Studies in Choice and Welfare

Marc Fleurbaey
Maurice Salles *Editors*

Conversations on Social Choice and Welfare Theory - Vol. 1

Springer

Studies in Choice and Welfare

More information about this series at http://www.springer.com/series/6869

Marc Fleurbaey · Maurice Salles
Editors

Conversations on Social Choice and Welfare Theory - Vol. 1

Springer

Editors
Marc Fleurbaey
Paris School of Economics
Paris, France

Maurice Salles
University of Caen-Normandy
Caen, France

ISSN 1614-0311 ISSN 2197-8530 (electronic)
Studies in Choice and Welfare
ISBN 978-3-030-62768-3 ISBN 978-3-030-62769-0 (eBook)
https://doi.org/10.1007/978-3-030-62769-0

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Contents

Contributors

Matthew D. Adler Duke Law School, Durham, NC, USA

Nick Baigent London School of Economics, London, UK

Taradas Bandyopadhyay University of California Riverside, Riverside, USA

Carmen Beviá Universidad de Alicante, San Vicente del Raspeig, Alicante, Spain

Felix Bierbrauer Center for Macroeconomic Research, University of Cologne, Cologne, Germany

Walter Bossert CIREQ, Montreal, Canada

Richard Bradley London School of Economics and Political Science, London, UK

Steven Brams New York University, New York, USA

Christopher P. Chambers Georgetown University, Washington, DC, USA

Youngsub Chun Seoul National University, Seoul, South Korea

Claude d'Aspremont Center for Operations Research and Econometrics, Université Catholique de Louvain, Louvain-la-Neuve, Belgium

Marc Fleurbaey CNRS, Paris, France;
Paris School of Economics, Paris, France

W. Gaertner Fachbereich Wirtschaftswissenschaften, Universität Osnabrück, Osnabrück, Germany

William Gehrlein University of Delaware, Newark, USA

Peter J. Hammond Department of Economics, Stanford University, Stanford, CA, USA

J. S. Kelly Department of Economics, Syracuse University, Syracuse, NY, USA

Jordi Massó Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Barcelona, Spain;
Barcelona GSE, Barcelona, Spain

Philippe Mongin GREGHEC, CNRS & HEC Paris, Jouy-en-Josas, France

P. K. Pattanaik Department of Economics, Faculty of Commerce and Social Science, The University of Birmingham, Birmingham, UK

Fred Roberts Rutgers University, New Brunswick, USA

Maurice Salles CREM (UMR-CNRS 6211), University of Caen-Normandy, Caen cedex, France;
CPNSS, London School of Economics, London, UK;
Murat Sertel Center for Advanced Economic Studies, Bilgi University, Istanbul, Turkey

Kotaro Suzumura Institute of Economic Research, Hitotsubashi University, Kunitachi, Tokyo, Japan

Karine Van der Straeten Toulouse School of Economics & CNRS, Toulouse, France

Roberto Veneziani Queen Mary University of London, London, USA

John A. Weymark Departments of Economics and Philosophy, Vanderbilt University, Nashville, TN, USA

Yongsheng Xu Georgia State University, Atlanta, USA

A Brief History of Social Choice and Welfare Theory

Marc Fleurbaey and Maurice Salles

JEL classification: D6 · D7 · I3

'Political Economy or Economics is a study of mankind in the ordinary business of life; it examines that part of individual and social action which is most closely connected with the attainment and with the use of the material requisites of wellbeing.' This is the first sentence of Alfred Marshall's *Principles of economics* (Marshall 1920) essentially outlining that economics as a discipline has to be some welfare theory. When our old masters, for instance Adam Smith, John Stuart Mill, Henry Sidgwick or Léon Walras define economics, they allude to studies related to a beneficial social organization. What is particularly interesting in Marshall's definition is precisely his use of the word 'wellbeing.'

Welfare economics is definitely established by the publication of Arthur Pigou's magisterial volume entitled *The Economics of welfare* (Pigou 1932) even though some previous editions appear as *Wealth and welfare*.

The modern theory of social choice is easily dated. We can identify three sources: Duncan Black's papers on majority rule (Black 1948a, b, 1958) and Kenneth Arrow's RAND discussion paper at the end of the 1940s (Arrow 1948), and Arrow's paper (1950), Georges Guilbaud's paper (1952) and Arrow's monograph at the beginning of the 1950s (Arrow 1951, 1963). Among these sources, Arrow's contributions stand

M. Fleurbaey (✉)
Paris School of Economics, Paris, France
e-mail: marc.fleurbaey@gmail.com

M. Salles (✉)
CREM (UMR-CNRS 6211), University of Caen-Normandy, 14032 Caen cedex, France
e-mail: maurice.salles@unicaen.fr

CPNSS, London School of Economics, Houghton Street, London WC2A 2AB, UK

Murat Sertel Center for Advanced Economic Studies, Bilgi University, Istanbul, Turkey

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_1

out for at least two reasons. First, Arrow's book is obviously in the welfare economics tradition even if voting is not absent (after all, social choice is about the selection of options by a society of individuals, and voting is a standard way to proceed to such a selection). Second, Arrow uses a relation framework (probably borrowed from the Polish logician Alfred Tarski), a framework which will become the standard framework of social choice theory (and, to some extent, part of the standard framework of microeconomic theory).

In the following part of this foreword, we will briefly sketch the history of social choice and welfare theory. Interested readers, for a more detailed survey, may consult the chapter on 'Social choice' by Maurice Salles and the chapter on 'Welfare economics' by Antoinette Baujard in the third volume of the *Handbook on the history of economic analysis* (Faccarello and Kurz (eds.), 2016).

1 History of Social Choice and Welfare Theory: A Brief Sketch

1.1 Precursors

There is surely much to explore in Plato and Aristotle regarding the general organization of social life. However, one must remember that they were both opposed to democracy. On the other hand, Aristotle has long developments on equality and the search of happiness or welfare (eudaemonia). Welfare economics too rarely refers to this concept and to Aristotle's analysis of it.

Precursors of social choice theory are generally associated with voting rather than welfare. Among these precursors, Pliny the Younger's reputation is due to Farquharson (1969). In his Letter 14 of Book VIII (Pliny 1969), Pliny the Younger describes a situation where strategic voting by a group of senators reveals to be beneficial (for the members of the group). Pliny's description is an example of the manipulation by a coalition of a voting procedure (here, plurality rule—also called 'first-past-the-post'), a theoretical question which will be solved (separately) in the 1970s by Allan Gibbard, Prasanta Pattanaik and Mark Satterthwaite.

Other precursors include the Catalan Ramon Lull and the German Nikolaus von Kues, respectively, in the 13th and the 15th century. Lull proposes, among other things, a rule known in the 20th century as the Copeland rule on the basis of binary majority voting and von Kues (also known under the Latin form of Cusanus) proposes a rule which is generally attributed to Borda, Borda's rule, where points are given to options on the basis of their position in the rankings of the voters (with k options, an option ranked first gets $k-1$ points, an option ranked second gets $k-2$, etc., and for each option, points are added, which determines a 'social' or 'collective' ranking based on the points obtained by each option. There are other precursors, and we do not doubt that further historical researches will lead to the discovery of heretofore unknown voting theory scholars.

1.2 First Foundation

1.2.1 Voting

Jean-Charles de Borda and Nicolas de Condorcet are rightly considered as the fathers of social choice and voting theory. The reason why they must be distinguished from those we call 'precursors' is the scientific aspects of their contributions. Borda's work on voting is quite limited, just a few pages in a mathematical paper in which he introduces what is now known as Borda's rule but where he also shows that plurality rule is flawed since it can select an option which is pairwise defeated (using majority rule) by all other options (being then a Condorcet loser according to modern terminology). However, there is, surprisingly, no proof that a Borda winner cannot be a Condorcet loser (Borda 1784).

Condorcet, on the other hand, devotes many pages (a book, several papers or pamphlets) to voting. His main contribution is the 1785 *Essai* (Condorcet 1785). This work is essentially devoted to the use of probability theory to solve specific questions such as what is now described as Condorcet jury theorem. But developments on this topic are rather recent and Condorcet is perhaps more famous for what he proposes in the preliminary part of his *Essai* in which one can discover the notions of Condorcet winner and the Condorcet paradox. The Condorcet paradox (or voting paradox) can be presented in a very simple way. Suppose there are three voters and three options a, b and c. Voter 1 ranks the options in the order abc (a first, b second and c third). Voter 2 ranks them in the order bca and voter 3 in the order cab. According to majority rule, x is declared socially preferred to y if the number of voters who prefer x to y is greater than the number of voters who prefer y to x. Then, one can observe that a is declared socially preferred to b, b socially preferred to c and c socially preferred to a. There is no Condorcet winner (an option which is socially preferred to all the other options, using majority rule). Condorcet's own example in the *Essai* is more complex. Another example provided by Condorcet consists in showing that Borda's rule can fail to select an existing Condorcet winner.[1]

Among the important contributions during the 19th century, we must mention Charles Dodgson (known under his pen name as Lewis Carroll) and Edward Nanson. In particular, Dodgson proposes various rules based on majority rule and Borda's rule, but he is now better known for a rule which has been studied by computer scientists. When there is no Condorcet winner, Dodgson suggests that we select the option which becomes a Condorcet winner after the smallest number of transpositions in the individual rankings. For instance, in the case of Condorcet paradox presented above, if voter 2 rather than having the ranking bca has a ranking bac, one can observe that a becomes a Condorcet winner after a unique transposition, ca becoming ac. However, in this precise example with its property of symmetry, one can see that similar transpositions would lead to b or c being Condorcet winners after a single transposition.

