GRiffin, HARsanyI AND OTHERS AT THE FUZZY BORDerLINE BETWEEN PHILOSOPHY AND ECONOMICS.*

mOzaffAR quIZILbASH
University of East Anglia, norwich

abstract

This paper shares Griffin’s points of view in Well-Being and the points of view that he has articulated in his most recent writings with some writing by John Harsany about ethic and the utilitarism. This exercise allows us understand better both the writings by Griffin and Harsany. It also helps us to illuminate the strong turn of the explanation of the informed desire and Griffin’s utilitarism in his recent workings of him. The comparative exercise also suggests that in his late writings the position of Harsany was found in a marked way close to the point of view and in the line of explanations of interpersonal comparissions that Griffin has defended recently. The discussion between Griffin and Harsany is therefore referred to the writings of certain contemporary economists.

Keywords: informed desire, utilitarism, economic theory, Harsany, Griffin

Paper prepared for the special issue of Tėloc in honour of James Griffin.
I would like to thank James Griffin for his encouragement when I was considering writing this paper. I began work on the paper on a short trip to the University of Aix-Marseille. I would particularly like to thank Miriam Teschl for her hospitality while I was there.

Resumen

Este artículo comparte los puntos de vista de James Griffin en Well-Being y los puntos de vista que ha articulado en sus más recientes escritos, con algunos de los escritos de John Harsany sobre ética y el utilitarismo. Este ejercicio nos permite comprender mejor tanto los escritos de Griffin como los de Harsany. También nos ayuda a iluminar el marcado giro de la explicación del deseo informado y el utilitarismo de Griffin en los trabajos recientes de éste. El ejercicio comparativo también sugiere que en sus últimos escritos, la posición de Harsany se encontraba señaladamente cercana al punto de vista y en la línea de las explicaciones de comparaciones interpersonales que Griffin ha defendido recientemente. La discusión de Griffin y de Harsany está por lo tanto referida a los escritos de ciertos economistas contemporáneos (Ken Binmore, Partha Dasgupta y Amartya Sen).

Palabras Clave: deseo informado, utilitarismo, economía, Harsany, Griffin

* Aceptación: Mayo 2001
1 Introduction

James Griffin’s work in moral philosophy constitutes one of the most important attempts in recent times to think about a range of issues which fall at the borderline of moral philosophy and economics. There is a gap of ten years between the publication of the two books which, in large part, embody his contribution - Well-Being: Its Meaning, Measurement and Moral Importance (or Well-Being, for short), published in 1986 and Value Judgement: Improving Our Ethical Beliefs (or Value Judgement, for short)\(^1\). There is also a marked difference between these books, which is mirrored in the contrast between their cover illustrations. The cover of Well-Being features an etching by Rembrandt van Rijn of a student reading by candlelight in a dark room. On the cover of the second book, we find a painting by Gustave Caillebotte, of a man looking out onto a sunny cityscape, viewing it with some detachment. The contrast is, perhaps, illuminating.

In the first book, Griffin deals with some very difficult issues in moral theory, especially issues which are of central importance for utilitarianism. In the second book, Griffin takes a more detached look at the state of moral theories at the end of the twentieth century and suggests that we need to break away from the major traditions - utilitarianism, deontology and virtue ethics - and to make a new start.

Griffin undoubtedly makes a major shift away from his earlier position in his recent writings. The shift is, nonetheless, rather obscured by the continuity in his thought over the years. In the first two sections of this paper, I try to bring the change in Griffin’s position into focus by contrasting his earlier and later position with the views of one of the great economist-philosophers of modern times - John Harsanyi. Harsanyi’s account of utilitarianism remained fairly constant throughout his academic career. It represents a useful reference point in thinking about Griffin’s work. Apart from bringing out the shift in Griffin’s thought, this exercise brings out the continuity in Griffin’s writings, and the distinctive features of the project on which he has embarked in his more recent writings. It turns out, I suggest, that the later Harsanyi came close, in some respects, to the views which Griffin has recently articulated. The remainder of the paper explores some of the potential implications of Griffin’s work, particularly the more recent work, for economics. In exploring some of these implications, I discuss the work of some very well-known contemporary economists - including Ken Binmore, Partha Dasgupta and Amartya Sen - all of whom have evolved distinct positions at the fuzzy frontier between philosophy and economics.

2. HARSANYI’S UTILITARIANISM AND GRIFFIN’S WELL-BEING

James Griffin’s Well-Being, like John Rawls’s A Theory of Justice, has been widely read by both philosophers and economists. Both books also engage with a very influential account of utilitarianism, developed by John Harsanyi. I shall not go into the details of Harsanyi’s theory here, but merely describe it in broad brush terms. Harsanyi categorises “ethics” as one of three branches of the study of rational behaviour. The three branches are: 1) decision theory, the theory of rational behaviour under conditions of certainty and conditions of risk at the individual level; 2) game theory, the theory of rational behaviour where there are strategic interactions between individuals and; 3) ethics, which on his view “is a theory of rational behaviour in the service of the common interests of society as a whole”. In Harsanyi’s view all three theories “essentially have the same method” which is that of axiomatic derivation or the use of a “constructive decision model”. They begin, on his view, with a primary concept of rationality, on which they build further. Harsanyi’s vision here involves a very significant attempt at the unification of the three areas of the study of rational behaviour. He tells us that: “[t]he common method that these normative disciplines use represents a unique combination of philosophical analysis and mathematical reasoning”. Harsanyi’s work, thus, begins a tradition of what one might term “moral mathematics”.

