Impartiality, Neutrality and Justice
Re-reading Brian Barry's *Justice as Impartiality*

THEORIES OF DISTRIBUTIVE JUSTICE

There have been three main competing normative political theories in our time: libertarianism, justice as fairness and utilitarianism. Brian Barry proposes a revision or variant of justice as fairness, which he calls justice as impartiality. There is a sense in which he is not proposing a new theory but is rather cleaning up the analysis and vocabulary of what he takes ‘to be the standard contemporary liberal theory’ (p. 124). The cleaning up, however, is substantial. There are three central elements of Barry’s theory that I wish to discuss. The first and simplest is that it is a second-order theory about institutions for achieving distributions for individuals, not a first-order theory for individual choice or individual outcomes. The second and third are that the content of its recommendations depends on impartiality and on reasonable agreement, terms that have manifold meanings.

Because of its general importance to virtually all discussions of justice, I will briefly consider the nature of second-order theories immediately below. Then, after canvassing some of the background of Barry’s contribution to theories of justice, I will discuss what he takes to be his main alternative category of theories: mutual advantage theories. Finally, I will turn to the two relatively novel claims of his own theory: impartiality and reasonable agreement. Almost all the

---

* This paper began as a commentary on Brian Barry’s *Justice as Impartiality*, Oxford, 1995, given at the American Political Science Association, San Francisco, 30 August 1996. I am grateful to the participants on that panel, including Brian Barry, and to Chandran Kukathas for discussions of an early draft.

play in Barry's account turns on that glorious word, reasonable. Surely all theorists think they are reasonable. Presumably, we may infer from Barry's persuasive redefinition of that term that most of them are, in some sense, unreasonable. I should perhaps note up front that I am, in his sense, an apparently unreasonable utilitarian.

SECOND-ORDER THEORIES

The central insight in second-order theories of the foundations of social institutions is that of Rawls in 'Two Concepts of Rules'. In this paper, Rawls wrote to defend utilitarianism as a two-stage theory against certain silly objections once popular. (Alas, in a discipline in which nothing is ever finally learned, these objections are still popular.) First we establish a principle of what outcome we want to achieve and then we design an institution that will help us achieve it best. Those with roles in the relevant institutions act according to the rules of their roles, not directly according to the supposed utility of their actions. At the level of the social actor who is not a role-holder in a particular institution we might recast this problem as one of resolving a collective action dilemma or a co-ordination problem. Each of us wants the generally good result but none of us might be motivated to act so as to achieve it, either by altering a co-ordination or by contributing to a collective provision. Our institution constrains or motivates our behaviour in relevant ways so that we do achieve it.

Utilitarianism is a two-level theory: both first and second order, that is to say, both moral and political. As a political theory it is a second-order theory in the sense in which impartiality and justice as fairness are second-order theories. In a second-order theory, the result we wish to achieve on the ground in the lives of individuals (the first order) determines the nature of the institutions for organizing society (the second order). As a political theory utilitarianism prescribes institutions or constraints on institutions. Utilitarianism is additionally a first-order theory in the strong sense that it directly prescribes ranges of behaviour for individuals. The chief difference between the second-

---

2 For example, that a sheriff should participate in or allow an unjust lynching of an innocent person if that lynching would calm a mob and prevent it from doing even worse things. The conclusion of this silly morality tale is supposedly that utilitarianism violates justice.
3 For one view of utilitarianism as a political theory, see Russell Hardin, Morality Within the Limits of Reason, Chicago, 1968, chs 3 and 4. It is sometimes supposed that if everyone acted as a utilitarian there would be no need for institutions. This supposition is probably false because, for example, co-ordination problems often require institutional resolution. But, in any case, the point is irrelevant for practical life because we cannot expect everyone to be moral.
4 Theories of Justice

In proposing a theory of justice grounded in impartiality Barry's purpose is, like that of Rawls, to construct a theory of justice as fairness with a strong element of mutual advantage. In Theories of Justice, the first volume of his larger Treatise, Barry canvasses these two main grounds, fairness and mutual advantage, on which contemporary theories of justice are based.