[1] Condorcet never mentions Borda, he mentions instead 'a famous Geometer' (un 'Géomètre célèbre').

1.2.2 Welfare

In the 19th century, the main contributions to British moral and political philosophy are certainly relative to the development of utilitarianism with the landmark work of Jeremy Bentham, John Stuart Mill and Henry Sidgwick. The fact that Mill and Sidgwick were also economists explains in part why utilitarianism permeates economics with the supposed possibility to measure utility and to make relevant interpersonal comparisons of such utilities. This is the basis of welfare economics à la Marshall or even à la Pigou.

1.3 Second Foundation

1.3.1 Welfare

In 1932, Lionel Robbins attacks these assumptions, in particular the comparability of utility across individuals. Robbins wants to clearly distinguish what is relevant to economics from what is relevant to moral and political philosophy. What remains is essentially ordinal utilities and Pareto criteria, probably an indirect and unintentional homage to the Lausanne school of economics represented by Léon Walras and Vilfredo Pareto! This attack defines, in some sense, the date of death of 'old' welfare economics (perhaps, in Robbins' view the definite death of welfare economics). However, some economists try to rescue a welfare economics based on Paretianism, and in the 1930s, several papers appear on so-called compensation tests (with contributions by, among others, John Hicks, Nicholas Kaldor and Tibor Scitovsky). Another approach is followed by Bergson (1938) with the introduction of the notion of social welfare function, a notion which will be later developed by Samuelson (1947). Social welfare is supposed to be a function of individual utility functions (in Samuelson's version). With the appropriate mathematical framework, Pareto criteria can take the form of the positivity of (first-order) partial derivatives.

Kenneth Arrow's 1951 monograph belongs to the descent of this literature. This assertion is exemplified by a few facts. First, Arrow uses the same phrase, viz. 'social welfare function'. Second, he devotes several pages of his short book to the compensation principles. However, his social welfare function is different from the Bergson–Samuelson variety. The Bergson–Samuelson function is a function of given individual utility functions which are themselves functions whose domain is the set of social states (some subset of a finite-dimensional Euclidean space). The utility functions can be considered as representing preference orderings so that these orderings are fixed. In Arrow's framework, each individual has any possible preference ordering over the set of social states, and accordingly, these individual orderings are not fixed: The variables are the lists of possible individual orderings (generally called profiles of individual orderings). The values taken by the social welfare function are social preference orderings. In the second edition of the monograph (Arrow 1963), the social welfare function is supposed to satisfy four conditions. First, individual preference orderings are not restricted. This means that if, for instance, the individual orderings are complete preorders of the social states (complete rankings with the possibility of ties, also called weak orderings), with three social states, then there are 13 possibilities. Second, in case of unanimous preferences of a social state *a* over a social state *b*, then the social state *a* must be socially preferred to social state *b* (Pareto criterion). Third, there is no individual whose (strict) preference becomes, whatever are the other individuals' preferences, the social strict preference (such an individual would be called a 'dictator'). Fourth, if in two lists (profiles) of individual preference orderings two social states *x* and *y* are pairwise preference-identical—for instance individual 1 prefers *x* to *y* in both profiles, individual 2 is indifferent between *x* and *y* in both profiles, etc.—then the two values taken by the social welfare function must be pairwise preference-identical—for instance *x* is socially preferred to *y* for both profiles. This fourth condition is called by Arrow 'independence of irrelevant alternatives'. Arrow proves that if there are at least two individuals (the number of individuals being finite) and three social states, there is no social welfare function satisfying the four conditions.

1.3.2 Social Choice, Voting, Mathematical Politics

Rather than 1951, 1948 is certainly the birth date of modern social choice. The two papers by Black mentioned above where majority and special majority rules are studied and the property of single-peakedness is introduced appear in 1948. However, if 1948 is the right date, it is primordially due to Arrow's RAND discussion paper P-41 in which one can find most of the theoretical results of the 1951 monograph and even some elements on binary relations and choice functions which will later appear in (1959) in *Economica*. What makes Arrow the father of modern social choice is not only the statement and proof of his impossibility theorem, but also the introduction of a mathematical framework based on relations (inspired by Tarski) to replace the usual instruments of mathematical analysis which, at the time, were used by mathematically inclined economists. The RAND discussion paper already includes an analysis of single-peakedness and majority decisions on the basis of a relational framework in contrast with what Black did.

If we call 'aggregation functions' functions whose domain are lists of individual preference orderings over social states and whose range is a set of (social) binary relations over these social states, we can emphasize the difference between the Arrovian and the Bergson–Samuelson versions. In Arrow's framework, the domain is a set of profiles. In Bergson–Samuelson's framework, there is a unique profile. According to Samuelson, the Arrovian analysis belongs to a scientific domain which could be called 'mathematical politics' but does not belong to economics. A dichotomous view due to Sen (1977) could be a (partial) solution to the question. A first interpretation of the aggregation problem consists in considering that the social relations are the results of amalgamating the data expressed by individuals (profiles of opinions, preferences, etc.) under a procedure or rule adopted by the society with the ultimate purpose to take decisions (this is the mathematical politics/voting view). A second interpretation is that the social relation reflects a specific individual's judgments about the goodness of the social states on the basis of the well-beings of all the individuals. This second interpretation is certainly related to what Bergson, Samuelson and others (such as John Harsanyi and Ian Little) were mostly concerned with. However, we think that the uniqueness of profile is not a necessary ingredient of this second interpretation.

A major work, somewhat neglected because written in French, is by Guilbaud and appears in 1952. In this long paper, Guilbaud recalls the importance of Condorcet and, among many developments regarding new ways to formally analyze the aggregation problem, provides what is surely the first combinatorial analysis of the Condorcet paradox, giving in a footnote a formula for the limit when the number of individuals tends to infinity, a formula which remained rather enigmatic until Peter Fishburn, William Gehrlein, Dominique Lepellley and others introduced probability models to calculate the number of occurrences of various pathologies of voting rules. Guilbaud also studies structures such as simple games, proper simple games, lattices and even suggests the possibility of fuzzy preferences. Furthermore, he proposes an analysis of judgment aggregation, a subject which will be further developed many years later by, among others, Franz Dietrich, Christian List and Philip Pettit.

1.3.3 Utilitarianism Revisited, Social Justice, Moral Philosophy

Harsanyi in (1953; 1955) reconsiders summation-based social welfare functions by introducing assumptions on individual utility functions and on the social utility function borrowed from an analysis which had been devised to deal with risky situations within game theory. Harsanyi assumes that the individual and social utility functions are von Neumann–Morgenstern utility functions. Accordingly, he has to transform the space of social states in an appropriate way. With the help of some kind of Paretian property, he is able to show that the social utility can be represented as a weighted sum of individual utilities. In his 1953 note, he proposes a kind of probabilistic veil of ignorance. This concept of veil of ignorance, generally associated with John Rawls, will be the object of discussion among moral philosophers and welfare economists starting with Rawls' *Philosophical Review* paper of (1963).

Harsanyi's analysis of utilitarianism was made precise later, in particular thanks to contributions by John Weymark and Philippe Mongin.

It was largely unnoticed by economists that some philosophers in the 1950s, as exemplified by Rawls's 'Justice as fairness' (Rawls 1958), were interested in social choice theory and welfare economics. Rawls refers to Arrow, Little, Hicks and Scitovsky.

1.4 Social Choice Theory Around 1970

In the years following the publication of Arrow's book, social choice studies hardly develop. This phenomenon is not exceptional. After, in some sense, 'revolutionary' publications time has to pass to permit these landmark publications to permeate the research world. For instance, this somehow happened to game theory as well. During many years after *Games and economic behavior* appeared, just a few papers, mainly mathematical papers, were written and appeared mostly in mathematical journals or special books such as the books in the well-known Princeton University Press series 'Annals of Mathematics Studies'. Regarding social choice, we must mention papers by Kenneth May on the characterization of majority rule and the relation between social and individual choice and by Julian Blau in which a formal (but benign) mistake in Arrow's proof is corrected (this will lead Arrow to modify the properties of his social welfare functions and to provide a new proof in the added chapter of the second edition of his book). There are also important papers by Ken-Ichi Inada. From 1964, there has been an upsurge of fundamental publications. These mainly concern the existence of solutions for the aggregation of preferences with majority rule, either the existence of Condorcet winners and/or properties of rationality of the social binary relation when individual preferences are restricted. Contributors include Sen, Inada, Pattanaik, Fishburn, Wulf Gaertner and, for other aggregation procedures such as simple games, Michael Dummett and Farquharson (with an earlier publication in 1961), Maurice Salles, Kenjiro Nakamura and Mamoru Kaneko, among others.

Arrow's theorem holds for aggregation functions where the assumed collective rationality property of the social preference is transitivity. Impossibility results are shown in the case of quasi-transitivity (transitivity of the strict component of the social preference) and acyclicity of this strict component by Allan Gibbard, and Andreu Mas-Colell and Hugo Sonnenschein, and in the case of semi-orders or interval orders by Douglas Blair, Blau, Georges Bordes, Jerry Kelly and Kotaro Suzumura, among others.

Structural properties of specific coalitions are discovered (for instance, filters by Bernard Monjardet). The necessity of the finiteness of the set of individuals is demonstrated by Fishburn and further studied by Alan Kirman and Dieter Sondermann.

