent badly off, even if she herself makes light of it. Before you can measure wellbeing properly, you must first work out what exactly you should measure. Philosophers have been discussing for millennia what wellbeing is. They are not going to arrive at a conclusion, because wellbeing is not the sort of concept that allows a conclusion. This does not mean the philosophers' work may be ignored. It means that any one-dimensional measure of wellbeing is bound to be inadequate. We cannot complacently think that we have good empirical measures.

5. You have been writing a lot on climate policy, and even been involved in the Intergovernmental Panel on Climate Change. What contribution are you trying to make in this domain, and what are the main points you would like people to know?

I am gloomy about our prospects. I don't think our governments are doing anything worthwhile to get on top of climate change. This has become more apparent in the last few years with the rise of populism, ignorance, and selfishness in politics. I now think that our only chance is to make use of selfishness. Climate change is bad for everyone, so in principle, everyone can be made better off by controlling climate change. This is what I would like people to know. I think we should develop institutions that could make this result achievable in practice, so it can be in everybody's interest to stop climate change. I learnt this way of thinking from Duncan Foley, another person to whom I owe a big debt. He supervised my thesis on general equilibrium theory.

But I do not intend to work much more on climate change. I am moving back towards my main academic interest, which is in normativity and rationality.

6. Could you tell us how you see the applied part of your thought and work, and what motivated you to keep an interest in applied policy?

Moral pressure and anger. When Nick Stern was starting work on the Stern Review, he persuaded me to make a small contribution. At one time, he made a gently sarcastic remarks to me about the relative importance of climate change versus the work I wanted to do on normativity and rationality. Next, when I read the reviews of the Stern Review written by some American economists, they angered me. These economists claimed that ethics has no place in economics and criticized Stern for placing it at the centre of his economics. They were wrong: ethics constitutes the foundation of welfare economics. Since I was by this time a moral philosopher, I even had a professional interest in making sure that the central place of ethics is recognized. So I wrote an article for the *Scientific American* explaining the importance of ethics in the economics of climate change. Since so many people are concerned about climate change, one result was that I received many invitations to talk and write on the subject. Furthermore, I thought that a moral philosopher should try to do something a bit useful before he dies, so I did not resist. I am a lapsed scientist, and the climate interests me anyway.

7. You have moved not only from economics to philosophy but also have been focusing lately on very abstract philosophical theory of intention, normative reasoning, and the practical implications of rationality. Is there a train of thought that logically connects your earlier work on value and goodness and this more recent research?

I'm inclined to think that value is ultimately derived from normativity (by which I mean, from ought). So there is a connection. However, I'm not working on this connection between value and normativity. I am working on the structure of normativity itself, and also on its connection with rationality. So within my own work, the connection is nugatory.

True, my interest in rationality arose from my interest in decision theory, which in turn arose from my interest in value. (I think that, despite its name, decision theory provides a better account of value than it does of decision making. I developed it as an account of value in my book *Weighing Goods*.) So there's a link there. But that link is now broken because I don't work on decision theory any more.

I do still write occasionally on value theory, chiefly in connection with climate change. So the two branches of my work are not much connected with each other.

2 Utilitarianism

8. Would you describe yourself as a utilitarian? In *Weighing Goods* you propose to incorporate a lot of egalitarianism in utilitarianism, via the measurement of utility including fairness. Can you explain why you stick to the utilitarian formalism rather than abandoning utilitarianism for a more popular approach such as prioritarianism?

Utilitarianism has several components. One is *teleology*, which is the view that you ought to do one of the best of the alternative acts that are available to you. I don't believe teleology and I don't disbelieve it either. In general, what you ought to do can be described by a choice function, and teleology is true if and only if the true choice function can be represented by a betterness relation. I don't know whether this is so.

Another component of utilitarianism is *consequentialism*, which is the view that the goodness of an act is determined by its consequences. Consequentialism comes in various versions, depending on what is included among the consequences of an act. If the consequences are taken to include the fact that the act is done, consequentialism is hard to deny. A very specific version of consequentialism is a view I call *distribution* (it is often called *welfarism*), which is the view that the goodness of an act is determined by the goodness of the distribution of wellbeing that results from it. I do not believe this. But it leads us to a further component of utilitarianism, which is the view that the goodness of a distribution of wellbeing is the arithmetic sum of people's wellbeings. This is the only component of utilitarianism that I do believe. It is an important component, of course. As you say, I believe it only under the condition that if a person suffers unfairness, that is treated as a negative component of her wellbeing.