There are three central planks of Harsanyi’s utilitarianism. Like any account of utilitarianism his requires: 1) an account of well-being; 2) an account of interpersonal comparisons of well-being and; 3) an account of the weight that is given to different people’s well-being in arriving at social or moral judgements. I shall focus here for the most part on a restatement of Harsanyi’s position in an essay entitled “Morality and the Theory of Rational Behaviour”. Harsanyi’s view, in that paper, is that the preferences which we should use in an account of utilitarianism are people’s “true

---

5 “Morality and the Theory of Rational Behaviour,” p. 43.
6 “Morality and the Theory of Rational Behaviour,” p. 44.
7 Around the same time that Harsanyi published the earlier versions of his theory, Kenneth Arrow also began a related line of research about the relation between individual and social value judgements, which also combined abstract philosophical argument with formal techniques and has the flavour of “moral mathematics”. See Kenneth J. Arrow, Collective Choice and Individual Values, New York: Wiley, 1951 as well as Amartya K. Sen, Collective Choice and Social Welfare, Amsterdarn: North Holland, 1979.
preferences”, preferences which someone “would have if he had all the relevant factual information, always reasoned with the greatest possible care and were in a state of mind most conducive to rational choice” ⁸. This position is not far from Griffin’s view in Well-Being. In that book, Griffin’s position is that well-being consists in the satisfaction of informed desires, where the criterion for a desire to count as informed is that it is formed with a proper appreciation of its object⁹. The information that is relevant to this account has to do with life plans, and such information is “full”—on Griffin’s account—when any more information will not advance a person’s life plan. The main advantage of this account is that it avoids certain well-known pitfalls associated with thinking about well-being in terms of people’s actual desires. Much later in his career, Harsanyi actually explicitly borrows Griffin’s language and talks of a person’s “informed preferences” which are the “hypothetical preferences he would have if he had all the relevant information and made full use of this information” ¹⁰. It is these preferences which Harsanyi thinks are relevant to welfare economics.

Griffin’s position in Well-Being does differ substantially from Harsanyi’s, nonetheless, since Griffin argues that we ought not to further restrict the kinds of desire which are used in moral theory¹¹. Harsanyi famously excludes anti-social preferences, involving sadism, envy, malice and resentment¹². This opens Harsanyi up to the criticism that he presupposes an account of morality, even at the stage where he is selecting the kinds of desire which are relevant to moral theory¹³. Harsanyi also makes a distinction between personal preferences—those which a person has in her ordinary behaviour—and moral preferences—which a person has when she takes a “special impersonal and impartial attitude”, in which she gives equal weight to all persons¹⁴. While Griffin does not follow Harsanyi in using this distinction—or in excluding anti-social preferences, they both disassociate the preferences that matter for moral theory from the desires people actually act according to.

Thus far I have been discussing well-being, and the desires which are appropriate to moral theory. If one has the account of well-being—or, more generally, of “advantage”—that is right for moral theory, one typically also

---

⁹ Well-Being, pp. 11-13.
¹¹ Well-Being, pp. 24-6.
needs an account of interpersonal comparisons of well-being or advantage. Harsanyi’s account of these is based on the notion of imaginative empathy. He supposes –like some others– that we can all think about each other’s situations and think about what it would be like to be in another person’s shoes. We can thus make comparisons of the following sort: would I prefer to be person a in situation x, or person b in situation y?15. This account involves approaching interpersonal comparisons cautiously, and reducing them, in some respects, to intra-personal comparisons. Given that we all have different preferences, this move does not overcome the problem that such an exercise might yield different results according to who actually engages in the exercise of imaginative empathy. Harsanyi’s attempt to deal with this complication invokes a similarity postulate, which states that once we allow for differences in people’s tastes and upbringing, their basic psychological reactions are the same. Their “extended preferences”, which are their preferences purified of the distortions of tastes and upbringing, are thus the same16. That is an important step in Harsanyi’s thought: if interpersonal comparisons of well-being are done in terms of such preferences it does not matter who does them.

In Well-Being Griffin rightly points out that when interpersonal comparisons are done in terms of “extended preferences”, they are hardly done in terms of the preferences of the person making the judgement17. Everything that is special to the particular person’s preferences has been removed. What Griffin does think is important in Harsanyi’s argument is the causal explanation of preferences. Starting from Harsanyi’s discussion, Griffin moves to his own view, which is that we have a picture of a distinctively human life and the components of a good life, which he calls “prudential values”. These values, Griffin thinks, make a human life go better, rather than being the sorts of things that make this or that life better. The genuine variation between persons, he thinks, comes from the specific way in which these values are manifest in lives –the specific ways in which values are realised– not in the profile of values. So, for example, if we take “enjoyment” to be a candidate prudential value, there are different ways in which enjoyment is realised in different people’s lives, because they enjoy different things. One person might enjoy crisps while another enjoys peanuts. The profile might also allow for considerable differences across ways of life and cultures, since there are different ways of enjoying oneself in different cultures. To make interpersonal comparisons, according to Griffin, we need, a profile of prudential values, as well as causal

17 Well-Being, p. 112.
knowledge about the world and information about specific people. The common ground, in this view of interpersonal comparisons, is the profile of prudential values. Unlike Harsanyi, Griffin does not need empathy in his account. However, clearly, as with the informed desire account itself, interpersonal comparisons require considerable information on Griffin's account of them.