In 1970, Sen, with a short paper (Sen 1970b) triggers a kind of revolution: the non-welfaristic approach to social choice and normative economics. He demonstrates the impossibility to have a rational social preference (in this case the acyclicity property) with the Pareto criterion (unanimity principle) and the fact that at least two individuals have each the power to impose their strict preference over two options, this power being interpreted as being over a personal sphere and, as such, as a freedom of choice property.

Another major result due (separately) to Gibbard, Satterthwaite and Pattanaik concerns the impossibility to have procedures which are immune to strategic behavior (except, of course, dictatorship). In popular parlance, this is often called the phenomenon of 'useful voting'. We can note that this is the question which Pliny the Younger was dealing with. This result, conjectured by Dummett and Farquharson (1961), was proven by Gibbard (1973), Pattanaik (1973) and Satterthwaite (1975).

One of the most important historical achievements of the period is the publication of a few books by Yasusuke Murakami (1968) outlining the relation between social choice and logic, Sen (1970a and, expanded edition, 2017), Pattanaik (1971) and Fishburn (1973). The book *Collective choice and social welfare* by Sen has to be distinguished for many reasons but mainly because it is at the same time a research monograph and a kind of textbook (and, as such, was crucial in the destiny of the authors of this foreword!). It also elegantly links formal and philosophical aspects of social choice theory.

1.5 *The Recent (Say, Approximately Post 1975) Development of Social Choice and Welfare Theory*

1.5.1 Arrovian Aggregation Theory

We consider here the legacy of Arrow's aggregation analysis leading to the impossibility theorem. Particularly noteworthy is the approach using topological ingredients. This approach is introduced in several papers by Graciela Chichilnisky from 1980 with further contributions by, among others, Geoffrey Heal, Nick Baigent, Luc Lauwers, Yuliy Baryshnikov, Charles Horvath and Shmuel Weinberger (with earlier contributions by Beno Eckmann). A basic characteristic of these contributions is the replacement of the condition of independence of irrelevant alternatives by a continuity property. This makes this kind of analysis more suitable for welfare theory than for voting since, typically, assuming continuity with, say, majority rule would not permit dramatic changes of the social outcomes (unless discontinuities are possible).

Arrovian aggregation when preferences are fuzzy is developed from 1980 by Richard Barrett, Pattanaik, Salles, Bhaskar Dutta, Ashley Piggins, Conal Duddy, Juan Perote-Penã and others (among whom a remarkable group of mathematicians in Cameroon–Nicolas Andjiga, Louis-Aimé Fono and others).

In direct descent from the Arrow–Black tradition of restricting individual preferences, one must outline a domain of interest for microeconomists: social choice in economic environments. A remarkable and detailed survey is by Le Breton and Weymark (2011).

1.5.2 Positive Political Theory and Voting Theory

The so-called economic environment has its counterpart in political science and can be called 'political environment.' The space of options has a rather similar structure, but the individual preferences in the political environment are very often so-called Euclidean preferences: There is an optimal point (where utility/preference is maximized) and moving away from this point makes the utility decrease. Among the major contributors, we must mention, among others, William Riker, Charles Plott, Richard McKelvey, David Austen–Smith, Jeffrey Banks, Norman Schofield and John Ferejohn.

Regarding voting, we can classify this large area according to several characteristics. We can consider a simple game structure in which powerful coalitions are defined a priori. A number of solution concepts are introduced in this area, such as the core (which, of course, was already studied in a microeconomic framework or as a kind of mathematical concept), the Banks set, the uncovered set (due to Nicholas Miller), the stability set (due to Ariel Rubinstein). Major contributions are due to Nakamura, Bezalel Peleg and Hans Peters,[2] and regarding the mathematical structure to Alan Taylor and William Zwicker, and to Moshé Machover and Simon Terrington. In particular, Nakamura (1979) introduces a number based on the structural properties of the voting games which plays a basic role in results regarding the non-emptiness of the solution concepts when this number is compared with the number of options or, in the case where the space of options is an appropriate subset of an Euclidean space of finite dimension, with this dimension (as shown by Joseph Greenberg (1979) for quota-games). Game-theoretic aspects are linked to the rather ancient analysis of voting power. Modern indices of power are due to, among others, Lionel Penrose, John Banzhaf, Lloyd Shapley and Martin Shubik [an excellent book on voting power is by Dan Felsenthal and Machover (1998)].

Powerful coalitions may be studied from another viewpoint. The question of coalition formation has been tackled by Demange and Wooders (2005) and is related to network theory, a topic which has known a spectacular development in the last few years.

We can consider specific voting procedures. Such procedures are, for instance, scoring rules such as plurality rule or Borda's rule. Borda's rule is prominent in the works of Dummett (1984, 1997) and Donald Saari who proposes an entirely new way to deal with voting theory with his 'geometry of voting' (1994, 2007). In our view, Donald Saari is one of the main contributors to voting theory in this post 1975 period.

Among other rules which have been proposed to replace procedures used in various countries, one must indicate 'approval voting' which has been studied by Steven Brams and Fishburn, and 'majority judgment' which has been devised and extensively studied by Michel Balinski and Rida Laraki (2010).

Already mentioned, an important domain stemming from the work of Fishburn and Gehrlein is the computation of the probability of pathological cases for specific voting procedures (Gehrlein and Lepelley 2017). Furthermore, Felsenthal and Hannu Nurmi provide several studies of voting rules detailing their pathologies and good properties [for instance, Felsenthal and Nurmi (2018)].

[2]See Peleg (1984), Taylor and Zwicker (1999), Peleg and Peters (2010).

1.5.3 Computational Social Choice

When we consider voting rules such as Dodgson's rule or Kemeny's rule, the problem is not to know whether there exists a solution (say, a winner), but to determine this solution (who won?). John Bartholdi, Craig Tovey and Michael Trick remark: 'we think Lewis Carroll [Dodgson's pen name] would have appreciated the idea that a candidate's mandate might have expired before it was ever recognized!' (Bartholdi et al. 1989). Although the foundational papers on this topic appear around 1990, it is only very recently that a whole domain takes off. It is exemplified by the publication of the *Handbook of computational social choice* in 2016 (Brandt et al. 2016). This handbook presents many chapters where the determinacy problem is crucial, be it in voting or in other areas where one has to determine a solution or to analyze the possibility for individuals to make a decision about how to proceed, as, for instance, the decision to undertake a strategic move.

As a by-product, a number of computer scientists have developed an interest in social choice theory in general, and many publications which are independent of the computational complexity problem have them as authors.

1.5.4 Logic and Social Choice, Judgment Aggregation

Another by-product is the consequence of the fruitful relations between formal logic and computer science. We mentioned that Arrow's analysis based on relations owes much to Tarski's theory of relations. So, at the origin, social choice was linked to logic. Later, Murakami used three-valued logic to formalize preferences. We have now some versions of Arrow's theorem based on formal/mathematical logic by Ulle Endriss, Umberto Grandi, Wesley Holliday and Eric Pacuit among others (see, for instance, Holliday and Pacuit 2020). There were precursors among whom Willard von Orman Quine himself, Rubinstein (1984) who demonstrated that, given appropriate assumptions, the multi-profile and single-profile impossibility theorems were essentially equivalent.[3]

Judgment aggregation has also known a major development under the impetus of scholars such as, among others, Dietrich, List, Pettit, Clemens Puppe, Klaus Nehring, Philippe Mongin, Elad Dokow and Ron Holzman. Their contributions can be attached to the formal logic domain since they sometimes deal with the aggregation of logic connectives (and their papers have definitely a formal logic flavor).

1.5.5 Strategy-Proofness and Implementation

Stemming from the papers by Gibbard, Pattanaik and Satterthwaite, many contributions to the area of strategic voting appear in the post-1975 period. The main

[3]Some recent textbooks on logic devote a chapter to social choice, for instance Hansson and Hendricks (2018) and de Swart (2018).

contributor to this area is Salvador Barberà. The classical proofs in the original works were based on the finiteness of the set of options. Important papers, including a paper by Barbera and Peleg (1990), deal with economic environments where one cannot assume finiteness.

The so-called Gibbard–Satterthwaite theorem can be interpreted in terms of implementation: The only mechanism that can truthfully implement a social choice function in dominant strategy is dictatorship. This might appear as a rather ad hoc way to transform a negative result (the impossibility to have a social choice function/voting rule which is individually strategy-proof (often said non-manipulable) and non-dictatorial) into a positive result. Of course, the many developments by, among others, Peleg, Peters, Matthew Jackson, Hervé Moulin deal with other frameworks (other ways to aggregate individual data and other solution concepts—such as, for instance, Nash equilibria). In the pursuit of this research, an important concept is devised by Peleg and Moulin, the concept of effectivity function. This concept is also central in some developments about rights and freedom.

1.5.6 Rights and Freedom

A rather large literature descends from Sen's impossibility theorem. However, some scholars were rather unsatisfied by the treatment of rights and freedom via an aggregation procedure. Two routes were followed. A first approach is due to Peter Gärdenfors (1981). Gärdenfors essentially uses effectivity functions, a concept that was forged after Gärdenfors' paper appeared. A second approach is via game forms and is a major element of an important paper by Gaertner et al. (1992).

The analysis of freedom in terms of choice is also a major research topic. The seminal paper where the cardinality of choice sets plays a crucial role is due to Pattanaik and Xu (1990).