Much of the fair division literature is about how self-interested players can achieve fair results. Barry discussed this issue at compelling length in Theories of Justice. This is a problem for a Rawlsian theory of justice only in the design of the institutions that are to implement it. It is not the problem of establishing the principles of justice that are to govern the purpose of those institutions. What we need at this prior stage is a principle of fairness per se. The fair division literature generally takes for granted what this principle is, since that literature generally deals in interpersonally comparable utilities of something nearly indistinguishable from what in ordinary parlance is called money. A fair division of x utiles or dollars among n persons is, ceteris paribus, x/n utiles per person. In the standard fair division literature, what is to be distributed is determined ex ante, so that the distribution is the only issue at stake.

In a real society, in which most of what is to be distributed must be produced, a rigorous fairness theory that required equality of distributions might run up against deep problems of motivation for those who produce what is to be distributed. Hence enforced egalitarianism might mean that overall production would be severely depressed, so that all would be equal but, alas, equally miserable. Somehow, then,
if incentives for production matter, we must consider introducing inequalities that lead to greater overall welfare, as in a mutual advantage theory.

Thomas Hobbes is the first great theorist of mutual advantage. Because his concern was only with order and welfare and not with justice, he was not interested in a normative notion of mutual advantage, in which essentially all are better off with respect to some alternative. He often wrote as though essentially all were better off with any order than with disorder, but he acknowledged that religious fanatics and glory seekers (certain aristocrats) might not be, so that his mutual advantage theory was essentially sociological rather than merely normative. David Gauthier’s theory of morals by agreement is largely a mutual advantage theory, but Gauthier requires supplementary moral principles to make his system work.6 Libertarian theories might often be cast as mutual advantage theories, although they are often grounded in supposed natural rights to property. A strictly ordinal utilitarianism is a mutual advantage theory. But, as such, it is inherently incomplete in what it can specify as the utilitarian outcomes because ordinal welfare cannot be aggregated in determinate ways.7 It shares this problem with the criterion of Pareto improvement, in which any change in allocations must make at least one person better off while not making anyone worse off. Often there are many such changes that could be made, but some are better for me while others are better for you, and neither the Pareto criterion nor ordinal utilitarianism can help us choose which of these to make.

The recent turn to resourcism in discussions of distributive justice may be motivated by an effort to reduce our problem to a similarly simple prima facie move to divide up something that is easily measured, as we may suppose that, at a first approximation, dollars are the measure of resources. A game theoretic account of mutual advantage shows the resourcist hope not to be generally coherent. There are, however, circumstances in which it may be straightforwardly meaningful. By somewhat tendentiously stipulating that some goods do not weigh in our welfare until other goods have been counted first, Rawls makes his Difference Principle sound like a principle of mutual advantage. But it is not really that, not least because mutual advantage is inherently indeterminate. It is his lexicographic invocation of egalitarianism that yields determinacy, at least in principle. Hence, his theory is fundamentally dependent on both mutual advantage, or welfarist, and on fairness considerations, although

many commentators and Rawls himself often refer to it as merely a fairness theory.

Resourcist theories seem to be inherently fairness theories. Contractarian theories are usually mutual advantage theories. The two larger classes generally do not fit well together. Yet Rawls has remarkably brought the two classes of theories together in one theory of distributive justice that seems to be reasonably coherent. Barry wishes to do the same under the rubric of reasonable agreement and impartiality.

Let us summarize and highlight the distinctive differences between these two kinds of theory. Fairness theories generally apply to distribution of a fixed sum once we have produced or come upon something, or, conversely, to a fixed burden such as the risks of military service, or to division of something that comes from a fixed production function that is not in itself a function of the nature of the division. Fairness therefore applies when one may reasonably suppose production is not at issue. If production is at issue, incentive effects foul ordinary analyses of fairness because they effect what there is to be distributed. Mutual advantage theories are inherently concerned with how we motivate production of whatever we are to distribute. One might think these two simply do not belong together, as traditionally they have not gone together.

Rawls’s theory is relatively original in bringing mutual advantage and fairness together. (This is a central thesis of Barry’s Theories of Justice.)8 He does not do so as some contemporary pluralists might, first on the one hand and then on the other. Rather he integrates them by placing them at different points in the deduction of his eventual distributive principles or, rather, his principles for the basic structure of a just society. Mutual advantage might be defended purely pragmatically, as it is by Hobbes. But it is a purely co-ordination criterion because, as in a Pareto allocation, there are many ways we could go, all of which would be mutually advantageous to some status quo. Rawls grounds the selection from these mutually advantageous states on their relative fairness. This is one of the few fundamentally original moves in Anglo-Saxon political theory since Hobbes.