As regards the leap from his views on well-being and interpersonal comparisons to moral theory, Harsanyi needs an account of the weights which are given to different people in arriving at social judgements. Harsanyi makes this leap by thinking of a situation in which people have an equal chance of being in any position in society, though they do not know which position they will occupy. They are then to choose between various societies, while supposing that they might occupy each position in society with equal probability. This is Harsanyi's well-known “equi-probability model of moral value judgements”18. The model is supposed to suppress self-interest and to generate impartial preferences. It is, of course, in some respects similar to (if distinct from) Rawls's “original position”. In Rawls' original position people (or their representatives) have no information about their chances of occupying different positions in society19. In Harsanyi's account, by contrast, probabilities are clearly assigned so that decision theory has some foothold. The unsurprising result is that in this situation agents choose to live in a society which maximizes average social well-being. Harsanyi goes on to consider the question of whether it is rule-utilitarianism or act-utilitarianism which ought to be preferred, and given his concern with strategic considerations, he falls on the side of rule-utilitarianism, largely because of the advantages of rule-utilitarianism in dealing with problems relating to coordination and cooperation20.

Griffin's position in Well-Being is more eclectic than Harsanyi's. While Griffin sees the strength of a moral principle which involves the maximization of social well-being, Griffin thinks of the moral point of view in terms of equal respect. It is this point of view which, he thinks, supports the idea of equal weight in the utilitarian maximization principle. Griffin thinks, nonetheless, that our moral intuitions are not always on the side of maximization. There are cases where the maximization principle seems to be less attractive than an egalitarian principle, and Griffin allows for the principle of maximization of social well-being to stand alongside egalitarian principles involving equal opportunities, or resources or welfare21. His position in Well-Being is pluralist both at the level of moral principles,

19 A Theory of Justice, p. 155.
as well as at the level of the values which constitute well-being. He admits that it is not clear if this position is correctly described as “utilitarian”\textsuperscript{22}. If one understands utilitarianism in this pluralist way, it is, perhaps, much more defensible than under various other guises.

2 Griffin’s Shift: Human Beings and The Search for A Morality Shaped for Them.

Since the publication of Well-Being, Griffin has turned away from many of the views he expressed in that book. The shift in his thought is, perhaps, most evident in a few papers published in the early nineties\textsuperscript{23}. In some respects, Value Judgement embodies the shift, though the continuity in his views is equally apparent. Harsanyi’s position, remained fairly stable over the years and since we know how close Griffin’s position in Well-Being was to Harsanyi’s rule-utilitarianism, comparing the later Griffin with Harsanyi allows us to see how far Griffin has travelled. It is also noticeable that in Value Judgement there is no reference to John Harsanyi. This is, in part, because, as I mentioned at the outset, the later book looks at moral theories with considerable detachment. However, it is partly due to the fact that much of what has changed in Griffin’s views of well-being and interpersonal comparisons is already clear in papers published between the two books –particularly in a very closely argued paper published in 1991 entitled “Against the Taste Model”– which covers some ground which Griffin does not go over again in Value Judgement.

In comparing Harsanyi with the later Griffin we can look again at three levels: the account of well-being; the account of interpersonal comparisons and; the account of morality. Firstly, consider the account of well-being. In his later work, Griffin argues that, if one takes the desire account to its logical conclusion, it leads one outside what he calls “the taste model”. The taste model sees something as valuable because it is desired. In Well-Being while Griffin does not defend the taste model\textsuperscript{24}, the term “informed desire account” does seem to put the emphasis on desire. The bridge between desires and what is good for a human being led Griffin and others in the direction of “rational” and “informed” desires\textsuperscript{25}. Nonetheless, in Well-Being,

\begin{itemize}
\item \textsuperscript{22} Well-Being, p. 212.
\item \textsuperscript{24} See, in particular, the discussion of the relationship between desire and value in \textit{Well-Being}, pp. 26-34.
\item \textsuperscript{25} The notion of “rational desire” was developed in Richard B. Brandt, “Rational Desires,” in R.B. Brandt, \textit{Morality, Utilitarianism and Rights}, Cambridge, U.K.: Cambridge University Press, 1992.
\end{itemize}
it is clear that what makes for a good life is that some value is realised when the informed desire is satisfied. Griffin hesitates over whether to label this position "the informed desire account" at all. In Value Judgement, Griffin is more explicit. He tells us that the fact that the "informed" or "rational desire" account "retains the word 'desire' does not show that much, or any, of the taste model is surviving." Even in the later work, nonetheless, Griffin does not actually want to draw value and desire too far apart. He simply wants to argue that when the desire account is developed in the natural way he earlier developed it, what makes something a component of well-being, or prudential value, is not the fact that it is an object of desire, but the fact that it makes a distinctively human life better. Even if one developed just the right formal account of the desires which connect with well-being, one would still only have a formal account. That would not get us very far, without an account of the substantive content of the good life. Since Griffin thinks that there are a number of values - such as enjoyment, deep personal relations and so on - which make lives better, his view is, then, a "list" view of well-being. In his recent replies to a number of essays, he is very explicit about this. He tells us: "my point in Well-Being - and more emphatically in Value Judgement - was that the informed desire account, fully worked out, becomes a list account." The list he actually gives includes values which are qualitatively quite distinct. The list includes: enjoyment; accomplishment; understanding; the components of a characteristically human life (minimum material provision, freedom from great pain and anxiety, basic capabilities, autonomy and liberty) and; deep personal relations. Griffin's more recent view is clearly better described as a list view than as a "desire account".

There are other reasons for rejecting the informed desire view. Some of these have to do with its rather stringent informational requirements. As we saw before, the informed desire account requires that people have all the information they need for the pursuit of their life plans. It requires that people have a proper appreciation of the object of desire. Yet, many of Griffin's recent essays focus on the limitations on human information and knowledge, and this is also clearly very relevant to the informed desire account. We rarely have all the information that is required by that account. Connie Rosati and David Sobel have made this point about "full information" views of well-being (and value). Since Harsanyi held to a

---

26 Well-Being, p. 34.
27 See Value Judgement, p. 23.
29 See Value Judgement, p. 80. Griffin does not insist that we all agree on this list.
view of this sort, this criticism also applies to his view. This is in some respects peculiar, since while Harsanyi is well-known for certain crucial contributions to the economics of incomplete information and the theory of games, he seems not to have appreciated the potentially serious implications of information limitations for his version of rule-utilitarianism.

