

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

1

An Introduction to *Justice, Political Liberalism,*
and Utilitarianism

Marc Fleurbaey, Maurice Salles, and John A. Weymark

The opposition between utilitarianism and liberal egalitarianism has triggered the most important developments in political philosophy in the twentieth century and has had a considerable effect on other subjects as well, such as law and economics. The turn of the new century has witnessed the death of two prominent scholars in these debates, John Harsanyi and John Rawls. Harsanyi and Rawls have undoubtedly been the leading figures in each of these schools of thought in recent decades. Building on the work of classical utilitarians such as Jeremy Bentham and John Stuart Mill, Harsanyi has provided decision-theoretic foundations for utilitarianism that have served as the touchstone for Rawls's own critique of utilitarianism. Rawls believes that utilitarianism fails to satisfy Immanuel Kant's maxim that individuals should be treated as ends in and of themselves, not just as means for promoting the social good. Drawing inspiration from the writings of social contract theorists such as John Locke and Jean-Jacques Rousseau, Rawls has fashioned a modern statement of liberal egalitarian principles for the design of the basic institutions of society that respect Kant's maxim.

The writings of Harsanyi and Rawls offer vigorous defenses of their theories, which their lively exchanges have done much to illuminate. Their theories draw on and provide support for widely shared values. Their contributions have been, and will continue to be, inspirational for scholars and others who seek to understand what social justice and ethical behavior require. The voluminous literature that has responded to Harsanyi's and Rawls's writings has drawn out many of the implications of their theories, has clarified and refined their most convincing arguments, and has pointed out ambiguities and weaknesses in their reasoning. Whether the divide between

We are grateful to Peter Vallentyne for his comments on an earlier version of this chapter.

1

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

these two schools of thought will eventually yield a convergence toward a consensual theory or whether it will be agreed that these theories are based on irreducibly opposed principles is premature to say. But the opposition itself is most useful in order to help us appreciate the difficulties of formulating a coherent account of social justice and social welfare. Furthermore, the ongoing debate between utilitarians and liberal egalitarians may simply reflect a deeper truth that it is utopian to believe that consensus can ever be reached on such fundamental social issues. As Alan Dershowitz has so eloquently said, “There are few, if any, moral truths (beyond meaningless platitudes) that have been accepted in all times and places. The active and never-ending processes of moralizing, truth searching, and justice seeking are far superior to the passive acceptance of one truth. . . . Conflicting moralities serve as checks against the tyranny of singular truths” (Dershowitz, 2006, p. 94).

The general principles of utilitarianism and liberal egalitarianism are not the only source of inspiration for recent reflections that have drawn on the work of Harsanyi and Rawls. The particulars of their theories have been challenging and illuminating in different ways. For example, Harsanyi’s theorems about utilitarianism have raised new questions about the relevance of individual preferences under uncertainty, and their representation with particular utility functions, for the evaluation of social states. Similarly, Rawls’s contractarian defense of his version of liberal egalitarianism and his frequent reference to the mutual gains of social cooperation has engaged theorists attracted by contractarian approaches to further explore what constitutes social justice, even though they have often ended up far from liberal egalitarian conclusions.

Harsanyi and Rawls deserve our most respectful tribute for their fundamental contributions to utilitarianism and liberal egalitarianism, respectively, and, more generally, for helping to bring questions of social justice to the fore after many years of relative neglect. The chapters in this volume honor Harsanyi and Rawls by investigating themes that figure prominently in their writings. In some cases, the chapters that follow take stock of what has been learned by exploring issues considered by Harsanyi and Rawls in more depth and from novel perspectives. However, much as it is valuable to understand and compare their theories, new approaches to dealing with problems of social justice have commanded attention in recent years. Many of the contributors to this volume use the work of Harsanyi and Rawls as points of departure for pursuing the construction of new theories for the evaluation of social justice and injustice.

In this chapter, we introduce the chapters in this collection and place them in the context of the literature. To address the questions considered in

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

subsequent chapters, some authors have employed mathematical arguments, which may be a formidable barrier for some readers with limited mathematical training. Nevertheless, the lessons learned from these “mathematical” articles are important both for evaluating what Harsanyi and Rawls have accomplished and for understanding some of the new approaches that their writings have inspired. Accordingly, we have attempted to provide relatively nontechnical (but not completely nonmathematical) summaries of them here, even though this results in the mathematical chapters receiving more attention than some of the more widely accessible chapters.