Note, however, that in his theory Rawls is concerned with fairness of an odd kind. Fairness in the allocation of the joint social product is an issue just because that product is joint. I might claim to ‘deserve’ a fair part of it because I contributed to its production. But Rawls does not want a simple desert model in which I get what I deserve as a result of my effort or whatever. If market wages and profits really mirrored contributions, allocations according to desert would reduce to little

---

6 Gauthier, Morals by Agreement.
7 Hardin, Morality Within the Limits of Reason, ch. 4.
8 Brian Barry, Theories of Justice, Berkeley, CA, 1989.
more than what markets do. Rawls clearly thinks market wages and profits do not mirror contributions. We cannot causally relate your effort to your share of the social product. Your wages are related to the supply and demand functions for your talents, rather than directly to your output. What you produce may stay constant while your wage changes, or vice versa. Rawls's theory of justice therefore generally entails a distribution other than what the market would produce.

Yet, his theory does not simply correct for the distortions that supply and demand might impose on your wage. Rawls does not tie your desert to what you produce. He ties it merely to the fact that you produce. The fact that you produce gives you a claim on a share of the joint product. The capitalist or well-paid professional may think she deserves what she gets from market relations. A Rawlsian and virtually any contemporary economist would reject such desert claims and would say that the capitalist or professional is merely in part very lucky to be in the right place at the right time — as, for example, the child of an immigrant to Silicon Valley might prosper radically better than her cousin back in Bangladesh. As John Stuart Mill noted, 'As civilization advances, every person becomes dependent, for more and more of what most nearly concerns him, not upon his own exertions, but upon the general arrangements of society.' Bill Gates is not in any plausibly meaningful sense worth the four billion dollars a year that he has recently been making, but that is what our (somewhat distorted) market yields him. Still, Rawls's theory retains an odd tie to desert or entitlement. Barry wants to break this tie with his theory of impartiality. Those who like desert talk may say he wants to base desert in simple humanity, not in any specific capacity or accomplishment. But in any standard view, Barry's theory, unlike that of Rawls, is freed of any connection to desert as merit of some kind.

Again, a central insight of Rawls is that fairness cannot stand alone as a theory of the good or the right because it is fully consistent with egalitarian misery. Inequality that produces greater overall welfare trumps pure equality under some circumstances. Those who dislike this assertion sometimes argue against it with an implicit dismissal of the possibility that inequality could have this effect or with claims that all we need is to correct aberrant psychologies to get people to produce for the general good rather than merely for selfish benefit (as in the theory of new Soviet man). Barry has his feet planted in the real world where one cannot sustain these claims even for those who are productive but especially not for those who are incapacitated in various ways.

**Mutual Advantage**

Although he gives specific attention to deontological libertarianism that is grounded in principles of rights, especially in the variant of Robert Nozick (pp. 202–5),[11] Barry's main alternative theories of justice are variants of mutual advantage: justice as reciprocity, which has a contractarian ring, and Hobbesian or Humean justice. The latter could be conceived as straight mutual advantage in a merely sociological and, hence, non-normative sense. Sociological mutual advantage is arguably the best going theory of actual liberal governments historically, because the way a typical democratic society is organized can be characterized as mutual advantage sociologically rather than normatively.[12] In comparison to substantial alternative ways of organizing these societies, it is advantageous for important, which is to say politically efficacious, major groups in the society.

Once this organization of the society is in place, it then is also advantageous for many or even all groups that might prefer a quite different organization. It is advantageous for these groups in the sociological sense that there is no reorganization they could cause that, taken together with the costs to them of changing to that alternative organization, would make them better off. This is essentially Hobbes's defence of any extant government for virtually all citizens.[13] As what Barry calls a proto-utilitarian, Hume makes mutual advantage a normative theory, because he supposes government arranged to support mutual advantage will produce good results.