When Griffin explicitly opts for a list view there is a shift in his view, though it is a very slight shift. The profile of prudential values which was already important in Well-Being remains central in his later work. The term “prudential value theory”, which is used in Well-Being, would perhaps have been a better label for the view he developed. On the other hand, while Harsanyi’s position remained fairly constant over the years, he nonetheless came closest to articulating a list view in his later writings. In discussing the good things in life or “prudential values”, he argues that all human beings have “much the same biological and psychological needs and therefore, have much the same basic desires for all these valuable things”. He also actually gives a list of such values which is clearly influenced by Griffin’s work. His list is: “[m]aterial comfort; physical security; freedom to control our lives; good health; a job suitable for our abilities and interests; deep personal relations in mutual love, in marriage, and in true friendship; to have children and be a good parent; to achieve better understanding of the world and of our place in the world; enjoyment of beauty in nature and art; to have worthwhile accomplishments of some kind; and to make our own behaviour consistent with our moral values”. Yet as he pursues this view, he refuses to accept any basis for such values except as objects of “basic desires”. What makes these values “intrinsically valuable to us is the fact that they are objects of our basic desires, which we largely share with other human beings, due to our common human nature and to our common biological and psychological needs”.

While Harsanyi appears to be groping to hold onto the taste model here, there is very little in his argument that distinguishes his view from Griffin’s list view. Clearly, the notion of basic desire that he is using would hardly be intelligible if there were not certain things which make a distinctively human life better. What makes a basic desire basic here is surely just that its object makes a distinctively human life better. That, I suggest, places Harsanyi outside the taste model.

31 Harsanyi won the Nobel prize in economics for these contributions. The July 2001 issue of the journal Games and Economic Behaviour is largely dedicated to Harsanyi and to the evaluation of his contributions.
32 See Well-Being, p. 72.
34 “Utilities, Preferences and Substantive Goods,” p. 139.
35 “Utilities, Preferences and Substantive Goods,” p. 139.
36 “Utilities, Preferences and Substantive Goods,” p. 140.
As regards interpersonal comparisons of well-being, Griffin’s views have not changed a great deal. It remains true that a profile of prudential values remains a central part of his account of how we do interpersonal comparisons. Griffin has, nonetheless, moved further in some ways from Harsanyi’s account. This is in part because of the very way in which Harsanyi approaches the problem. It was because of the so-called “problem of other minds” and the influence of logical positivism that many economists thought that interpersonal comparisons were meaningless, and thus, “unscientific.” Harsanyi’s contribution was remarkable given the prevailing views of economists at that time. He was adamant that “[o]ne would think that after so many years the time had come to escape the narrow confines of a long-obsolete logical-positivist orthodoxy and to have a fresh look at the problem of interpersonal utility comparisons.” He argued that there is sufficient common ground between human beings, and it is that ground which allows him to argue for the similarity postulate in developing his view of interpersonal comparisons. Nonetheless, by modelling interpersonal comparisons on intrapersonal comparisons, and adding the similarity postulate, Harsanyi’s position remains problematic. The most serious problem with Harsanyi’s discussion is that he fails to see something which, on Griffin’s later account, Ludwig Wittgenstein and Donald Davidson both saw. Both argued that in making sense of each others’ utterances we make certain sorts of interpersonal comparisons of desire and assume a background of, at least some, shared values. Interpersonal comparisons of desire and a bedrock of certain shared values (like the avoidance of pain) ground our mutual intelligibility as human beings. They are where we need to start in thinking about values, rather than where we end up. Griffin has pursued this line of thought, which fits well with his prudential value theory. This approach demystifies the problem of interpersonal comparisons as it appears in Harsanyi. The “jump” required in making interpersonal comparisons is not one we need to make through empathy, it is one we already make in making sense of each other.

Inasmuch as the views articulated in the later Harsanyi come close to a list view of the good, one can nonetheless see how Harsanyi’s views about interpersonal comparisons could have ended up rather closer to Griffin’s than they did, even in his late writings. By the time of the publication of

38 “Against the Taste Model,” pp. 54-56.
42 “Against the Taste Model,” pp. 51 and 57.
43 See, in particular, “Against the Taste Model,” p. 57.
"Utilities, Preferences and Substantive Goods", it is clear that Harsanyi’s claim about a “common humanity” is made in the context of the objects of basic desire. He thinks that the “extensive uniformity in our basic desires...is an interesting empirical fact about human nature, which seems to be of some importance for ethics”. In “Morality and the Theory of Rational Behaviour”, as we saw, the argument about a common humanity and an underlying uniformity in our basic reactions (when purified in various ways) is used in the defence of the similarity postulate, and is crucial to the way in which interpersonal comparisons are understood. Even in the later essay, Harsanyi retains arguments about empathy and extended preferences in his discussion of interpersonal comparisons. His later account has moved so far in the direction of a list view, however, that he hardly needs the account of extended preferences. In particular, he no longer needs extended preferences to find a level at which our preferences or desires are the much same. “Basic desires” provide such a level, and this is indeed important for ethics.