I cannot understand prioritarianism as it is generally presented. To make sense of it, we have to have two cardinal scales: one a scale that measures a person's wellbeing and the other a scale that measures how much a person's wellbeing contributes to the overall value of a distribution. The latter is supposed to be a strictly concave transform of the former. According to a theorem of Harsanyi's, utility (which is defined within decision theory as the value of a function that represents preferences expectationally) is a scale that measures how much a person's wellbeing contributes to general value. So prioritarianism implies that utility is a strictly concave transform of the scale of wellbeing. Yet the prioritarians I know seem to assume that utility is itself a scale of wellbeing, and anyway, they offer no other scale. Utility can't be both a scale of wellbeing and also a strictly concave transform of a scale of wellbeing. So I cannot understand their view without some further explanation.

The word 'utility' causes no end of confusion in economics. In real life, it means 'usefulness'. Bentham and the other classical utilitarians used it to refer to a special sort of usefulness, namely usefulness in promoting people's wellbeing. Sometime in the decades around 1900, economists started using 'utility' for wellbeing itself. Then, another fifty years on, decision theorists and economists came to define utility as the value of a function that represents preferences. I have always regarded this as the official definition in economics. Of course, a person's utility defined this way does not necessarily measure her wellbeing. Yet economists continued to use 'utility' for wellbeing, despite the official definition. Because they used the same word with two different meanings, many of them seem to have become confused between the two. Very unfortunately, philosophers have recently begun to copy economists in their use of 'utility'. Some prioritarian philosophers may have fallen into the same confusion between two meanings of 'utility'.

To make sense of prioritarianism, there are two options. One is to deny the premises of Harsanyi's Theorem. That makes good sense, but it doesn't impress me because I think the premises are secure. The other option is to find another scale of wellbeing besides utility. This also makes sense. However, it does imply that the truth of prioritarianism is not a substantive issue. The difference between utilitarianism and prioritarianism understood this way makes no difference to the relative goodness of different worlds. Both theories agree that relative goodness is determined by the sum of utilities. Instead, the difference between theories is an issue about what is an appropriate way to set up a cardinal notion of wellbeing.

There are definitely some available cardinal notions that are alternatives to utility. At least, one is attractive. This one is modelled on the QALY, or quality-adjusted life year. The idea is that if your life continues at a constant quality, your lifetime wellbeing is proportional to the length of your life. This seems plausible. It might be good for you to be risk-averse about your QALYs, which would mean that your utility is a strictly concave transform of your wellbeing as measured by QALYs. Then prioritarianism would be true: what your wellbeing contributes to general good is measured by a strictly concave transform of the scale of wellbeing. This strictly concave transform is your utility. I have no objection to this view.

9. Uncertainty is an important element of your analysis of social goodness. Can you explain why you give it such a key role? How did you come across Harsanyi's theorem, and what made you realize its importance?

The economist's standard account of the value of human life depends on uncertainty. Economists typically say that they are not truly setting a value on life, but only on risk to life. This is a funny idea because what is bad about being exposed to a risk to your life is that you may lose your actual life. But it does mean that uncertainty is central to their theory.

The standard economist's measure of the badness of a risk to a person's life is what the person would be willing to pay to avoid the risk. This measure is not proportional to the size of the risk—to the probability of dying that the risk imposes on her. But our standard theory of value under uncertainty is decision theory, in which the value of a risk of dying is the badness of dying multiplied by the probability it will happen. This product is proportional to the probability. So the economists' standard measure of the value of life is inconsistent with our standard theory of value under uncertainty. Since I was interested in the value of life, I had to be interested in decision theory.

I then realized that the theory of uncertainty provides a useful analytical tool in value theory. This was Harsanyi's discovery. I remember working on it when I was a visitor in All Souls College, Oxford, in 1982. I was fascinated that Harsanyi's Theorem could derive such a powerful conclusion from such seemingly anodyne assumptions. I found the mathematics almost magical. The additive structure of decision theory and the additive structure of the utilitarian theory of value emerge from premises that do not mention additivity. Moreover, I discovered this is true also

in the Bolker–Jeffrey version of decision theory, which has quite different mathematical foundations. Additivity evidently has deep mathematical roots that I still don't really understand.