1.1 Themes from Rawls

Rawls’s theory of social justice is multifaceted, with different elements of the theory cohering in a complex way. His ideas have been challenged in many ways, even by those who are in broad agreement with his approach to questions of justice. The chapters in Part 1 of this volume consider three issues related to the work of Rawls: (i) his use of absolute priority rules, (ii) the role that merit and personal responsibility play in his theory of justice, and (iii) the role that moral intuitions should play in justifying ethical beliefs.¹ Echoing the vast literature devoted to analyzing Rawls’s writings, these contributions range from critiques of some fundamental features of his theory to proposals for amending his ideas so as to overcome some of its shortcomings.

The main features of Rawls’s theory of justice, first set out in detail in *A Theory of Justice* (Rawls, 1971) and further developed in *Political Liberalism* (Rawls, 1993) and *Justice as Fairness: A Restatement* (Rawls, 2001), can be briefly summarized. Rawls’s theory of justice focuses on the basic structure of social institutions, requiring them to be organized in such a way as to favor the most destitute members of society, that is, those individuals who have the fewest basic social goods (what Rawls calls primary goods) such as rights and liberties, power and opportunities, self-respect, and income and wealth. Moreover, Rawls considers equal liberties and a fair equality of opportunities (in particular, for access to careers and related advantages) to be of paramount importance, so that the right to equal liberties and equal opportunities for all should have priority over the advancement of the socio-economic status of the poor, with priority in turn given to liberties over equality of opportunity. These principles were originally developed for a society that shares similar moral values, but in his later

¹ Additional Rawlsian themes are considered elsewhere in this volume.

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

writings, Rawls has defended them as a political conception of justice (hence the name, “political liberalism”) for a pluralistic society whose members can endorse his principles even though they differ fundamentally in their religious, philosophical, and moral beliefs. In this revised account of his theory of justice, Rawls places greater emphasis on distinguishing political liberties from other kinds of liberties and requires that they be of comparable worth in the sense that everyone has an equal opportunity to hold political office and to influence the decisions made in the political sphere.²

Rawls’s focus on the basic structure of society and on primary goods is related to the liberal features of his theory. Specifically, Rawls regards society as having the duty to provide everyone with a fair share of resources and opportunities. However, society does not have the right to interfere with private uses of these resources that result from personal conceptions of what constitutes a good life, which each member of society is free to develop and revise as he wishes. A liberal society should not dictate to its members how life must be lived, and from this, Rawls derives the far-reaching implication that social institutions should not in any way refer to particular conceptions of the good life and, therefore, should not try to compare individual success by any metric of the good that would involve such a conception.

1.1.1 Harsanyi on Rawls

In Chapter 2, the late John Harsanyi, pursues and, sadly, closes a debate with Rawls that began in Harsanyi (1975) and Rawls (1974). The initial debate focused on the appropriate specification of an impartial perspective, what Rawls calls an original position, from which an impartial observer (to use Harsanyi’s terminology) identifies the basic features of their theories.³ The use of this device plays a substantial justificatory role in the work of both Harsanyi and Rawls, and it raises fundamental issues about what constitutes rational criteria for impartial decision making. More will be said about original position arguments in the next section.

Here, Harsanyi focuses on two other issues. Echoing the earlier exchange, he first objects to the absolute priority accorded to the worst-off social group, or maximin principle, that is embraced by Rawls. With this principle, one social arrangement is judged to be better than a second if the situation of the

² For precise statements of Rawls’s principles, see Rawls (1971, section 46) and Rawls (1993, pp. 5–6).

³ Strictly speaking, Rawls views his original position as a forum for multiperson bargaining, but, as we shall see in Section 1.2.2, it can also be viewed in terms of a single impartial observer choosing principles of justice.

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

worst off in the first alternative is better than the situation of the worst off in the second. For Rawls, an individual's situation is assessed by the expected value of an index of primary goods holdings over his lifetime.⁴ More so than in his earlier critique of Rawls, Harsanyi questions the policy implications at the social level of the maximin criterion. According to Harsanyi, such an extreme criterion can only have extreme social consequences in terms of redistributive policies, possibly triggering a civil war. Although Harsanyi does not mention it, the maximin criterion has often been employed in the analysis of optimal redistributive taxation (see, e.g., Atkinson, 1973, 1995; Choné and Laroque, 2005) without inducing extreme consequences because incentive constraints preserve the interests of the wealthy members of society better than an army. This example, however, does not rule out the possibility of extreme consequences with other policies.