Griffin’s discussion of Harsanyi’s account of interpersonal comparisons also relates to, what he now thinks are, the limits of most contemporary moral theories. I shall limit myself to utilitarianism here. Griffin’s account of interpersonal comparisons involves the use of the profile of prudential values, in combination with information about particular people and knowledge of the world. If we had the right account of well-being and all the relevant information and calculative powers to make the relevant interpersonal comparisons, it may look as if there is no serious problem for utilitarianism. However, this is where Griffin’s shift is actually very marked in his recent writings. He has been insistent about the importance of the limits of human beings - which relate to our calculative abilities and information constraints. Informational constraints on human beings, of course, mean that there are serious limits to our capacity for making interpersonal comparisons of well-being. The idea that there might be full comparability, if there were perfect information and knowledge is irrelevant - limited information and knowledge are simply a fact of human life, and Griffin thinks that we need a morality which fits human beings as they are, taking account of the limits of their capacities. In a memorable line, he argues that “[m]oral norms must be made to fit the human moral torso”. If utilitarianism requires that the well-being of all members of society is comparable, and thus generates rather excessive informational requirements, then it will falter, simply because of this problem. Even if

46 “The Human Good and the Ambitions of Consequentialism,” p. 131. See also Value Judgement, p. 105.
it can produce an answer sometimes, it will say little about much of our moral lives. It will not be *imperfect* so much as rather *vacuous*.

On Griffin's views similar flaws are to be found elsewhere in the demands that utilitarianism makes on our capacities. If we follow Harsanyi and others, for example, and suppose that morality involves following the rules (motives etc.) that maximize total or average utility, there remains the question of how to work out which rules (motives etc.) maximize (average or total) social well-being. The limits of our knowledge and calculative capacities do, Griffin thinks, put that calculation beyond us. A similar point can be made about impartiality. Griffin's view is that, given the nature of the good life - which he thinks involves certain important commitments - we cannot be entirely impartial. Given the limits to our impartiality, our calculative powers and our knowledge, the principle of maximization of (average or total) social well-being may not provide much guidance. If it cannot give us much at all, it does not, in the end, provide us with a serious account of how human beings ought to act, or indeed, much of a criterion of right action. Griffin uses similar arguments against the other major traditions in moral theory - deontology and virtue ethics - which he thinks are often overambitious.

It must be clear, by now, quite how far Griffin has moved from his earlier views, and also, how far he has moved from Harsanyi's position. Harsanyi was, of course, aware of some of these problems, and it is worth dwelling for a moment on what he wrote. In "Morality and the Theory of Rational Behaviour" we find the following passage about imperfect information and interpersonal comparisons:

[I]f we have enough information about a given person, and make a real effort to attain an imaginative empathy with him, we can probably make reasonably good estimates of the utilities and disutilities he would obtain from various alternatives. But if we have little information about him, our estimates might be quite wrong.

In any case, utilitarian theory does not involve the assumption that people are very good at making interpersonal utility comparisons. It involves only the assumption that, in many cases, people have to make

---

47 It can be argued that it might nonetheless act as a "criterion" of right and wrong, even if it cannot play the role of a decision procedure. However, my feeling is that Griffin is right in thinking that it will fail on both counts. See *Value Judgement*, pp. 105-6. I have discussed this issue, and Griffin's views about the limits of human beings and the relevance of these for utilitarianism in my "Obligation, Human Frailty and Utilitarianism," *Utilitas*, 7, 1995, pp. 145-156.

48 *Value Judgement*, p. 106.

certain decisions - however badly they may make them. If I am trying to decide which member of my family is in greatest need of food, I may sometimes badly misjudge the situation. But I simply have to make some decision. I cannot let all members of my family go hungry because I have philosophical scruples about interpersonal comparisons and cannot make up my mind50.

The strength of the argument here is, surely, that morality requires that one act in a tragic situation, and that we must decide something or other, rather than fail to decide at all, when such indecision leads to worse consequences than any positive decision that is on offer. This is the well-known lesson of the tale of Buridan’s ass: the ass died because of its failure to choose between two bales of hay which were either as good as each other, or incomparable with each other, according to the particular version of the tale. That is a very general lesson and not specific to utilitarianism or interpersonal comparisons. It is not clear, however, that in the example Harsanyi gives, one ought to—or “has to”—make an interpersonal comparison when one cannot, given the limits of information. That suggestion would violate the principle of “ought implies can”: if the relevant person could not make the comparison in the relevant situation, then he cannot be blamed for failing to make it51. Rather he would need some way of deciding without making utility calculations. That would put him outside utilitarian rationality, rather than make him an imperfect utilitarian. Harsanyi’s purpose in the above passage is probably just to argue that we ought not to be put off by the fact that we cannot make perfect comparisons, because of imperfect information. Utilitarian rationality would merely be imperfect in that case. An imperfection in rationality is less troublesome, however, than a gap in rationality—or “incompleteness”—which is involved in a failure of our ability to make comparisons at all52.

Griffin’s shift leads him to articulate a general dissatisfaction with most contemporary moral theory. It leads him to develop his own proposal. He thinks we need a new way of thinking about morality, which takes the nature of human agents and their limitations seriously. That leads him to limit the role of rationality in moral theory. He tells us that we have to do