Mutual advantage theories are commonly focused on what Sidgwick derides as justice as 'order'.[14] Some libertarians, Hobbes, Hume and Adam Smith seem to think the main function — or, for the libertarians, the only function — of government is maintaining a system of law and order. Adam Smith asserted that only the appropriate function of justice is 'commutative'.[15] But the sometime reason for this focus, as in the arguments of Hobbes most specifically and of many libertarians, is that justice as order is the institutional guarantor of mutual advan-

---

[12] This is a central argument of Russell Hardin, Liberalism, Constitutionalism and Democracy, Oxford, forthcoming.
tage. In this sense, justice as order is a second-order theory. Its point is not the achievement of justice in your relationship with me through, for example, forcing you to restore what you have taken from me. Moreover, the proponents of such a theory are generally not concerned with the rightness or wrongness of individual actions (although deontological libertarians generally are). Indeed, Hobbes famously argued that there could be no meaning of right and wrong without a government that has a system to enforce law.

One could, of course, suppose that our only purpose in commutative justice is that of righting or punishing wrongs. And one might want a system of law to accomplish this first-order purpose merely for the pragmatic reason that there is no other plausible way to accomplish it well. But Hobbes, Hume and Smith want a system of law and order for the second-order reason that it structures the kind of society they thought we must want—a society graced with stable expectations and commitments that make for prosperity and, in Hume’s and Smith’s eighteenth-century vocabulary, contentment.

Let us turn now to the motivational structure of mutual advantage theories. They typically are grounded in motivations of self-interest on the part of citizens and are therefore not normative for citizens in their own personal actions. Note the striking difference with justice as reciprocity, which requires, in addition to calculation of what serves mutual advantage, a motivation beyond merely self-interest: concern not to be unfair (p. 112). Once we settle on justice as reciprocity, either by contract or by convention, we must then act normatively against our own interests to defend the position of those too weak to defend their own interests (pp. 48–51). Barry cites the example of American Indians during the nineteenth century. Whites repeatedly made treaties when peace was advantageous to overcome Indian disruptions and then broke them when Indian capacities faltered (p. 45n). For Hobbesian mutual advantage theory, the weak do better under a regime that largely ignores them than they would do if the regime failed to maintain order. Therefore, they have an interest in being obedient to the regime even though no one else is motivated to defend the interests of the weak.

One might object that the difference between justice as reciprocity and justice as impartiality is not motivationally very significant, that, indeed, it is plainly a matter of philosophical or conceptual interest. The difference is this. Justice as reciprocity requires deduction of what system of rules we should have from the interests of all concerned, from nothing more than what would serve mutual advantage if those rules are instantiated. Justice as impartiality requires deduction of the rules of justice from what reasonable agreement (motivated by fairness) would entail. The latter is an internally consistent theory because the actual backing of the system of justice that it defines would follow adequately from the same motivation, namely fairness, as that from which it was deduced. The former, justice as reciprocity, is internally inconsistent in that it requires for its maintenance a different motivation, fairness, from that, namely interest, from which its structure is determined.

Does this difference in internal consistency of motivations matter substantially? Perhaps the answer is a matter of taste. Of course, philosophical taste is not a concern of much practical significance. This is a fact to be applauded by those who consider very extensively the range of philosophical tastes in vogue historically. It may be especially applauded by those appalled by much of contemporary philosophical taste, with its nihilistic, elitist bent and its vocabulary that oddly defies understanding by anyone other than sycophantic insiders even while it is anti-intellectual. Should those concerned with justice on the ground be bothered by the difference between the two classes of theory? Only if they have significantly different implications for the actual structure of our institutions and the actual content of our policies. Do they? Unfortunately, this question is very hard to answer. Its answer turns heavily on the implications of “reasonable agreement.” It is a striking omission of the work of Thomas Scanlon, Barry and others who argue for reasonable agreement as an apparent criterion for choosing structures of justice that they do not either define the term or even very much constrain it.

Note, incidentally, that many of us actually even do act from Barry’s supposedly incoherent pair of motivations. We support tax and other law changes that would make us pay more to enable welfare transfers. But we also figure our own taxes sharply, striving to pay no more than the law mandates. Is this inconsistent? Well, no. We have many quite different motivations for the actions we take in our lives. Of course, we also have many ends, most of which compete with each other for our time and resources. Any theory that supposes this is inconsistent should give us hard argument for its implausible claim.