10. Harsanyi shared with Kolm the idea that at a fundamental level, people are alike and the differences in their preferences can be traced to different characteristics. They conclude from this claim that at a fundamental level, preferences are the same. You strongly objected to that view. Do you recall the debate, and how do you see the issue now?

Yes I do recall it. We can agree that the difference between people's preferences is explicable by their different characteristics. The authors argued that it follows different people's preferences which are 'fundamentally' the same. In their argument, they muddled the causes of a person's preferences with the objects of her preferences—what her preferences are about. You may have preferences about what career to follow—that is one thing. And the career you follow may affect what preferences you have—that is another thing. Harsanyi and Kolm confused the two.

What could they have meant by 'fundamentally the same'? They were trying to find some sort of universal preferences that could be used as a basis for interpersonal comparisons of wellbeing. They couldn't have meant merely that if two people had the same characteristics, they would have the same preferences. That doesn't yield universal preferences, but only preferences that may vary according to characteristics. I've no idea what a fundamental preference is supposed to be.

I think you must be asking this question because of the claim that appears in my book *Weighing Lives* that there is a single scale of goodness for lives. A life lived by one person is exactly as good for that person as the same life would be for a different person if she were to live it (which is not usually possible). Perhaps you think this is in some way inconsistent with what I said about the argument from Harsanyi and Kolm. There is no inconsistency. If Harsanyi and Kolm had merely meant that there is a single scale of goodness for lives, I would have applauded them. But their aim was to derive interpersonal comparisons from universal preferences. They had the economist's predilection for deriving value from preferences. The result was a confused argument.

3 Bernoulli's Hypothesis and the Representation of Betterness

11. In *Weighing Lives* you say that you doubt that there is anything more to the idea of goodness than betterness. But considerations of betterness alone don't seem to be sufficient to determine the unique measure of goodness implied by Bernoulli's hypothesis. Can you explain how this measure is to be constructed?

Take a particular person and the relation of betterness for this person. This relation can be represented in the standard expectational way by a utility function. The question arises whether utility defined this way measures the person's good cardinally, or whether there is some other cardinal measure of good. If utility measures good, Bernoulli's hypothesis is true for this person; if it does not, Bernoulli's hypothesis is false for her. But the goodness measure, whatever it is, is not determined by the person's betterness relation alone. Any goodness measure, so long as it's an increasing transform of utility, is consistent with the betterness relation. This is your point.

The measure of goodness for a person makes no difference to her intrapersonal betterness—to the person's own betterness relation. But it does make a difference to interpersonal betterness, to the general betterness relation. Suppose we hold fixed the function that relates general good to individual goods. For example, this might be the utilitarian function. Then the goodness measures for the individuals affect the general betterness relation. Personal goodness in this respect reduces to interpersonal betterness rather than intrapersonal betterness.

True, this is only in the context of a particular theory of general good such as utilitarianism. If we allow arbitrary theories, personal goodness would no longer be fixed by betterness. That is to be expected. What we mean by 'goodness' depends on how we use goodness in assessing betterness.

12. Why do you use Savage's framework for investigating betterness rather than say von Neumann and Morgenstern's? Is it because you don't think the probabilities that determine the relative goodness of prospects are objective? Or because of some other feature of Savage's framework?

John Harsanyi proved his theorem on the assumption of objective probabilities. Since objective probabilities are rare in the world, this severely weakens its significance. Since the theorem can be proved without that assumption, it is better not to make it. However, dropping this assumption does not get us far forward if we interpret Harsanyi's Theorem as Harsanyi himself did: as a theorem about aggregating people's preferences. The premises of the theorem—the Pareto principle and expected utility for individual and social preferences—together imply that everyone agrees about the probabilities of every state of nature. Probabilities are embedded in each person's preferences, and the same probabilities must be embedded in each. I call this 'the probability agreement theorem'. Since agreement about all probabilities is as rare in the real world as objective probabilities, one of the theorem's premises has to be false.

We must therefore give the theorem a different interpretation. I interpret it as a theorem about aggregating people's goods—what is good for each person—rather than their preferences. The probability agreement theorem tells us that anyone who is trying to aggregate good must apply the same set of probabilities throughout her calculation. She must evaluate the good of each person on the basis of her—the evaluator's—probabilities, rather than the person's. This is exactly what we should expect.