51 One needs to use the principle of “ought implies can” with some care. On this see my “Obligation, Human Frailty and Utilitarianism”.
52 The issues of comparability of well-being and comparability of prudential values are distinct. Griffin has tended to argue that there is very little, if any, genuine incomparability of prudential values. His best known discussion of the problem of incomparability is in chapter five of Well-Being. For a recent version of his views see: “Incommensurability: What’s the Problem?” in R. Chang (Ed), Incommensurability, Incomparability and Practical Reason, Cambridge, Mass., Harvard University Press, 1997. I have discussed Griffin’s views on a number of occasions. See, in particular, my ?Comparability of Values, Rough Equality and Vagueness: Griffin and Broome on Incommensurability?, Utilitas, 12, 2000, pp. 223-240.
without “the extensive background rationality that most utilitarians and some deontologists think is available to us\(^{53}\). However, even in the limited form in which he has articulated this proposal, there are some things that are clear at the outset. Firstly, Griffin thinks that moral philosophers have to be more modest - much that is relevant to the kinds of norms that human beings can live by may be found in other disciplines. Secondly, Griffin thinks that while some moral norms - such as a convention against torture - may relate directly to a prudential value, like the avoidance of pain, others may not. Others may answer to the “demands of social life” and require a concern with groups and strategic interactions\(^{54}\). Here, perhaps, we still have a link with Harsanyi, inasmuch as Harsanyi thought that game theoretic explanation has an important role to play in our understanding of morality. There are issues which remain about the role of traditional moral theories in this proposal. Are these moral theories entirely discredited by Griffin’s arguments and thus of no relevance (except inasmuch as they show us which paths in moral philosophy are largely fruitless) for his proposal? Or is it rather the case that their claims are exaggerated, and that they only account for some of our intuitions?\(^{55}\) I shall not pursue these questions here, but turn to the relevance of Griffin’s writings for contemporary economics, where his work has already had some influence.

3 Griffin’s Relevance to Some Contemporary Economist-Philosophers

Griffin’s influence on contemporary economics is most obvious in the writings of those economists who have, at various times, accepted something like the informed desire account. Harsanyi was, as we have seen, one such economist, though Harsanyi was also deeply influenced by Richard Brandt’s rational desire account. There are, however, many others\(^{56}\). Griffin’s influence and relevance is, nonetheless much broader and relates to a great deal more than the foundational issues about utilitarianism which have been discussed above. It is unsurprising, nonetheless, that it has been economists at the borderline of economics and philosophy who have been most influenced by Griffin. Inevitably, his influence has been deepest on those whose central concern has been well-being.

\(^{53}\) Value Judgement, p. 119.

\(^{54}\) Value Judgement, p. 93.


Griffin’s influence is, I think, evident in Partha Dasgupta’s *An Inquiry into Well-Being and Destitution*. His influence is nonetheless more limited than it appears at a first reading. Dasgupta’s work does not actually directly borrow Griffin’s prudential value theory. It involves a species of pluralism, as regards the components of well-being, which echoes Griffin’s views. One of the distinctive features of Dasgupta’s account is that he allows the “well-being functions” he works with to include information on preference satisfaction as well as “freedom”, where “freedom” is related to the size of a person’s “action set” (the set of actions she can choose from)\(^5^7\). This approach has the potential advantage that, by placing freedom at the heart of well-being, it defuses possible criticisms to the effect that an approach that focusses on well-being does not give much importance, or an intrinsic value, to freedom. As we saw, Griffin himself places liberty and autonomy at the heart of his list of prudential values. Unsurprisingly, these values are central to the account of human rights which he has recently developed\(^5^8\). By listing “enjoyment” – which is close to the hedonistic utilitarian notion of “utility” – alongside liberty and autonomy and other prudential values, Griffin makes room for both the traditional concerns of utilitarians, as well as the concerns of those, like Amartya Sen, who think that freedom is an important consideration in judging the quality of a person’s life\(^5^9\). Here Griffin and Dasgupta are close.

Another point on which Griffin and Dasgupta look close has to do with the arbitrariness of norms. Dasgupta explains certain norms in terms of self-fulfilling beliefs (what are called “bootstraps” equilibria in the economics literature). He uses this form of explanation, for example, in the context of choices about how many children parents choose to have\(^6^0\). The argument is, roughly, that if you like to do as others do, and if you believe that others will have lots of children, then you will also choose to have lots of children. This idea is modelled theoretically in the form of a game. The norm which results will be arbitrary in the sense that it depends on a belief about the average number of children in households, which is self-fulfilling: Griffin thinks that any account a morality which is shaped for human

---


\(^5^9\) Amartya, K. Sen, “Well-Being, Agency and Freedom: The Dewey Lectures 1984”, *Journal of Philosophy*, 82, 1985, pp. 189-221. Sen often distinguishes different aspects of freedom and, in particular, distinguishes opportunity and autonomy. Griffin also gives some importance to opportunity, particularly (as we saw) in the context of moral principles. In that context, he is willing to allow for a principle of equality of opportunity for well-being, which is close to Sen’s idea of capability equality.

\(^6^0\) *An Inquiry into Well-Being and Destitution*, chapters 12 and 12.
beings must allow for a certain amount of arbitrariness about norms\(^{61}\). However, the arbitrariness of norms in this part of Dasgupta’s book is of a different sort, since Dasgupta is arguing that there is scope to change or improve norms (say, by reducing average household size) on the grounds that the ability to change people’s beliefs itself might allow governments to alter norms. Griffin probably had something different in mind when he wrote about the arbitrariness of norms which fit the human frame. He probably had in mind something more like the arbitrariness which goes with *indeterminacy* about which norm is best. Where it is not clear what the right norm is we may simply have to go for some arbitrary solution. Dasgupta’s argument is, perhaps, more relevant to those parts of Griffin’s work where he focusses on issues of social cooperation and prisoner’s dilemma games.