1.5.7 Bargaining Theory, Allocation Rules, Cost Sharing, Matchings

This is a rather heterogenous title, but this grouping may be justified by the fact that the formal methods are similar and the contributors are often the same persons. Another justification is that these areas stem from the works of three mathematicians/game theorists who were graduate students at Princeton University: John Nash, David Gale and Lloyd Shapley.

The theory of fair allocation rules [which is strongly linked to welfare theory as seen in Fleurbaey and François Maniquet (2011)] is well documented with books by Kolm (1972, 1988), H. Peyton Young (1994), Moulin (2003) and Julius Barbanel (2005), and the domain is surveyed in papers by William Thomson [see, in particular, Thomson (2011)], a major contributor to the area.

1.5.8 Welfare Economics, Interpersonal Comparability, Inequality, Poverty

Periodically, the death of welfare economics is announced. Atkinson (2009) speaks of 'strange disappearance'. A rather recent and supposedly fatal blow to the concept of Bergson–Samuelson social welfare function is the discovery of a single-profile Arrovian impossibility theorem. However, many including us deny this. The single-profile approach entails the use of properties such as neutrality with a strong welfaristic flavor. Furthermore, we think that, on the contrary, welfare economics is still alive and rapidly developing again under various guises. We already mentioned the theory of fair allocation. But we should also include renewed analyses of utilitarianism (Bayesian aggregation à la Mongin and Richard Bradley (see Bradley 2017), utilitarianism à la Blackorby et al. (2005), or the numerous and important contributions of Peter Hammond). The problems related to the measurement questions such as interpersonal comparability are the object of an impressive literature with Sen and, among others, Claude d'Aspremont, Louis Gevers (see d'Aspremont and Gevers 2002), and from the philosophical side Derek Parfit (1984), John Broome (1991) and Ruth Chang (1998).

The measurement aspects of inequality and poverty are also topics which are within welfare economics with the major contributions of Atkinson, Kolm, François Bourguignon, Le Breton, Alain Trannoy, Patrick Moyes, Udo Ebert, Peter Lambert, K.-Y. Tsui, Bhaskar Dutta, Satya Chakravarty, James Foster, Martin Ravallion and Claudio Zoli among others.[4]

The approach using capabilities (largely due to Sen and Martha Nussbaum) has known an important and rather independent development, but we think that it belongs to welfare economics. The works on this capabilities and human development approach have essentially focused on specific 'functionings' but we certainly need a formal and general analysis along the lines initially developed by Sen.

1.5.9 Distributive Justice

Major contributors to the distributive justice area are economists who are also recognized as philosophers such as Sen, Kolm, Broome or John Roemer. Philosophers such as Patrick Suppes, Donald Davidson, Gibbard, Robert Nozick, Nussbaum, Dummett, Daniel Hausman, Ronald Dworkin, Rawls, Bernard Williams, Gerald Cohen made crucial incursions into social choice and welfare theory, and it often seems impossible to assert whether the contributors to the social choice and welfare theory literature are more philosophers than economists or the reverse. What seems important is the recent publication of volumes by Roemer (1996), Kolm (1972, 1988, 1997, 2005),

[4]Many important recent studies in inequality and poverty such as those of, for instance, Thomas Piketty and Esther Duflo pertain to positive rather than normative economics even if, for such topics, this dichotomous partition can be problematic because the frontier between the positive and the normative is rather fuzzy.

Sen (2009), Fleurbaey (1996), Samuel Freeman (2018), Cohen (2008) where the word 'justice' is the main term of the title. This fact makes us think that academic research on social justice and what can be called 'social ethics' is not out of fashion as it could have appeared one or two decades ago. One can imagine that welfare economics as a sub-domain of social justice will know a new renaissance.

1.5.10 Experimental and Behavioral (welfare) Economics

The first paper ever published in the journal *Social Choice and Welfare* is an experimental paper on social justice (Menahem Yaari and Maya Bar-Hillel 1984). Although we hardly had other publications of this type in any journals during many years, there has been a change, and we can now observe many such papers either in *Social Choice and Welfare* or elsewhere, as exemplified by the references in the book by Erik Schokkaert and Gaertner (2012).

Regarding behavioral economics, there certainly is a problem to reconcile a so-called positive approach with the normative economics approach pervading social choice and welfare theory, but, as shown in a special issue of *Social Choice and Welfare*, Ben McQuillin and Robert Sugden seem to be optimistic.

2 The Conversations

Several interviews were conducted from the 1980s and appeared in *Social Choice and Welfare*. Arrow was interviewed by Jerry Kelly, and their conversation appeared in 1987, just a few years after the creation of this journal. This was followed by conversations with Sen, Harsanyi, Little, Dummett, Kolm, Samuelson and Suzumura.

Given the positive feedback from the readers and the intrinsic interest of these conversations regarding the history of thought (economic or else), the decision was taken to pursue this enterprize and to devote special volumes to it in the Springer series entitled 'Choice and Welfare.' At this time, we intend to devote two volumes to this project.

This volume, the first one, is organized in two parts and within each part by the alphabetic order of the scholars who were interviewed.

The first part includes four 'chapters' with conversations with scholars who can be considered as the founding fathers of the modern theory of social choice and welfare: Kenneth Arrow, John Harsanyi, Paul Samuelson and Amartya Sen. We can observe that these four are Nobel laureates, and for Arrow and Sen, the prize was, at least in part, linked to their work in social choice.

In the second part, we gathered scholars whose main contributions appeared after 1970 (this does not mean that they did not publish important contributions before 1970—this is the case, for instance, of Peter Fishburn and Prasanta Pattanaik). Furthermore, the chosen date of 1970 must be taken as rather vague: It can be 1969 or 1971! Incidentally, some of them were quite young in 1970, and it is difficult to imagine that they could have written significant scientific papers at such an early age.

There are eleven scholars in this second part of the book: Salvador Barber à, John Broome, Gabrielle Demange, David Donaldson, Peter Fishburn, Allan Gibbard, Peter Hammond, Prasanta Pattanaik, John Roemer, William Thomson and John Weymark.

We already have plans for the second volume which will include conversations already published in *Social Choice and Welfare* as previously indicated (Little, Dummett, Kolm and Suzumura) and new conversations which are actively organized at this time.

We have a deep regret. We initially planned interviews of Jan de V. Graaff and of Patrick Suppes (Patrick Suppes had even been contacted and had kindly accepted). Unfortunately, we were unable to carry out our plans regarding these two great minds.

References

Arrow, K. J. (1948). *The possibility of a universal social welfare function*. Document P-41, Sept 26, 1948, RAND Corporation.

Arrow, K. J. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*, 328–346.

Arrow, K. J. (1951). *Social choice and individual values*. New York: Wiley.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, *26*, 121–127.

Arrow, K. J. (1963). *Social choice and individual values* (2nd ed.). New York: Wiley.

Atkinson, A. B. (2009). Economics as a moral science. *Economica*, *76*, 791–804.

Balinski, M., & Laraki, R. (2010). *Majority judgment. Measuring, ranking, and electing*. Cambridge (Mass.): M.I.T. Press.

Barbanel, J. B. (2005). *The geometry of efficient fair division*. Cambridge: Cambridge University Press.

Barberà, S., & Peleg, B. (1990). Strategy-proof voting schemes with continuous preferences. *Social Choice and Welfare*, *7*, 31–38.

Bartholdi III, J. J., Tovey, C. A., & Trick, M. A. (1989). The computational difficulty of manipulating an election. *Social Choice and Welfare*, *6*, 227–241.

Baujard, A. (2016). Welfare economics. In G. Faccarello & H. D. Kurz (Eds.), *Handbook on the history of economic analysis, Volume III-Developments in major fields of economics* (pp. 611–623). Cheltenham: Edward Elgar.

Bergson, A. (1938). A reformulation of certain aspects of welfare economics. *Quarterly Journal of Economics*, *52*, 310–334.

Black, D. (1948a). On the rationale of group decision making. *Journal of Political Economy*, *56*, 23–34.

Black, D. (1948b). The decisions of a committee using a special majority. *Econometrica*, *16*, 245–261.

Black, D. (1958). *The theory of committees and elections*. Cambridge: Cambridge University Press.

Borda, J.-C., Chevalier de. (1784). Mémoire sur les élections au scrutin. *Histoire de l'Académie des Sciences pour 1781* . Paris: Imprimerie Royale.

Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social choice theory, welfare economics and ethics*. Cambridge: Cambridge University Press.

Bradley, R. (2017). *Decision theory with a human face*. Cambridge: Cambridge University Press.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.). (2016). *Handbook of computational social choice*. Cambridge: Cambridge University Press.