Harsanyi also critically examines other absolute priorities that are granted, in Rawls's approach, to certain primary goods over others, in particular liberties over socioeconomic advantage. Harsanyi argues that, in general, individuals acting collectively employ finite trade-offs, whether between social groups or between goods, although he does allow for the absolute preeminence of certain values over others (e.g., moral duties over nonmoral interests). How reasonable it is to give absolute priority to the worse off over the better off, or to liberties over socioeconomic status, is likely to remain a contested topic for some time. Following the publication of Rawls's *A Theory of Justice*, the maximin criterion attracted considerable interest and, in some cases, support from welfare economists. Interest in giving priority to the worst off has ebbed and flowed over the years, but it appears to have enjoyed a recent resurgence (see, e.g., Maniquet and Sprumont, 2004; Tungodden, 2000). Nevertheless, Harsanyi's skepticism about absolute priorities and his preference for middle-ground criteria are very natural and widely shared.

⁴ Rawls refers to this principle as the "difference principle," but the term "maximin" is both more standard and a more transparent label for a principle that seeks to maximize the smallest value of some attribute. For example, the maximin principle applied to utility first identifies the utility of the person with the smallest utility in each of a set of distributions of individual utilities and then chooses a distribution for which this utility value is largest. With the lexicographic version of Rawls's maximin principle, leximin for short, the ranking of two alternatives is determined by the value of the index of primary goods of the worst-off group for which the value of this index is not the same in the two alternatives. With both the maximin and leximin principles, groups are defined by their ranks (in terms of primary goods holdings), not by the names of the individuals that comprise them. Although it is in this lexicographic form that the maximin principle is typically employed, for simplicity, we ignore this refinement in the subsequent discussion.

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

The second set of considerations raised by Harsanyi has to do with Rawls's complex treatment of the notions of merit and personal responsibility in his theory of justice. As previously discussed, Rawls advocates a division of labor between social institutions and individuals. The former provide resources and opportunities, the latter are responsible for how these resources and opportunities are used to pursue their own conceptions of what makes a life valuable. It is not the job of social institutions to track individual merit and responsibility and to reward them accordingly because institutions only have to take care of the distribution of fair shares of resources. One then obtains the somewhat paradoxical situation in Rawls's theory of a concept of responsibility that plays a key role in delineating the limited role of social distribution, while being totally absent from the principles of distribution themselves. Only incentive considerations can justify differential rewards in this view. Rawls (1971, section 48) famously argues that even effort cannot be the moral basis for superior claims over resources because the propensity to work hard is largely inherited and is nurtured in a favorable family environment. Both of these contingencies are morally arbitrary.

In a discussion that appears to be based on commonsense morality as much as on utilitarian maximization of the social good, Harsanyi opposes this view. He argues that it is essential that society publicly recognizes the intrinsic and social value of (i) moral behavior that results from a good character and (ii) the development and employment of talents for the common good even if these talents and characters are partly inherited or nurtured by a caring family. Otherwise, human excellence cannot flourish. Harsanyi abhors the vision of a society in which all kinds of moral characters would be considered equally nonpraiseworthy. He understands Rawls's view as connected to hard determinism, a doctrine that denies the existence of free will and responsibility on the assumption that all causal laws are deterministic. Harsanyi defends a compatibilist approach to the problem of free will, that is, a view that accepts physical determinism but nonetheless carves a place for personal responsibility and moral praise and blame. In this view, while an individual's moral attitudes are heredity and environment dependent, nevertheless, they are subject to choice and, therefore, his actions are subject to moral commendation or discredit.

1.1.2 Liberal Egalitarian Approaches to Personal Responsibility

The issue of individual responsibility has been the subject of much attention in the philosophical literature since the publication of *A Theory of Justice*, most notably by Arneson (1989), Cohen (1989), and Dworkin (1981, 2000).

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

Each of these scholars has proposed variants of liberal egalitarianism that put personal responsibility at the core of the definition of individual advantage.