There are many levels at which Dasgupta’s and Griffin’s writings are actually, nonetheless, rather far apart. Dasgupta treats preference satisfaction and freedom as distinct components of well-being. “Utility” is associated with the first of these. It is thus understood in a manner which is quite close to the classical utilitarian tradition - it is treated as a *substantive* value. This is a glaring difference between Griffin and Dasgupta, since in Griffin’s work desires and preferences are just used in a formal manner - all substantive values are *objects* of preference, and preference satisfaction is not itself a substantive value. There are further differences. Dasgupta uses his account of well-being to develop a contractualist account of morality, with a game theoretic foundation\(^{62}\). The “social well-being function” – the mathematical representation of the ranking of social states – he discusses is thought of as the outcome of a bargain. Dasgupta probably gives much more prominence to game theory than Griffin might wish for in developing his account of morality. While no doubt Griffin would agree with much of the discussion of norms in Dasgupta’s book, Griffin’s own recent writings give much more importance to particular prudential values in grounding certain norms.

In spite of Dasgupta’s bold steps in the direction of a new account of well-being and in spite of his attempt to merge the insights of his account of well-being with a contractualist view of morality, Dasgupta’s views mark only a limited departure from orthodox economics. The power of his *An Inquiry into Well-Being and Destitution*, relates both to the novelty of his account of well-being, and to the fact that he keeps many of the key insights of mainstream economics. It is a heady, if sometimes odd, combination. My suspicion is that the departure from the standard desire account (or “taste model”) in the later Griffin has deeper and, possibly,

\(^{61}\) Value Judgement, p. 96.

\(^{62}\) *An Inquiry into Well-Being and Destitution*, p. 65.
more serious consequences for economics. One important reason for this relates to the use of arguments relating to the efficiency of markets in normative economics, particularly the “fundamental theorem of welfare economics”\(^63\). These arguments often assume that people (or consumers) act according to their actual tastes or preferences—which are typically assumed to be directly related to welfare. Dasgupta treads carefully when discussing this issue\(^64\), but he fails to note that the notion of “preference” which is used in this debate is typically simply the notion which operates in the theory of consumer choice - it relates to the preferences on the basis of which people act. These clearly differ from any notion of preference which relates to a person’s interests. A close examination of philosophical debates relating to well-being, and of Griffin’s work in particular, must lead to a more serious reevaluation of orthodox normative economics. Dasgupta’s writings do nonetheless, sometimes arrive at an analysis which is fairly close to that which one might expect to arrive at by following Griffin’s own approach\(^65\).

Ken Binmore shares with Partha Dasgupta the desire to articulate an account of morality which is contractualist and game theoretic. Binmore’s project is far more influenced by John Harsanyi than Dasgupta’s is. The link with Harsanyi, nonetheless, allows us to see the relevance of Griffin’s writings. Binmore’s account, like Dasgupta’s, combines bargaining theory with an account of moral norms. It attempts to combine Harsanyi’s use of decision theory with the use of a “veil of ignorance”. The result, Binmore thinks, is a qualified vindication of Rawls’s *A Theory of Justice*, or at least of the well-known “difference principle”\(^66\).

Harsanyi’s influence on Binmore is most evident in Binmore’s use of “empathetic preferences”. Binmore writes that “[i]f an individual says that he would rather be one person under one set of circumstances rather than another person under a second set of circumstances, then he is expressing an empathetic preference”\(^67\). Binmore’s account of the social contract uses an original position in which the parties involved in bargaining choose according to their empathetic preferences. This move is very important in Binmore’s theory, because it is also assumed that people make interpersonal comparisons on the basis of empathetic preferences. This is entire-

---

\(^{63}\) The relevant theorem is sometimes treated as having two parts, while sometimes the two parts are treated as being different theorems.

\(^{64}\) *An Inquiry into Well-Being and Destitution*, pp. 170-1.

\(^{65}\) See, for example, the attempt to use prudential value theory in the context of well-being rankings in my “Pluralism and Well-Being Indices,” *World Development*, 25, 1997, pp. 2009-2026.


ly in keeping with Harsanyi’s account of interpersonal comparisons, and as with Harsanyi, there is a close relation, here, between interpersonal comparisons and intrapersonal comparisons. Finally, Binmore also employs the assumption that when they are flexible enough to change (over the “medium run” in his theory) people’s empathetic preferences are much the same. Rather than using a similarity postulate, or assuming anything like a “common human nature”, Binmore argues that people evolve in such a way that, over time, they end up with identical empathetic preferences. That allows him, like Harsanyi, to be unconcerned about the particular person whose preferences are used in making interpersonal comparisons. He then goes on to distinguish a number of the major positions in moral philosophy in terms of the distribution of “utilities” where “utilities” relate to (a mathematical representation of) empathetic preferences. One of the positions involved is, of course, utilitarianism. Yet, as Griffin remarks in his discussion of Harsanyi, for the extended, or empathetic, preferences, which are involved to be connected with well-being, they must be preferences of a certain sort, preferences which have to do with what is good for someone. Yet, Binmore does not restrict the nature of preferences in any way, at least as regards their content. We then have the following problem: the description of “utilitarianism” in Binmore—which involves adding up “utilities” may have little to do with utilitarianism, since the relevant “utilities” may have nothing to do with well-being. This does not, of course, necessarily undermine Binmore’s key arguments. Rather it reminds us that we must be careful in interpreting Binmore’s views, and, in particular, his characterisation of “utilitarianism”.

The economist-philosopher I want to end with is Amartya Sen. I shall have the least to say about Sen, in part because so much has been written about him already. The two aspects of his work I want to focus on are his critique of utilitarianism, and his account of the quality of life. Both aspects are central to his writings over the last two decades. Sen breaks down utilitarianism into three parts: 1) welfarism—the view that the only information that matters to moral judgements is “utility” information (however that is defined); 2) consequentialism—the view that all the information that is relevant to moral judgements relates to consequences, broadly defined and; 3) sum-ranking, the view that to make moral judgements we must rank states of affairs in terms of the sum of utility in each state of affairs. This is, of course, really only one way of characterising utilitarianism. Sen’s critiques of utilitarianism have been levelled at 1) and

68 “Against the Taste Model”, p. 56.
3. Sen has argued that consequentialism - or at least a focus on consequences - can be understood in such a way that the chief critiques with utilitarianism do not undermine it. Yet, Griffin’s arguments about the limitations of human beings, and about information and knowledge constraints are, mostly, arguments against consequentialism. That is, they relate to our ability to work out what actions, rules or motives would lead to the best consequences, as much as to the actions, rules etc. which would maximize the sum of utility. So Sen and Griffin have quite different reasons for rejecting utilitarianism.