Broome, J. (1991). *Weighing goods*. Oxford: Basil Blackwell.
Chang, R. (1998). *Incommensurability, incomparability, and practical reason*. Cambridge (Mass.): Harvard University Press.
Cohen, G. A. (2008). *Rescuing justice and equality*. Cambridge (Mass.): Harvard University Press.
Condorcet, M.-J.-A.-N. Caritat, Marquis de (1785). *Essai sur l'application de l'analyse à la probabilité des décisions rendues à la pluralité des voix*. Paris: Imprimerie Royale.
d'Aspremont, C., & Gevers, L. (2002). Social welfare functionals and interpersonal comparability. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. 1). Amsterdam: North-Holland.
Demange, G., & Wooders, M. (Eds.). (2005). *Group formation in economics. Networks, clubs and coalitions*. Cambridge: Cambridge University Press.
de Swart, H. (2018). *Philosophical and mathematical logic*. Cham: Springer.
Dummett, M. (1984). *Voting procedures*. Oxford: Oxford University Press.
Dummett, M. (1997). *Principles of electoral reform*. Oxford: Oxford University Press.
Dummett, M., & Farquharson, R. (1961). Stability in voting. *Econometrica*, *29*, 33–43.
Farquharson, R. (1969). *Theory of voting*. Oxford: Basil Blackwell.
Felsenthal, D. S., & Machover, M. (1998). *The measurement of voting power. Theory and practice, problems and paradoxes*. Cheltenham: Edward Elgar.
Felsenthal, D. S., & Nurmi, H. (2018). *Voting procedures for electing a single candidate. Proving their (in)vulnerability to various voting paradoxes*. Cham: Springer.
Fishburn, P. C. (1973). *The theory of social choice*. Princeton: Princeton University Press.
Fleurbaey, M. (1996). *Théories économiques de la justice*. Paris: Economica.
Fleurbaey, M., & Maniquet, F. (2011). *A theory of fairness and social welfare*. Cambridge: Cambridge University Press.
Freeman, S. (2018). *Liberalism and distributive justice*. New York: Oxford University Press.
Gaertner, W., Pattanaik, P. K., & Suzumura, K. (1992). Individual rights revisited. *Economica*, *59*, 161–177.
Gärdenfors, P. (1981). Rights, games and social choice. *Noûs*, *15*, 341–356.
Gehrlein, W. V., & Lepelley, D. (2017). *Elections, voting rules and paradoxical outcomes*. Cham: Springer.
Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*, 587–601.
Greenberg, J. (1979). Consistent majority rules over compact sets of alternatives. *Econometrica*, *47*, 627–636.
Guilbaud, G-Th. (1952). Les théories de l'intérêt général et le problème logique de l'agrégation. *Economie Appliquée*, *5*, 501–551.
Hansson, S. O., & Hendricks, V. F. (Eds.). (2018). *Introduction to formal philosophy*. Cham: Springer.
Harsanyi, J. C. (1953). Cardinal utility in welfare economics and the theory of risk-taking. *Journal of Political Economy*, *61*, 434–435.
Harsanyi, J. C. (1955). Cardinal utility, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*, 309–321.
Holliday, W., & Pacuit, E. (2020). Arrow's decisive coalitions. *Social Choice and Welfare, 54*, 463–505.
Kolm, S.-C. (1972). *Justice et équité*. Paris: Editions du C.N.R.S.
Kolm, S.-C. (1988). *Justice and equity*. Cambridge (Mass.): M.I.T. Press.
Kolm, S.-C. (1997). *Modern theories of justice*. Cambridge (Mass): M.I.T. Press.
Kolm, S.-C. (2005). *Macrojustice*. Cambridge: Cambridge University Press.
Le Breton, M., & Weymark, J. A. (2011). Arrovian social choice theory on economic domains. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. 2). Amsterdam: North-Holland.
Marshall, A. (1920). *Principles of economics* (8th ed.). London: Macmillan.
Moulin, H. (2003). *Fair division and collective welfare*. Cambridge (Mass.): MIT Press.
Murakami, Y. (1968). *Logic and social choice*. London: Routledge.

Nakamura, K. (1979). The vetoers in a simple game with ordinal preferences. *International Journal of Game Theory*, *8*, 55–61.

Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.

Pattanaik, P. K. (1971). *Voting and collective choice*. Cambridge: Cambridge University Press.

Pattanaik, P. K. (1973). On the stability of sincere voting situations. *Journal of Economic Theory*, *6*, 558–574.

Pattanaik, P. K., & Xu, Y. (1990). On ranking opportunity sets in terms of freedom of choice. *Recherches Economiques de Louvain*, *56*, 383–390.

Peleg, B. (1984). *Game theoretic analysis of voting in committees*. Cambridge: Cambridge University Press.

Peleg, B., & Peters, H. (2010). *Strategic social choice. Stable representations of constitutions*. Heidelberg: Springer.

Pigou, A. C. (1932). *The economics of welfare* (4th ed.). London: Macmillan.

Pliny. (1969). *Letters, Book VIII-X, Panegyricus* (B. Radice, Trans.). Cambridge (Mass.): Harvard University Press (Loeb Classic Library).

Rawls, J. (1958). Justice as fairness. *Philosophical Review*, *67*, 164–194.

Rawls, J. (1963). The sense of justice. *Philosophical Review*, *72*, 281–305.

Roemer, J. E. (1996). *Theories of distributive justice*. Cambridge (Mass.): Harvard University Press.

Rubinstein, A. (1984). The single profile analogues to multi profile theorems: Mathematical logic's approach. *International Economic Review*, *25*, 719–730.

Saari, D. G. (1994). *Geometry of voting*. Heidelberg: Springer.

Saari, D. G. (2007). *Disposing dictators, demystifying voting paradoxes. Social choice analysis*. Cambridge: Cambridge University Press.

Salles, M. (2016). Social choice. In G. Faccarello & H. D. Kurz (Eds.), *Handbook on the history of economic analysis, Volume III-Developments in major fields of economics* (pp. 518–537). Cheltenham: Edward elgar.

Samuelson, P. A. (1947). *Foundations of economic analysis*. Cambridge (Mass.): Harvard university Press.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, *10*, 187–217.

Schokkaert, E., & Gaertner, W. (2012). *Empirical social choice. Questionnaire-experimental studies on distributive justice*. Cambridge: Cambridge University Press.

Sen, A. K. (1970a). *Collective choice and social welfare*. San Francisco: Holden-Day.

Sen, A. K. (1970b). The impossibility of a Paretian liberal. *Journal of Political Economy*, *78*, 152–157.

Sen, A. K. (1977). Social choice theory: A re-examination. *Econometrica*, *45*, 53–89.

Sen, A. K. (2009). *The idea of justice*. Cambridge (Mass.): Harvard University Press.

Sen, A. K. (2017). *Collective choice and social welfare: An expanded edition*. Cambridge (Mass.): Harvard University Press.

Taylor, A. D., & Zwicker, W. S. (1999). *Simple games. Desirability relations, trading, pseudoweightings*. Princeton: Princeton University Press.

Thomson, W. (2011). Fair allocation rules. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. 2). Amsterdam: North-Holland.

Yaari, M. E., & Bar-Hillel, M. (1984). On dividing justly. *Social Choice and Welfare*, *1*, 1–24.

Young, H. P. (1994). *Equity in theory and practice*. Princeton: Princeton University Press.

Foundations

Kenneth J. Arrow

J. S. Kelly

The following is an edited transcript of an interview conducted on March 4, 1986, with Professor Arrow while he was visiting Syracuse University to deliver the Frank W. Abrams Lecture Series to be published as *The Uncertain Future and Present Action* by Syracuse University Press. This interview was to elaborate on his description, presented in Volume 1 of his *Collected Papers* (Harvard University Press, 1983), of the origins of his work in collective choice theory.

JK. *You started off the story in the collected papers with remarks about studying relational logic while you were in Townsend-Harris High School in New York City.*

KA. Not in high school in the sense of in my high school courses, but during this period I was an omnivorous reader and got into all sorts of things. One of them was Bertrand Russell's *Introduction to Mathematical Philosophy* and it made a tremendous impression on me. It was the idea of logic that was in there. I don't really recall, for example, if there was a formal definition of a relation as a set of ordered pairs, but I learned the ideas of mathematical logic and its applications to mathematics in Russell's book. It seems to me that I also read one or two other logic books around that time.

JK. *Later, when you went to City College of New York as a mathematics major, you encountered more mathematical logic.*

KA. Yes, but again the logic study was on my own, there were no courses in it. I don't really remember exactly what I read. I remember once taking out the *Principia Mathematica* but of course it's not the sort of thing one really reads from. I was looking up some theorems in it and things like that. I really am not prepared to tell you what I read, but at some point things like the idea of defining rational numbers by ordered pairs and equivalence classes by ordered pairs was something I got to

This chapter was previously published in the journal *Social Choice and Welfare* (1987) 4:43–62.

J. S. Kelly (✉)
Department of Economics, Syracuse University, Maxwell Hall, Syracuse, NY 13210, USA
e-mail: jskelly@maxwell.syr.edu

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_2

know. I was fascinated by this and used to aggravate my professors by writing out proofs in very strictly logical form, avoiding words as much as possible and things of that kind.

JK. You did take a formal course with Tarski in the Philosophy Department; how did you happen to take that course?

KA. Yes. Well, I *knew* that Alfred Tarski was a great and famous logician and there he was in my last term in school and obviously I was going to take a course with Alfred Tarski. It turned out he had two courses. One was a kind of introductory course and I felt I knew more than *that*. The other course he gave was in the calculus of relations. To say it was in the *calculus* of relations meant that he gave an axiomatic treatment of relations, although he motivated it of course by motivating the axioms. You never had xRy; you only had R and S and T. You see, he never mentioned *individuals* in the formal theory. He had an axiomatic theory like an axiomatic treatment of set theory. Relations have some special aspects, in particular the idea of relative product, RS. If there is a z such that xRz and zSy, then $xRSy$. The relative square, $R^2 = RR$ is especially interesting; if the relative square is included in R you have transitivity.

So it was a fascinating thing, although it was really very elementary; really very easy. The concepts were not very subtle compared with the deep things he was working on like the truth principle.

JK. At this point you were involved in translating some of Tarski's work.