Also note that, as part of his defence of the superiority of justice as impartiality over mutual advantage because of the greater coherence of its motivational structure, Barry appeals to the game-theoretic representations of the two theories. He says mutual advantage has the structure of a prisoner’s dilemma while justice as impartiality has the structure of an assurance game (p. 51). The best individual outcome for a large-number prisoner’s dilemma is to have all others co-operate while the lone individual defects. The best individual outcome for an assurance game is the same as the best outcome for the collective. This

---

18 Hardin, Morality Within the Limits of Reason, pp. 100–5.
discussion is slightly odd, however, because the ordinary game theoretic structure of an assurance game is that, in it, while players act from the usual game-theoretic motivation (self-interest), they produce a generally good overall result. In an assurance game, ordinary self-interest produces an optimal outcome; in a prisoner’s dilemma, self-interest produces a suboptimal outcome. It is only with a very different, not a standard game-theoretic, motivation that justice as impartiality is an assurance game.

In the end, the coherence of motivations in the theory of justice as impartiality is what we might call sector-coherence. I am not to have coherent motivations across all sectors of my life—this would be perverse, because I should, for a trivial example, play to win at chess or poker if these endeavours are to be worth the effort and, contrariwise, I should probably act with great altruism in familial and other relationships of love (perhaps you with your values should not, but I with my values certainly should). Rather, I am only to have coherent motivations in the sector of my life concerned with social justice, in which my single motivation is ‘to behave fairly’ (p. 51). Just what is the force of this concern? Why not let motivations be relatively coherent or even incoherent in similar ways across all sectors of my actual behaviour while I plump for institutions that align those motivations with just and otherwise good outcomes? This is, of course, Hume’s vision of the roles of government and citizenship.

The main complaint by Hobbes, Hume and many other theorists against a theory such as reciprocity or Barry’s impartiality is that it depends on a motivation that is, though perhaps not rare, nevertheless too limited to underwrite government. The pragmatic appeal of Hobbesian and Humean theory is not grounded in a nice philosophical concern with the tastefulness of internal coherence. It is grounded in the deep conviction that such motivations of morality as altruism, fairness or even justice somehow defined simply will not carry the day in real life. Barry, Rawls and others speak of the extent to which a just social structure will lead to different motivations in citizens over time. No doubt, there are many considerations that explain the failure of the theory of the new Soviet man during the seven decades of Communist rule in the Soviet Union. But even the most optimistic backers of such a theory must count that seven-decade experience as fundamentally distressing. It seems unlikely that any contemporary philosopher with a normative theory of justice that depends on richly normative motivations on the part of all or virtually all citizens could genuinely believe the theory would ever work on the ground.

Let us stipulate that an explicitly contractarian version of justice as reciprocity is not a compelling theory for the reason that it seems implausible that contractarian argument works at all, either practically or normatively. The central question for reasonable agreement is therefore whether it would lead to institutions that, in their working, are distinctively different from institutions created by either a theory or a politics of mutual advantage. The crux of this question is arguably whether there would be a difference in the treatment of a stable minority whose interests differ substantially from those of some majority. Given that Barry and other reasonable agreement theorists seem to agree with at least Madisonian mutual advantage theorists that our institutions will be roughly democratic and majoritarian, we have to ask whether we can count on majorities to protect the interests of minorities as defined in an initial mutual advantage or fairness vision.

**IMPARTIALITY**

Barry’s principle of impartiality can be used to argue against three positions that might, in different ways, be called partial: solipsism, the group level analogue of solipsism or group solipsism, and any particular theory of the good. No one supposes a solipsist account of justice is credible, reasonable or compelling. Barry also rejects, virtually by analogy, any group solipsist account (p. 8). In the face of the contemporary onslaught against universalism, impartial theories, this is, of course, a crucial move. As it happens, this move is conceptually related to Barry’s claims about neutrality toward variant conceptions of the good. Let me briefly discuss these two matters: group solipsism and commitment to a particular (hence, partial, not in the sense of incomplete but in the sense of biased) theory of the good.