Sen’s defence of a focus on consequences does, nonetheless, seem to allow for some of the problems involved in Griffin’s discussion. Sen allows for “bounded rationality”, especially where there are problems with respect to ranking, which arise from epistemic limitations. Sen has recently argued that we should characterize rationality in terms of what he calls “maximization”, which allows people to act when information constraints or ignorance makes it difficult or impossible to rank alternatives. I am not sure that this move actually goes far enough in terms of recasting human agents in economics and game theory. It merely allows for the possibility that humans can act rationally even when they cannot rank alternatives. It allows an agent to choose any alternative, as long as there is no other option which is positively better than that alternative. That means that we can choose rationally when there is non-comparability. This is a departure from anything like standard consequentialist rationality.

In developing his own “capability view” of the quality of life, Sen suggests that we should focus on a person’s ability, opportunity or “positive freedom” to lead the life she values. Sen has characterised this in terms of a person’s capability to “achieve” (or realise) the (valuable) “beings and doings”—or “functionings”—which constitute a flourishing life. He has actually stopped short of giving a full list of the valuable functionings which constitute flourishing lives. He has hesitated because he thinks that people tend to value different things, and because there are likely to be limits to the extent of consensus about the objects of value. His concern with consensus derives, no doubt, from the influence of John Rawls and the Rawlsian account of primary goods (i.e. all purpose means). Rawls and others argue that the stuff of good lives cannot form the basis of any agree-

72 There is, of course, a large literature in economics on “bounded rationality”, which is related to Griffin’s project. See Herbert Simon, 1952, Models of Bounded Rationality, Cambridge, Mass., MIT Press.
ment and so focus instead on the means to lead good lives\textsuperscript{74}. Sen’s view, on the other hand, is that we ought to focus on the ability to lead valuable lives\textsuperscript{75}. This might be intrinsically valuable to people, whatever their views of the good life.

Yet, in applying the capability view, at the level of comparisons of the quality of life, between nations, regions or persons, the capability theorist needs some account of what valuable “functionings”, or components of the good life, we ought to focus on. Accounts like Griffin’s prudential value list view which explicitly focus on the components of well-being\textsuperscript{76}, are thus relevant to those who want to apply anything like a capability view. Inevitably, Griffin’s list has become a contender amongst the various lists of the stuff of good lives, which are increasingly being discussed in the debate surrounding Sen and others (notably Martha Nussbaum)\textsuperscript{77}.

The importance of Griffin’s work in this context derives from the fact that he has always stressed the importance of deliberation about substantive values. Griffin thinks that we need to understand this kind of deliberation better. Some economists and philosophers have, in strong contrast, been more concerned with the formal properties, or formal analysis, of well-being, or the good, without being too interested in what well-being actually consists in\textsuperscript{78}. Griffin’s writings suggest that we cannot say much about well-being, without thinking about what it consists in. This enduring concern with prudential deliberation may, in the end, be Griffin’s most important contribution to the literature on well-being.

4 Conclusions.

When one looks at the writings of James Griffin and John Harsanyi over the years, it becomes clear that as they develop the informed desire account, they move very far from the standard desire account. While Griffin has explicitly abandoned the taste model, Harsanyi clung onto something like the informed preference view to the end of his life. Nonetheless,

\textsuperscript{74} I discuss this further in my “Needs, Incommensurability and Well-Being.”

\textsuperscript{75} Griffin’s recent writings suggest of course that it is in part the limits of our capabilities that define our humanity. This point is relevant to capability theory, especially when it combined with a maximisation view of rationality which allows for bounds to rationality (as Sen’s view does).

\textsuperscript{76} John Finnis’s position also falls into this category. See John Finnis, Natural Law and Natural Rights, Oxford: Oxford University Press, 1979.

\textsuperscript{77} Nussbaum’s list of human capabilities has been an important contribution to this debate. Her latest version of this list is in Women and Human Development, pp. 77-80. I discuss her list, and other such lists, in relation to prudential value lists at greater length in my “Development, Common Foes and Shared Values,” forthcoming, Review of Political Economy, 2002. My own list is most fully articulated in my “The Concept of Well-Being,” Economics and Philosophy, 14, 1998, pp. 51-73. See also Sabina Alkire’s discussion in her Valuing Freedoms, forthcoming, Oxford: Oxford University Press, 2002.

Harsanyi’s position at the end of his life is perhaps better understood as a list view. I have suggested that the later Griffin has moved quite far from his earlier writings, and that the shift is much clearer when we compare his later position to Harsanyi’s account of utilitarianism. While Harsanyi’s later work still insists on empathy as the ground for interpersonal comparisons, his account of basic desires provides such a ground, without any need for empathy and extended preferences. In Griffin’s more recent writings, some shared values are a prerequisite of our mutual understanding and language, while for the later Harsanyi what we have in common—our basic desires—remains, chiefly, a matter of empirical observation. Inevitably, since most economists have used a desire account of well-being, the writings of James Griffin and John Harsanyi remain very relevant both to those who stick to this account as well as to those who have attempted to develop alternative ways of thinking about well-being.