KA. He wrote a textbook called *Introduction to Logic* (Tarski 1941) which is one of the modern treatments, modern as of 1940. It had been published in German, may even have been originally published in Polish. I didn't translate it. What happened was he had a translator and I read the proofs. I was just finishing college and he asked me to read the proofs for him. He didn't know any English, you see. This was the interesting thing. He came to this country in September, 1939 for some kind of congress or conference and was trapped here by the outbreak of the war. He knew Polish, he knew German, but he didn't know any English so he spent the Fall term learning some English so he could teach us in the Spring. At first we couldn't understand a word he was saying but after about a week or so we began to catch on and we realized it wasn't *his* rate of progress it was *our* rate of progress that was relevant. His stresses were all wrong. He was aware of this and therefore felt he couldn't proofread in English. It's rather interesting as a coincidence that the translator was a German philosopher named Olaf Helmer and Helmer comes back into my story eight years later.

It's interesting… Tarski, although his English was weak, had a very good sense of language and he kept on asking me "Is that really good English?" Not in the sense of being grammatically correct, but, well for example, Helmer was very fond of using the word "tantamount" and Tarski got the feeling that somehow it's not a word used very often. Actually his instincts for language were extremely good. I suppose that was connected with his general work on formalizations and metalanguages. Anyway, I was just a proofreader.

JK. You write that as a graduate student at Columbia you spent time as an exercise translating consumer theory in the logic of relations and orderings. What got you started on that and what did you get out of it?

KA. I went to Columbia because …well there were several problems. One was that we were extremely poor and the question of going anywhere depended on resources. Columbia had the great advantage, of course, that I could live at home, which wasn't true anywhere else. I didn't get any financial support for my first year, none at all.

But another of the things I had learned on my own at college was mathematical statistics and I really had become fascinated with it. There was a course in statistics [at CCNY]; the teacher, a man by the name of Robinson, had no *real* knowledge of it I would say, basically—I won't even say he had a good reading list—but he did list one book, Kenney (1939) if I remember correctly, which happened to have an excellent bibliography. It was not one of those cookbooks in statistics but actually did have some attempts at mathematics. Kenney had references to R. A. Fisher and gave you enough to get you interested. So I started reading Fisher and one of the first things was trying to work out his derivation of the distribution of the correlation coefficient under the null hypothesis, which was an integration in n-dimensional space. In Fisher it was done by intuition. I mean it's rigorous if you're sufficiently sophisticated; to me it was gibberish. But I knew enough multivariate calculus to be able to translate it into rigorous form, at least a form that I understood, and then I could see that he really was right. But I couldn't see it the way he wrote it. Then I suppose because of my logical background what was really important was reading the Neyman-Pearson papers which were then new and written in rather obscure places, but they were available in the [CCNY] library. From Fisher alone, I think I would have been hopelessly confused about the logic of statistical tests, although Fisher was great on deriving distributions.

So, I knew I wanted to study mathematical statistics, which however was not a field, not a Department at Columbia. It was spread out in other Departments. I knew that Hotelling was one of the major figures, but he was in the Economics Department. I rather naively thought I would study mathematics and then would take the statistics from Hotelling. I had no interest in Economics.

I was in the Mathematics Department, taking courses like Functions of a Real Variable, but I was going to take courses from Hotelling. In the first term he happened to give a course in Mathematical Economics. So out of curiosity I took this and got completely transformed.

The course to an extent revolved around Hotelling's own papers. But, as it happens, they were kind of central. He gave a rigorous derivation of supply and demand. There was one paper on the theory of the firm, one on the theory of the consumer (Hotelling 1932, 1935). And he gave a rigorous derivation of demand functions in the consumer theory paper and derived the Slutsky equations. I think he knew about Slutsky's work, though I'm not sure he actually referred to Slutsky. So, anyway, this was one of the best papers around at the time. It's now a staple of our literature but then really was novel. One of the things, he was a very, *very* strong ordinalist, emphasized that all these results were invariant under monotone transformations, which was not a normal practice in economics at that time. Of course, all those who were coming of age, like Paul Samuelson, would jump to that position; it was the normal position of the avant garde.

Well, the idea was that it was an *ordering*. It was clear that what they were saying was "x is better than y" and that this is a transitive relationship. And I recognized that there were certain continuity axioms that had to be added to that. I was already familiar with that because there were certain similar things in the foundations of probability theory. In fact I think I worked that out for myself. I was playing around once in college trying to work out an axiom system for probability theory, that was work on an Honors paper or something, and I ran across a set of axioms by Karl Popper. Research methods were pretty primitive; I looked through the Union catalog and there was a reference to an article (Popper 1938) in *Mind* by Popper. I realized that his axiom system really couldn't explain certain things that we take for granted like the fact that cumulative distributions have a one-sided continuity property. So I realized that you need some kind of extra continuity axiom and I sort of invented countable additivity all by myself. Later, of course, I found that Kolmogoroff and others had done this, but I could see there had to be an axiom.

So I was kind of familiar from having worked it out there that you needed these continuity axioms in order to close your preference theory system. It was easy to provide and I suppose others were doing the same. I could also see that while it was clarifying for me, it was hardly a contribution to knowledge because all I was doing was translating to a language that I knew. At least it got me thinking; whenever I saw a U for a utility function I translated to a preference ordering.

In fact one thing that struck me as an interesting problem—this is digressing a bit, but not entirely—why should there be a utility function representing an ordering? Hotelling had never really asked that question. Although he emphasized that the indifference map was the primitive, and the utility function only represented it, he didn't really ask "Why should you have a representation in terms of numbers?" I was really thinking about this problem when I happened to run across some papers (Wold 1943) by Herman Wold who gave what he called a "Synthesis" in some papers in *Skandinavisk Aktuarietidskrift* which gave a long treatment of demand analysis which did have essentially an axiomatic point of view. There he said you've got to prove there is a utility function representation. He was the first person I know to realize, in print, that this was a problem. He gave an answer, extremely weak because he needed strong assumptions.

Anyway, then I switched to Economics from Mathematics. I had gone to Hotelling asking for a letter of recommendation for a fellowship in the Mathematics Department and he said, "Well, I'm sure I dont't have any influence in the Mathematics Department, but if you should enroll in Economics, I've found in the past they are willing to give one of my students a fellowship." I was bought.

Incidentically, I impressed him on about the second day of the class because he was fascinated by Edgeworth's taxation paradox; in fact his paper on the theory of the firm was called "Edgeworth's Taxation Paradox and the Nature of Supply and Demand Theory" (Hotelling 1932). Consider a case where there are first class and third class railroad tickets as in the English system. It turns out that if you impose a tax on one ticket then, with suitable demand functions, you could lower the price of both commodities. At the time there was a lot of excitement about that; the public finance people were pooh-poohing it, saying, "How can this be?" It had to do with the nature of inter-related demand curves and that was the big thing Hotelling stressed, that demand functions depended on *n* variables, not one variable. But he said he was puzzled by the fact that he had never been able to produce an example of Edgeworth's paradox with linear demand functions. So I sat down and wrote out the conditions for linear demand functions to yield the paradox; these conditions were certain inequalities on the coefficients and the inequalities were inconsistent. So I came in the next day and showed it to him. Really it was just a few lines, but from that point on he was really impressed with me. It was an extremely easy calculation, but thinking in inequality terms was not common. Little pieces were quite easy to prove, but you couldn't do it in the mechanical fashion which you were doing with, say, solving simultaneous equations or maximizations.

Anyway, I enrolled in Economics and one of the things I read was a brand new book, Hick's *Value and Capital* (Hicks 1939). You know, after reading though the mish-mash like Marshall and things like that, suddenly there was this clear, well-organized view, you knew exactly what was happening. Just the sort of thing to appeal to me. There was a whole, messy, confused literature on capital theory; all those great debates between Knight and von Hayek and all that. And now here was just the idea of dated commodities and suddenly scales fell from your eyes. A simple idea like dated commodities made whole issues transparent.

But as I read Hicks, I could see there were things left out. I turned to this again when I returned from the War, which was really pretty much of a hiatus in any work I was doing—I was gone and very busy for about three and a half years. I had done all my examinations before I had left. So now it was just a question of my thesis. I decided to take *Value and Capital* and redo it properly. I could see all kinds of specific points that were of concern. I wanted to combine it with Samuelson's stability theory, which he had developed in the meanwhile, the papers on dynamic stability in '41 and '42 (Samuelson 1941). Maybe I would add some stochastic elements to the story because as a student of probability and statistics theory I could see noise in the system. Well, it was a lifetime of work, really; it was a very unrealistic thesis.

Hotelling was primarily interested in statistics at this time and then he left for the University of North Carolina. And Abraham Wald wasn't interested in Economics anymore, either. The one I was closest to was Albert Hart, who was regarded then as a very promising theorist, but somehow wasn't able to do what he was capable of. Now people haven't even heard of Albert Gailord Hart. He had a good analysis of flexibility in a Festschrift for Schultz (Lange et al. 1942), another figure who has faded, but I was never impressed by Henry Schultz. Hart considered a problem where you're thinking of buying a durable machine and you're uncertain as to the second period output; the trade-off is between a first machine that would be beautifully optimizing if you knew exactly the output but is not very good at slightly different outputs and a second machine that has costs that are fairly uniform along a wide range. The second machine might sometimes be preferred. He gave a sequential analysis; the idea that your choice today can be dependent on your uncertainty about tomorrow. Elementary as that point may seem, it just hadn't been expressed anywhere; it was very revelatory and came out in this Festschrift for Schultz (Hart 1942), published in '42, '43 or '44.