**Anti-Solipsism**

First, note that justice, unlike a theory of the good, seems to be an inherently social notion. Hence, it is incoherent even to think it could be solipsist, whereas a theory of the good could be (egoism is one of Sidgwick’s methods of ethics). But it is not analogously incoherent to think justice could be group-specific. To assert impartiality over group specific norms seem ex cathedra, just as to assert the specialness of my group seems ex cathedra. There may be considerable force to the more general position of those who think abstractly. There may contrariwise be considerable force to the group-specific vision for those who think more concretely and pragmatically. The contemporary communitarian claim that there is no ground on which to stand to weigh the one position against the other is therefore not silly. Reason is the slave of the passions, Hume said. You have a passion for abstract visions; I have a passion for my tribe. My reason, therefore, disagrees with yours.

---

17 Hardin, *Liberalism, Constitutionalism and Democracy*, ch. 4.
and there appears to be no knock-down argument either of us can deploy against the other. Or, at least, no such argument has made it into currency.

Note that I think communitarianism has reprehensible implications that cannot be handled from within the communitarian vision but that must be addressed only from a more universal perspective. The woeful implications of communitarianism are, of course, not a matter of logic and they are not of the essence of communitarianism. Rather, they are in its all too often realized potential to be reinforced with norms of exclusion, often to the point of violence against those outside the community. This goes both ways, of course. As Murray Kempton remarks, a community is where you leave it you get beat up. That suggests the good side of community. It is unfortunately implicit in Kempton’s quip that community is also a place where when you enter it as an outsider you get beat up. On its bad side community has led to the violent persecution and deaths of tens of millions of people in only the twentieth century and to the moral and social suppression of plausibly hundreds of millions.

The usual defence of communitarians to such complaints is that there is also a good side to community, as in good communities. To critics of communitarianism this sounds like the trivially true and interesting claim that good community is good. The thesis that good community is good does not merit the extraordinary outpouring of blissful paens to community that have cluttered journals and bookshelves in the past decade or so, many of those paens vigorously directed against the universalism of Rawls, Kant and utilitarianism. It does not seem tendentious to read much of this odd literature as holding community good per se. But I do not have an argument against this view that should convince anyone who continues to assert the inherent goodness of the communitarian vision even after seeing the evidence of Rwanda, Bosnia, the Ku Klux Klan, the Nazis and various other group-solipsist, communally driven groups and populations. In so far as it presumes that community is inherently good, however, I think that the current vogue of communitarian thinking and argument is morally incomprehensible and pernicious.

Theories of the Good

Now turn to specific theories of the good and their breach of impartiality. With the caveat noted below, reasonable agreement makes some sense as a political theory. It makes much less sense as a moral theory. It makes sense as a political theory for the contingent reason that it seems likely that pushing for a particular theory of the good would often not actually put that theory into political practice and hence would not bring about that good. A utilitarian or an autonomy theorist might readily therefore think it more likely to enhance utility or autonomy to follow something like the Scanlon and Barry reasonable-agreement approach in constructing a constitution for a society. But then, of course, the reasonable-agreement or fairness theory does not trump utilitarianism or autonomy but may merely be a practical means toward instantiating as much as practically possible of either of these.

And this is the caveat on reasonable agreement: what is ‘reasonable’ to me, with my theory of the good, is how much of that good I can get out of our political arrangements. How could I plausibly say my theory of the good should be trumped by a theory of mere fairness? I might, however, have no option to seeing it trumped by an array of contrary political forces. This is a pragmatic, not a moral matter.

One might suppose that my sticking to my theory of the good come what may is a variant of solipsism. To paraphrase G. K. Chesterton’s criticism of patriotism, it is a defence of my theory drunk or sober just because it is, after all, mine. But that is wrong. I back that theory because I think it is compelling and correct, not because it is merely mine. The connection between my thinking it right and its being mine is merely logical, not solipsist or otherwise psychological. My willingness to compromise my theory of the good in political decisions is not a choice grounded in the alternative morality of fairness but is a pragmatic choice grounded in my theory of the good. We could properly rewrite the title of one of Rawls’s papers from ‘Justice as Fairness: Political not Metaphysical’ to the title of this essay ‘Reasonable Agreement: Political not Normative’.