It does raise the question of what are the right probabilities to apply, given that there are generally no objective probabilities to go on. Different probabilities will lead to different judgements about aggregate good, so which should we choose? To be sure, they should be probabilities that are supported by the evidence. But the available evidence rarely determines probabilities fully. This seems to leave us with nothing to go on apart from our own subjective prior probabilities. That is plainly unsatisfactory, but I admit that I don't know what to do about it. Chapter 3 of my book *Rationality Through Reasoning* discusses this problem.

13. In your work, you cardinalize goodness by means of 'risk-neutral' weighing under uncertainty. Do you think other ways of cardinalizing goodness are possible, or is there an intrinsic connection between goodness and this form of weighing?

Harsanyi's Theorem tells us that two different means of cardinalizing give the same result: cardinalizing by uncertainty and cardinalizing by aggregation across people. It is convenient to use the cardinalization they agree on. But I've nothing against alternative cardinalizations. I mentioned one in answering question 8: we could cardinalize by length of time. Suppose your quality of life is a . Suppose that having it raised to a better quality b for one week is just as good for you as having it raised to a quality c for two weeks. Then we conclude that the difference between b and a is twice the difference between c and a . I haven't worked out this approach to cardinalization in detail, but I have no objection to it.

4 Interpersonal Addition

The interpersonal addition theorem tells us that if personal and general betterness are coherent and jointly satisfy the principle of personal good, then there exists an expectational utility function V representing the general betterness relations and expectational utility functions V_i for the personal betterness relations such that V is the sum of the V_i .

14. How do we get the from interpersonal addition theorem to the utilitarian principle of distribution?

As I understand the argument in *Weighing Goods*, it goes as follows. Bernoulli's hypothesis tells us that one of the expectational utilities representing a person's goodness relation measures goodness for her, but not which one. By choosing a sum-of-individual-utility representation of general betterness, a particular choice is forced upon us. *That choice determines the meaning of personal goodness*. So we shouldn't ask 'how do we know if each individual's good counts equally in overall goodness? Because what an individual's good is, is determined by such impartial interpersonal weighing. This fact also grounds the comparability of the good of different individuals.

15. Do we have this right? One possible objection is that it undermines the whole idea of providing a concrete way of constructing a measure of social goodness from scratch, since it seems to rely on a given notion of social betterness.

It's pretty much right as a report on *Weighing Goods*. You could have added some preliminary sentences. By telling us that two different means of cardinalizing good give the same result, Harsanyi's Theorem gives us some grounds for adopting their cardinalization. So it gives us some grounds for accepting Bernoulli's hypothesis, whereas there were no grounds up to that point in the book.

I never thought of constructing a notion of social betterness from scratch. I took the question to be whether we could find a coherent theory of value that fits our various intuitions about value reasonably well. It's the method of reflective equilibrium involving concepts as well as substantive theories. Formulating a quantitative notion of good is a part of this work. We must expect our notion of good to be influenced by what we do with the notion.

16. Indeed, in *Weighing Lives* you seem to reject this argument, observing that if it were true, it would literally make no sense to say that future goodness should be discounted (or that the goodness of the less well-off should count for more). But what replaces this argument in the derivation of the utilitarian principle?

Yes, by the time I wrote *Weighing Lives*, I had come to the conclusion that the method for making interpersonal comparisons of good that I adopted in *Weighing Goods* did not account properly for our intuitions. It was not in reflective equilibrium. In *Weighing Goods* I claimed that if a benefit to one person counts equally in general good as a benefit to another, that means these are equal benefits. In *Weighing Lives* I pointed out that even if these benefits are actually equal, this is not because being equal *means* counting equally. We can make good sense of the possibility that two equal benefits do not necessarily count equally in general good. The idea of pure discounting is that a benefit that comes earlier in time counts more than a benefit that comes later, even if the two benefits are equal in size. Pure discounting may be wrong, but we can make sense of it. So in *Weighing Lives*, I gave a different account of interpersonal comparisons of good. It is based on the idea that if two people live lives that are the same in all respects that affect their good, they are equally well-off.

But I continue to adopt Bernoulli's hypothesis. So I didn't have to do much more to get the utilitarian principle. I made the additional assumption that general good is impartial between the goods of different people. That is to say, permuting quantities of good among people leaves general good unchanged. That did it.