Dworkin has not diverged much from Rawls's view that individuals should be held responsible for their conceptions of the good life and their ambitions, but, unlike Rawls, he believes that individuals are responsible for certain kinds of unlucky outcomes. Specifically, Dworkin distinguishes between option luck, which is concerned with the outcomes of deliberate risky choices, and brute luck, which is not. He holds individuals responsible for a bad outcome in the former (given fair initial conditions), but not the latter, case because the adverse consequences of a deliberate risky choice may be mitigated by purchasing insurance, and, if this is not possible, individuals could have refrained from making a risky decision in the first place.

Arneson and Cohen fully embrace the commonsense view that individuals can be held responsible only for what lies within their genuine control, and define social justice in terms of equal opportunities in a radical sense – equality of what is under the control of individuals to achieve.⁵ They differ on how the metric of achievement should be defined for comparisons across individuals, but these differences need not concern us here.⁶

These developments have had a substantial impact on welfare economics, where studies of freedom and opportunities have flourished in recent years. An important strand of this literature is concerned with the fair distribution of resources and opportunities when account is taken of the responsibility individuals have for making choices. Fleurbaey (1998), for example, distinguishes between the objective of neutralizing the effects of factors outside of an individual's control and the various possible objectives that may be adopted to reward an individual's exercise of responsibility. See also Roemer (1998) and the contributions in Laslier, Fleurbaey, Gravel, and Trannoy (1998) for further explorations of these and related issues.

1.1.3 Arneson on Personal Responsibility

In Chapter 3, Richard Arneson undertakes a detailed analysis of the role that personal responsibility plays in Rawls's theory of justice in light of

⁵ Rawls's notion of equal opportunities is closer to the ordinary sense of the term, namely, nondiscrimination in the access to positions of authority and responsibility.

⁶ The way that Arneson and Cohen view achievement has a close affinity to Sen's theory of capabilities (see Sen, 1985, 1992). While Sen's theory is primarily defended by him in terms of the freedom to choose between alternative options (e.g., lives), factors that individuals cannot be held accountable for also play a fundamental role in assessing their circumstances. See, for example, Sen (1992, section 5.3).

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

these more recent developments. This analysis illuminates the evolution of Arneson's thought from his theory of equal opportunities (Arneson, 1989) to his desert-sensitive theory of justice (Arneson, 2000).

Arneson observes that Rawls walks on a tightrope because he denies that social institutions should be devoted to rewarding the deserving, while retaining a key role in his theory for individual agency and responsibility. Unlike Harsanyi, Arneson does not interpret Rawls as endorsing hard determinism, nor does he believe that it is necessary to determine the extent to which a person has a free will or is responsible for the outcomes of his actions in real-world situations before any conclusions can be reached on principles of justice. Arneson takes as his starting point "the limiting principle that we should be held responsible at most for what lies within our power to control" (p. 98) and argues that independently of all conceptual and practical difficulties in the definition and measurement of control, it is worth pursuing the ethical analysis of ideal principles in order to derive their consequences before any consideration of practical implementation is raised. In this view, the free will problem is concerned with implementation, not principles, because society often lacks the information needed to assess personal responsibility.

Arneson identifies two main shortcomings with Rawls's account of the role that merit and deservingness play in the design of institutions that shape the distribution of resources. First, he argues that Rawls's use of the maximin principle, with its emphasis on the maximization of the level of primary goods that the worst-off group in society can be expected to enjoy over their lifetimes, fails to distinguish correctly between inequalities that are a matter of choice and those that are not. For example, future wealth (one of the primary goods) depends on one's choice of employment. Second, he argues that Rawls's dismissal of the view that benefits and burdens should be distributed in proportion to moral worth because moral worth cannot be defined independently of the content of the norms of justice is based on a false premise. For Arneson, desert does have an independent specification.

Arneson is not content to show the failings of Rawls's theory compared with a more refined desert-based principle of responsibility. He also suggests a possible amendment to Rawls's theory in which the maximin principle is expressed in terms of the expected potential lifetime holdings of primary goods, not the expected value of their actual holdings, at the onset of adulthood. If individuals do not achieve their potentials, that is their responsibility. As Arneson notes, this proposal has much in common with Dworkin's views discussed earlier. However, he ultimately concludes that this move

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

is not successful because it fails to consider that the ability to make good choices and stick to them is an unchosen characteristic.