A political theory of reasonable agreement is an analogue of Locke’s argument for religious toleration. Locke did not assert fairness as a prior or trumping principle against any particular religious value. Rather, he argued for accommodation as a practical antecedent to achieving any religious value. It is the practical consideration of opposing forces that makes agreement to less than one’s full theory of the good reasonable. Let us turn to Barry’s reasonable agreement.

reasonable agreement

Barry writes that ‘A theory of justice which makes it turn on the terms of reasonable agreement I call a theory of justice as impartiality’ (p. 7). The terminology is slightly confusing, as virtually all terminology is in political theory, because there are theories other than his, Rawls’s and Scanlon’s that are also grounded in impartiality (or, in a slight twist in vocabulary, impartialism): Kantian and utilitarian theories (p. 232). What Barry’s formulation of his theory of justice urgently demands, then, is an account of reasonable agreement that lets us know how
that theory differs from those other impartial theories. The difference is somehow in that term, reasonable agreement, because there must be a normative term somewhere in the account or, by Hume’s structures on deducing values from mere facts, we are cheating ourselves into believing values reside in nature. We cannot reach a normatively grounded principle of reasonable agreement without imposing some normative principles. But agreement is merely a factual matter when it happens. Hence, to paraphrase a quip of William Godwin’s that Barry uses, we must unpack the magic in the adjective reasonable (p. 223). In that term lies all the good of justice as impartiality.

The starting point of the account of reasonable agreement is Scanlon’s formulation of his contractualist variant of contractarianism. Hume contributed a theory of co-ordination mostly in a series of long footnotes. A pithy statement of Scanlon’s central contribution to political theory is one footnote. The full relevant text is: ‘On this view [Scanlon’s contractualism] what is fundamental to morality is the desire for reasonable agreement, not the pursuit of mutual advantage.’

Merely pragmatic principles will not do for the reasonable-agreement crowd because they sound too much like mutual advantage (p. 193). This is arguably the implicit reason many critics of liberalism assert that the reasonable-agreement programme is inherently not value neutral. In Scanlon’s concern for reasonable agreement, the condition is negative: I must accept any arrangement that I cannot reasonably reject. Barry — rightly, I think — rejects this view of the condition as merely negative (pp. 69–70). Either way, the condition is a co-ordination condition of reasonable agreement. As Barry writes, ‘the theory of justice as impartiality will rarely prescribe a unique outcome’ (p. 197). It was once an aspiration of utilitarians, such as Bentham, to reach definitive conclusions. Those who have understood the complications and difficulties of welfarist value theory in its development over the past two centuries can no longer ground such an aspiration, and utilitarianism, especially in its ordinal mutual advantage form, similarly sets little more than co-ordination conditions.

Reasonable agreement seems to be a matter of judgement. It is not Kantian rationalist deduction from transcendental principles. It is perhaps rather like Aristotelian or Millian judgment that depends on a broad range of experience and analysis. That makes it relatively unsatisfactory for a newcomer to the enterprise. One might even

---


---

question whether the theory is impartial to theorists (admittedly not a group whose welfare is of great concern), some of whom are evidently qualified to apply the term and some of whom are not qualified. Presumably, for example, Nozick thinks (or once thought) his libertarian principle reasonable. I most certainly think my utilitarian principles to be as reasonable as any other principles in moral or political argument, as Barry (and even Rawls) perhaps once did also. Evidently, Nozick and I lack relevant judgement. Part of what Barry means by reasonable is fair. But both libertarians and utilitarians can make sensible claims for the fairness of their visions, although they would not usually start from fairness as their first principle.

Rawls, perhaps uniquely among major theorists of our time, openly starts from two first principles: mutual advantage and fairness. There is debate over whether the one or the other of these is evidently prior morally, although, since both are necessary, priority is an odd concern. (Both RNA and DNA are necessary for human life. Which is prior?) Utilitarians start from the single value of welfare, although the singularity of this value is problematic if there are multiple contributing factors to welfare, as there must be. Libertarians typically start from a putatively single principle of something like pre-Adamic or natural rights, although the singularity is clouded by the sheer number of these rights, which libertarians toss out with seeming abandon. If there is more than one non-redundant value or right, then there must be the logical possibility of conflict between them, which is a difficult problem for multi-principle theories, a problem that Rawls attempts to avoid. Rawls’s attempt conspicuously fails in a way fully analogous to the apparent failure of singularity in utilitarianism because his category of basic goods includes multiple goods and lacks a functional form for combining them.