Hart was friendly and respectful but not very mathematical. One of the things he brought to my attention was that firms are, after all, multiowner objects. It is true that all the owners are interested in the same thing, maximizing profit; however, from a Hicksian point of view, the owners might have different expectations. Then their recommended investment policies today would be different. Each one, trying to maximize the same thing, expected profit, would nevertheless have a different choice. Now, from a modern point of view, we would probably take a different position. In fact the whole idea on my part was wrong; I didn't take account of the very simple point that owners could sell their stock. I didn't think about until I read Modigliani-Miller (1958) years later and realized that my whole attack had been wrong. Those who were most optimistic about the firm would, in effect, buy it out from everybody else, who would do something else with their money. So, in fact, in that context, the voting paradox is irrelevant. But I did not think of it that way, I thought of owners as glued to the firm, and just did not think about the stock market. Within that context I thought, well, how would they decide between two actions. A reasonable thing is to assume it goes by majority vote, a majority of shares, of course. I started writing this down and it occurred to me that I have a preference of the firm defined by the statement that a majority prefers investment A to investment B. Then, from my background, a natural question is, is this relation transitive? Well, it didn't take more than a couple of tries to see that this is not true.

The minute I saw it, I thought: This must be well known. In fact, I thought I might have seen it before. I have no idea from that day to this what I could have seen. However, it is the sort of thing which would appear maybe even in a puzzle page of a newspaper. It could have appeared in some quite trifling way. Anyway, I thought I'd seen it before and I didn't think it was major; all I thought was it was a nuisance because it was spoiling my theory. I ended up trying to develop a theory on the basis of maximizing profits weighted by share numbers, using that as a maximand. Later I gave up the whole thing because it seemed very unwieldly and didn't cohere.

JK. Gave up that section or the whole dissertation?

KA. The whole dissertation. Large parts were in no way novel. At best, I learned some ideas about myopia in investment, things that in later years I pursued. At one point I had about thirty pages of outline of ideas and results but somehow it didn't send me. I felt somehow it was supposed to lead ultimately to empirical work. I was rather discouraged because I was spending quite a bit of time at this. It was a couple of years.

JK. Now comes the Hicks lecture …

KA. Intransitivity was something I had discovered, but it was not on my mind; it was something I had dismissed. Hicks gave a lecture during the winter of '46–'47. He had a very interesting idea. He was trying to find a definition of welfare inequality which was nevertheless consistent with ordinalism. What was meant by saying "individual A is better off than individual B?" Hicks' statement was the following: Suppose individual A prefers his own bundle to B's and individual B also prefers A's bundle to his own. Then Hicks would say A is better off than B. Of course, this definition surfaced again about twenty years later in the work of Foley (1967), but I'd never heard of it before. I think I once found some indication that Trygve

Haavelmo had had that idea and I've tried to find the reference again but I've never succeeded in tracking it down. Hicks presented this lecture and went through a lot of variations of this point but he said Joan Robinson had criticized it and he was a little worried about it. He said, of course, that people might be non-comparable. A could prefer his bundle to B's and B could prefer his own to A's—he recognized that. I thought about it sitting there and finally said, "Would you want this property of 'being better off than' to be transitive?" I defined what I meant by transitive. "Well, in the case, your definition won't satisfy that, because the comparison between A and B is based on A and B's orderings, the comparison between B and C is based on B and C's orderings, and the comparison between A and C is based on A and C's orderings. And it's possible to have A better off than B, B better off than C and C better off than A by this." So I guess somehow this idea of intransitivity fascinated me. Hicks said, "What? What? What?" I don't think he quite got the point. Hart was very quick; he was chairing the meeting, got the point and tried explaining it. Interestingly, Hicks never published this. Now all this may be beside the point, because you may not want transitivity even though it seems natural. In Foley's work, for example, this issue doesn't arise. Foley just said a point is fair if nobody is better off than anybody else. But this story does show intransitivity was bubbling around inside me even though I wasn't consciously aware of pursuing this line as a subject of research.

At this time, I received an invitation to the Cowles Commission. At first I postponed a move because I was trying to finish my Hicksian dissertation before I went there, but I finally settled on finishing it there.

JK. Wasn't it unusual then to leave graduate school before finishing your dissertation?

KA. You know, my knowledge of what was typical wasn't very good. The people I knew were at Columbia. I didn't know what was going on at Harvard or Chicago. It was really very provincial. In fact from what I now know the Columbia situation was unusually chaotic. One problem at Columbia, and I think it's true to this day, is that the sense of community among students and faculty is very weak; they're all dispersed. In particular the National Bureau of Research was a very strong organization and some of the leading Department members went off there, especially the great Wesley Clair Mitchell. So they weren't available. The Bureau was not near the University and so they were just simply physically somewhere else. One had the feeling, in fact, that they never talked to each other.

JK. What led to your invitation to Cowles?

KA. They came around and asked Wald and he recommended me. While he was primarily trained as a statistician, nevertheless he was interested in economics. There weren't many in that category and so they asked him.

Cowles was a funny kind of place because they were kind of a persecuted sect; the mathematical and quantitative emphasis was exceptional and distrusted. My salary was $3200 per calendar year—it was a calendar year appointment. Even by the standards of 1947, that wasn't very much.

JK. What were you hired to do?

KA. What they really wanted me to do was work on statistical problems but it was a free-wheeling place. At the moment the emphasis was on the development of the econometrics of large-scale models. So-called larg: three equations, five equations. Larry Klein ended up with a 20 equation model. Now Tinbergen had even bigger models in his League of Nations study but this was simultaneous equations estimation which made the computational burden very much greater. I had some idea of using higher-order approximations. Others had gotten the asymptotic distributions which were normal and I had been taking a course in Edgeworth-Cramer expansions which you get from higher-order approximations. These ideas were originated by Edgeworth and quite ignored; interestingly, Edgeworth authored quite a few new ideas in statistics, most of which were ignored and then rediscovered. Edgeworth had this method of Cramer rediscovered it. Actually, it was pretty high-powered mathematics and it really was probably beyond me. I knew how to do it, but mathematical estimates of the error term in the approximation were a very subtle and complex matter.

But I was there to do anything I pleased and I was very obviously interested in theory. There was a feeling that theoretical foundations were also an essential part. Finishing my thesis could fit into this.

JK. While you were at Cowles you worked on the single-peakedness result.

KA. I really spent a year there not doing much of anything, to tell you the truth. I wrote a few tiny papers, none of which amounted to anything. I was a great contributor to discussions: argumentative, finding exceptions, errors and counterexamples. But I really felt very discouraged. Once, at lunch, we were talking about politics, left parties and right parties, and I remember drawing on a piece of paper the idea that a voter might have preferences over the parties. It wasn't so much that I saw the ideas—it was the only way I could think about it. It was not that I thought why don't we represent voters as having preferences—as soon as I thought about the question, it couldn't occur to me there was any other way of doing it. So I wrote this thing down and started looking at the question of majorities. It's really hard to describe it. All I can say is, once you've seen it, its obvious; it takes an hour or two. If you ask the question, the answer is fairly obvious. I spent a day or two working it up as a formal proof. And in my usual way, I sort of stalled about a month on writing it up for publication. No, but that doesn't make any difference. I can't say a lost anything. It would just establish that I had the idea independently. But in a sense it didn't matter because within about a month I picked up the *JPE* and there's the paper (Black 1948) by Duncan Black that had exactly that idea.

The coincidence I regard as an extremely interesting point in the history of thought. It's an idea which could easily have occurred to Condorcet. It doesn't depend in any way on the development of mathematics in the last 150 years.

JK. Except to a sensitivity to the logic of relations?

KA. I suppose so. Well, let me put it this way. The logic of relations was worked out at great length in the latter part of the nineteenth Century. Let's say it depends on Boole and the idea of relations as ordered pairs—even that goes back to 1910 or 1911. But while the idea of relations as ordered pairs is a comforting idea, in the sense that you have a logical foundation, it isn't necessary. The fellow who developed a lot of the ideas about logical relations, rather sophisticatedly, was Charles Peirce, the philosopher and logician, founder of pragmatism around 1880–1910. It was picked up by a German named Schroder who in good Germanic fashion wrote three large volumes (Schroder 1890) around 1890. All the apparatus, all the sensitivity was there. If you needed more, the *Principia* surely supplied all that was needed. Nobody asked that question; that's all I can say. Black of course, had been building up to it; he really did have the idea of voting as a mechanism.

JK. Had you read anything of Black's before this paper?

KA. I honestly can't tell you. The stuff before was awfully formal and obvious. This was the only paper of his I seriously regard as having some excitement in it. So, anyway, there was my third encounter with orderings. But that really developed out of amusement.

Then that summer I went to Rand Corporation—again through sheer accident. My wife, who I met as a graduate student in Chicago, had previously worked in the Agriculture Department. She'd arrived there as a clerk and became a professional, a statistician. Her boss was a very distinguished mathematical statistician named M. A. Girshick. Sol was friendly with Girshick who had gone to the Rand Corporation when it was started. The Air Force needed someone to tell them what was going on in the world, so they took all these wild characters and unleashed them. Girshick was one of those invited to go out there and he commenced to spend a couple of years. He often visited Chicago. He had been in contact with the Cowles Commission anyway, because some of his work in multivariate analysis really was very close mathematically, more than mathematically, close conceptually, to simultaneous equation estimation. He was giving advice to them; in fact, he had some ingenious ideas. He contributed a good deal to the development of the limited information method and was never really given full credit for that. Anyway, Girshick had this connection with Cowles independent of us, but when he came to Chicago he visited us.