5 Personal Goodness and Interpersonal Comparisons

In *Weighing Goods*, you suggest that it is the fact that weighing gives meaning to personal goodness that grounds interpersonal comparisons. But as mentioned

above, you reject this in Weighing Lives and offer a different basis for interpersonal comparisons. In essence, as we understand it, the goodness of different persons is comparable in virtue of the fact that the goodness of a life is independent of who lives it, and hence, that everyone's good is measured on the same scale. (This requires that lives are maximally specific with regard to all facts concerning both the individual and what happens to her that are relevant to the goodness of the life, potentially including characteristics of the agent such as her personal values.)

17. In what sense are the personal betterness relations personal if they are all the same?

The ranking of lives is the same for each person, as you say. It is a personal betterness ranking because the betterness in question is betterness for the person who lives the life. It is not general betterness, or betterness for society or betterness for anyone else.

You might think that this makes little difference, because the principle of personal good tells us that what is better for a person is also generally better. You might even think that we can ignore personal good because general goodness is fully determined by the goodness of the lives that are lived, quite independently of the identities of the people who live those lives. You might think that we could attend to betterness among distributions of lives and ignore whose lives they are.

Betterness among distributions of lives is indeed independent of whose lives they are; this is a consequence of impartiality. But if we attend to the identities of people, we can gain information about the betterness of distributions that we could not otherwise get. For example, we can gain access to Harsanyi's utilitarian argument. So personal betterness cannot be ignored.

Here is a slight example that hints at what can be done. Let m be one life and n another, and compare the various prospects below. Each vector shows the lives lived by two people; in each case, it is the same two people in the same order. Assume the coin is fair.

A: (m, m) if heads; (n, n) if tails

B: (m, n) if heads, (n, m) if tails

C: (m, n) for sure

D: (n, m) for sure

E: (m, m) for sure

F: (n, n) for sure

A and B are equally good for the first person: in both she gets m if heads and n if tails. A and B are equally good for the second person: in both she gets an equal chance of m or n . So the principle of personal good tells us that the gambles A and B are equally good. Impartiality tells us that C is equally as good as D. Given that, the sure-thing principle tells us that C is equally as good as B, which is a fair gamble between C and D. Since we already know that B is equally as good as A, we can conclude that C is equally as good as A, which is a fair gamble between E and F.

It follows that when utilities are assigned to represent general betterness among gambles on distributions, C's utility must lie half-way between the utilities of E and F.

This is the first step on the road to showing that general utility is additively separable in individual utilities, which is Harsanyi's conclusion. We can take this step only because we can show that A and B are equally good. This conclusion depended on the identities of the two people, which made it possible to apply the principle of personal good.

6 The Intuition of Neutrality

18. One particularly striking argument you make in *Weighing Lives* is that the intuition of neutrality—that adding people to a population does not in itself make things better—is false. My students typically feel the pull of the intuition but reject the translation of it into the principle of equal existence: roughly that if two distributions differ only in that population of one is a superset of the other, but not in the wellbeing of those individuals who are in both, then they are equally good. They argue that the notion of adding people requires reference to a status quo point from which possible changes in population are evaluated. Do you reject any such relativization of what is better to a reference point of view?

Yes. I argued against this sort of relativism in *Weighing Lives*. I used Partha Dasgupta's relativist theory as an example because it was the only example I had. Relativism is the idea that the same thing may differ in its value according to the point of view it is evaluated from. For example, from a parent's point of view, her own children's good counts for more than another parent's children's good, whereas the opposite is the case from the point of view of the other parent. I did not argue against relativism in general, but I did argue against those particular sorts of relativism in which one person occupies different points of view at different times. Relativism of this sort implies that values from the point of view of a person change over time. This makes for an incoherent life. It may turn out wrong to do at a later time what, at an earlier time, you rightly commit yourself to do. Furthermore, you may know this at the earlier time.

For instance, suppose you know on Monday that from the point of view you will occupy on Friday, it would be best to leave town that day. Suppose you also know that Monday is the last day you can get a ticket to leave town on Friday. But suppose that from the point of view you occupy on Monday, it is better for you not to leave town on Friday. Then on Monday, you ought not to buy a ticket to leave town on Friday, even though you know that this will prevent you from doing on Friday what it will be the case on Friday that you ought to do. This is incoherent.

Population relativity threatens to lead to this sort of incoherence, because a single person is a member of several successive populations as some people are created and others die. Dasgupta proposes a way of overcoming the resulting incoherence, but I argued he is not successful.

There may be a more successful relativist theory, but I doubt it.