Barry is less worried to work out this issue of the multiplicity of underlying concerns in his theory, although his accounts of the specific political policies here (and presumably also in the promised third volume of *A Treatise of Social Justice*) suggest that he too has problems of multiplicity in the underlying considerations to which reasonableness leads us. In fact, any theory that does not have such problems is probably derelict and worthy of no attention. More fundamentally, Barry’s project is specifically motivated by the multiplicity of extant conceptions of the good and by the concern to erect a system of justice that does not impose one such conception on everyone. But in this aspect, his theory is directed at an outward multiplicity of values. The difficult issue, as with utilitarian, libertarian and Rawlsian theories, is the internal multiplicity of considerations. Until we have more of a grasp on this issue, we cannot confidently claim to know the magic of the adjective reasonable.
In conclusion, consider two examples of what one might call inferences from reasonableness, two examples that must be contested and that therefore want further discussion from reasonable agreement theorists. The first example is essentially about a conceptual problem. The second is about the content of agreement even among those who share Barry's programme. We would need to know more about the meaning of 'reasonable' if we are to resolve these and related issues.

First, the conceptual problem. Barry argues against what Bentham called the pre-Adamical rights that are assumed in most libertarian theory (p. 125) and he generally rejects at least Nozickian libertarianism as outside the range of reasonable agreement (pp. 200–4). The latter move is perhaps the severest test he offers of the notion of reasonable agreement. Many serious philosophers and others would counter that the libertarian vision is eminently reasonable. They might readily concede that others would not agree with that vision. Mill's claims against entitlement might seem to knock the stilts from under libertarian rights. But one might argue for them on grounds that they are the best way to go rather than that they are 'natural'. They could be the best way to go because they serve mutual advantage or because they serve autonomy. Nozickian libertarians seem to think they serve autonomy or even that they are defined by autonomy. How is autonomy only a specific theory of the good rather than a relational notion comparable to fairness? The point of autonomy seems typically to be to let individuals choose their own good just as Barry's impartiality with respect to conceptions of the good is intended to do.

Second, the content of reasonable agreement. It is a peculiarity of Rawls's theory that the welfare component of his value theory has a remarkable shape. Initial income matters a great deal but after some level further income matters not at all. The curve that relates welfare to income flattens out. (Further income might matter causally if it affects other components of the value theory, such as equal political voice. But it does not matter as a contribution to further value.) The implication of this is that the staggering inequalities among, say, the upper eighty per cent of people in American society are of no concern to Rawlsian distributive justice. It is only the inequality of these eighty per cent as a class and the remainder of the population — some of whom suffer lacks of some of Rawls's basic goods — that is unjust. Is this a conclusion on which we can reasonably agree? We do not know from their writings what Scanlon and Barry might say; Rawls evidently has spoken.

CONCLUDING REMARKS

Return for a moment to the brief discussion of the priority of principles when there are more than one. Some readers of Rawls suppose that fairness is morally prior to mutual advantage. Yet many theorists make mutual advantage prior. One can understand why an anarchist or one who is committed to self-interest as the right motivation would elevate mutual advantage above fairness. The move that some contemporary contractarians make is to give concern with coercion moral priority over concern with welfare and other plausible principles. They generally do not face the full implications of this move for priority. Does a government act unjustly if it coercively organizes production and distribution of food and other necessities in the face of grim poverty and failures of collective action? Or is the only morally correct response for such a government to let the society and its people be destroyed? Such conditions of destitution fall outside the realm of application of Rawls's theory, as he explicitly asserts. This means, of course, that his principles of justice are not prior to all other considerations.

Hume and various writers before him recognized the artificiality of various principles of justice in the sense of their being constructions that apply to particular human circumstances as opposed to a priori or natural principles that apply in all circumstances. In correctly asserting that justice is an artificial virtue in this sense Hume offended his lazier readers, who were offended with the notion that a virtue could be artificial rather than absolute (and perhaps directly apprehended by those benighted souls who know such things). Barry grants that his own theory is artificial in this sense as well and he lays out the circumstances of impartiality that make the theory workable. Is there a prior moral position we can take in judging this theory? For a utilitarian theory of anything, the prior theory is clear enough. Because utilitarianism is a full two-level theory, we can at least rough out the structure of its second-order political theory from its first-order moral theory. For justice as impartiality, we jump into the fray in medias res and try to analyze justice under certain circumstances. That makes it a strangely partial theory. This conclusion is, seemingly, a demand of reasonableness.