After Arneson considers and rebuts some variants of the thesis that the choice of ends individuals make and their consequences are not matters of justice, provided some threshold of rationality is attained, he critically examines the responsibility principle to determine how one could obtain a reasonably acceptable notion of justice that provides a role for both individual and social responsibility in the choice of ends. Arneson, like Rawls, believes that individuals should be responsible for their freely chosen ends, but he also believes that society should undertake any measures that can help improve the quality of people's responsible choices. He concludes that justice requires allocating resources at the onset of adulthood to maximize the effective opportunities for well-being of the most disadvantaged, where effective well-being is measured in terms of the well-being that could be obtained if an individual acts as prudently as one could reasonably expect. Justice does not demand compensation for bad consequences of a rational choice, but it may be required if the individual making a choice is not completely responsible, which would be the case if society failed to provide an adequate environment to nurture the ability to make prudent choices.

Arneson also examines whether there is a case for assigning liabilities for an adverse outcome in a way that diverges from the costs that individuals are responsible for and that may depend on factors outside their control. In other words, are there circumstances in which individuals should be asked to share in the costs of events or decisions for which they are not strictly responsible? Such sharing is a common response to a natural catastrophe and can be justified by the obligation we all have to compensate the affected individuals for events outside their control. It can also be envisaged for the costs of responsible decisions made by individuals if pooling the costs promotes general fairness and efficiency better than a fine-grained sorting out of personal responsibilities.

In going beyond the pure opportunity-based approach to resource allocation by, in some circumstances, considering the consequences of poor decision making when determining whether compensation is merited, Arneson seeks to avoid the common criticism that the pure opportunity approach to compensation can be too unforgiving to individuals who suffer misfortunes apparently as a result of their own choices. This line of reasoning opens the way for the desert-catering prioritarian theory that he develops in Arneson (2000). In this theory, a failure to seize opportunities reduces the moral value of providing a compensatory benefit to the concerned individual, but does not necessarily nullify it.

Cambridge University Press

978-0-521-64093-0 - Justice, Political Liberalism, and Utilitarianism: Themes from Harsanyi and Rawls

Edited by Marc Fleurbaey, Maurice Salles and John A. Weymark

Excerpt

[More information](#)

1.1.4 Griffin on Moral Intuition

Turning now from issues of priority and responsibility, James Griffin, in Chapter 4, raises a third set of issues dealing with the role that moral intuitions play in justifying a set of ethical beliefs. Rawls's arguments in defense of his principles of justice often refer to intuition about what seems reasonable. He refrains from using the common strategy employed in political philosophy of testing general principles by artfully conceived, but sometimes contrived, examples that appeal to our intuitions. Instead, he proposes the concept of a reflective equilibrium (Rawls, 1971, sections 4 and 9) as the archetype of the support that a normative theory should seek to obtain from its double confrontation with reasoning and intuition. A reflective equilibrium in favor of normative principles (of justice or of morality) occurs when abstract analysis (for instance, a description of an impartial observer's reasoning) yields conclusions that fully agree with one's well-considered judgments, that is, with the normative beliefs that one would hold once one's initial beliefs have been revised after having considered alternative normative principles. In Rawls's case, his objective is to justify his three principles of justice and the priority accorded by them of first securing equal liberties for all, secondly providing fair equality of opportunity, and, finally, maximizing the expected holdings of primary goods of the least advantaged. Rawls's hope was, of course, that his theory of justice, with its combination of reasoning about what principles of justice would be agreed to in his original position and its appeal to the intuitive reasonableness of these principles and their priority ranking, would produce such an equilibrium.

Griffin is concerned with the justification of ethical beliefs in general, not simply the justification of normative principles of justice. He is refreshingly critical about the piecemeal approach to ethical reasoning based on intuitive consideration of hypothetical examples, arguing that this procedure gives too much weight to intuition. He also dismisses the other extreme of deriving substantive moral principles without any appeal to intuition as being unsuccessful.

Griffin considers the lessons to be learned for moral reasoning from the coherence theory of justification found in the natural sciences in some detail. As in an ethical theory, a scientific theory distinguishes good from bad beliefs and tries to make them cohere. In the natural sciences, it is empirical observations, the inferences that can be made from them, and how well these observations and inferences describe how the world works that separate credible beliefs from ones that are not.