One of the things Rand was doing was inviting large numbers of visitors for the Summer so Girshick urged me to come. Summer in Santa Monica didn't seem like a bad idea to me and it turned out to be far more intellectually exciting than anything I had planned because the halls were filled with people working on game theory. Everybody was fooling with zero-sum games, how to calculate them, the fundamental definition of the concepts; it was work at the conceptual level and at the technical level.

JK. Was game theory something you'd studied before you went to Rand?

KA. Not really. I mean I knew about the book and had been vaguely pecking away at it, but I hadn't really studied it at all carefully. It was not a big topic at the Cowles Commission, although Marschak had written a review of it.

Anyway, Olaf Helmer was among those who had been brought to Rand; a philosopher. There were several people there as a matter of fact who were basically philosophers; Abraham Kaplan was another. Helmer said to me one day: "There's one thing that disturbs me." They were taking game theory and applying it especially to Soviet-U.S. relations: diplomatic conflict, potential tactical situations, war. However, the payoff functions were defined in terms of utility functions, as Von Neumann and Morgenstern argued in their appendix, and these were derived on the basis of the individual. The trouble was, the Soviet Union and the U. S. were not individuals. What is the meaning of this? You've got to give him credit for proposing the problem.

Now I hadn't spent a lot of time or attention on welfare economics. I had really been trying to work on descriptive theory and general equilibrium theory considered as descriptive, rather than as normative theory. But I did read. One of the things I did when I was supposed to be working on my thesis was read and read and read. It was the typical thing I was doing. I still do it to this day. I'm supposed to be doing something and I find myself picking up something allegedly relevant and reading it. You pick up a lot of information that way at times. I had read Oscar Lange's expository article (Lange 1942) on the foundations of welfare economics. Lange was extremely clear. It was only afterward I began to feel his clarity was purchased at the price of depth. He set forth very clearly the conditions for a Pareto Optimum. But then he referred to the fact that one could consider maximizing a welfare function. You started off with $U_1, \ldots, U_n$, utility functions of individuals, and you want to maximize $W(U_1, \ldots, U_n)$ but then there are a whole subset of the maximizing conditions that don't involve W—those are essentially the conditions that define Pareto Optimality. If I recall correctly, he was quite clear on distinguishing these concepts but he did have this W function and he gave a rather casual reference to Bergson. By this time, Samuelson's *Foundations* had been published and he gives a very full account of welfare economics in Chap. 8 which he bases on Bergson's paper.

So I gave a quick reply to Helmer: "Economists have thought about that and it's really explained by Bergson's social welfare function." "Oh, is that so," he said, "Why don't you write it up? I think it would be nice for us all to have an exposition of how the Bergson social welfare function settles this."

JK. So you just started to write this up.

KA. Well, of course, I dropped the *U's* which I never liked because I knew the *U's* were just disguises for *R's* for preference relations. I thought, while I was at it, I'd do an exposition starting from just the orderings. Then I started musing about what information is conveyed. Welfare comparisons could be regarded as a series of pairwise votes and I was obviously interested in elections so that just seemed like the natural language to use. One natural method of taking a bunch of *R's* and putting them together would be by pairwise comparisons by majority voting. And I already knew *that* was going to lead to trouble! So I figured, well, majority voting was just one of a very large number of possibilities, you just have to be more ingenious. I started to write various possibilities down.

JK. For example …

KA. I'm pretty sure the Borda method came to my mind, because that was a very well known method. I didn't know it was Borda, you understand.

JK. It was something you'd encountered before?

KA. That was very well known; it was a widely used custom. A club might do that. It was something that was done in practice.

JK. Did you ask any political scientists at Rand about voting procedures?

KA. I don't think at first they did have any political scientists. I don't know anybody I would have regarded as a political scientist. I don't think Rand was interested in traditional political science. Later I think they had more traditional political scientists but even when they did they had people interested in area studies, Soviet experts. I don't think they ever had theoretical political scientists.

JK. So this experimentation was totally isolated. You didn't ask anybody about rules.

KA. Right, but that also reflected me. It seems to me I was trying at some point to systematically go through all possible rules. I took some examples and then consider related examples. Beyond that, I can't tell you which rules I explored. I did grasp that, at some point in this procedure, that part of the point was I was only using information on the alternatives under consideration. But then that struck me as a very natural thing to do.

JK. Very crucial.

KA. That turned out to be *very* crucial. In fact, if I had realized how crucial it was, I might have been more disturbed. It seemed very natural. Afterwards, when I formalized it, I saw the importance of it. Now it was obvious enough that if you let *one* person make the decision, there isn't any particular problem. So I was assuming non-dictatorship. But, I think it was a lucky thing, I assumed non-dictatorship in a very strong form. You know at the start I was only looking at triples, because that had the essential problem. I was assuming there wasn't any dictator or any pairwise choice. And then you see, you need two or more individuals to be decisive. A little calculation shows that you can always produce orderings that violate this. It really is the Condorcet paradox restated. And then when I tried to extend it to more than triples, I retained this postulate.

So, in this first version, which I did show about a week later to Abraham Kaplan, there was this idea of showing that these conditions were incompatible. But I thought I had really put a lot of emphasis on the assumption that there was no dictator on any pair; so, I thought, well there must *be* a solution if you allow different dictators on different pairs. This seemed to be absolutely crucial to the argument, and I thought about this for awhile. I thought it would be easy to produce an example where you have a different dictator on every pair. Now I didn't think that was a suitable solution anyways, I thought an assumption of non-dictatorship was a correct assumption, even on one pair, so I wasn't too disturbed. But I thought, to clean out the exposition, to show exactly what was meant, I ought to produce an example where you have different dictators on different pairs. But no, if there's transitivity that means dictatorship sort of propagates itself. And finally one night when I wasn't sleeping too well, I could see the whole proof, you know after playing around with it for awhile. And that was a couple of weeks later, that I had the idea that the non-dictatorship condition could be stated in this much weaker way, that the whole ordering can't be determined by one dictator.

JK. What did you feel at this point?

KA. I felt this was very exciting. I thought *"This* is a dissertation." You know, it's a funny thing. One of my problems had been feeling that one has to be serious and every time I'd thought about these voting questions they seemed like amusing diversions from the real gritty problem of developing a good descriptive theory. And in some sense I still have a little bit of that feeling. But when I got the result, I felt it was significant. I really did. It clearly didn't conform to my preordained ideas about what was significant. I would have said a priori if somebody told me about this, my temptation would be to say, "Well, that's very nice, but what importance is it?"

But when I did it, I felt, yes, this is something. This was at least asking some very fundamental questions about the whole nature of social intercourse and particularly about legitimation of collective action.

This wasn't just a technical issue in game theory. The technical and the philosophical were intimately merged. My whole work in general, not only in this field but in others, has tended to deny the idea we can take off the technique and put it here and put the deep issues there. Some of the so-called technical issues are really of the essence of the so-called deep issues and you really can't separate them at all. Each one illuminates the other. In fact they fuse together and in some cases they're identical. And nothing can better exemplify this than social choice theory where the central issues and the technical issues were identical.

JK. Let's go down the list of some of the names you've acknowledged and tell us what they contributed. First the people at Rand: Abraham Kaplan.

KA. Well, he was one of the first persons to whom I showed the results. He was the only one who combined the philosophical side and at least some of the technical capacity to appreciate this. I don't recall getting anything specific from him, just an appreciation that it really was important.

JK. Youngs.

KA. I don't know why I thanked Youngs. Youngs was a mathematician with whom I discussed some issues of preference orderings and the like. I had been in close contact with him and I discussed some other aspects about preference orderings, really about individual preference orderings. I felt a kind of general intellectual debt.

JK. David Blackwell.

KA. Again, well, Blackwell was a genius. He and I worked very closely on other matters. He, Girshick and I wrote a paper on sequential analysis (Arrow et al. 1949). He did contribute one other thing. There was this chapter [in *Social Choice and Individual Values*] which has never been followed up. I was raising the question about if the orderings were restricted in some ways, when does the paradox exist. If you go to the extreme of single-peakedness we know the answer. So the question is, supposing you have *some* restrictions on orderings but there is a lot of freedom left. I had a result, where the technical point was when could you extend a quasi-ordering to a full ordering and it was Blackwell who told me about Szpilrajn's Theorem. However, unfortunately, my proof is not correct, because it suffers from the problem that Blau pointed out. I suspect the theorem is correct or some theorem like it is correct but nobody's ever stated it and I've never gone back to it.

JK. J. C. C. McKinsey.

KA. McKinsey was a very interesting fellow. He was the one who educated all of us to what game theory was all about. So the influence was indirect, but in a way it was there. He was a beautiful expositor. He was a logician of considerable power and had done some work earlier on the formalization of logic; he was a disciple of Tarski's. The whole game theory ambience, and therefore in particular McKinsey—and Blackwell for that matter on the technical side—were influential in setting the whole tone to this.

JK. The next names are from Chicago: Tjalling Koopmans, Herbert Simon, Franco Modigliani, T. W. Anderson, Milton Friedman, David Easton.

KA. The exposition of the book was developed in the next year back in Chicago. I presented the material over a number of seminars. I was grateful to these people because they thought it was a good idea, encouraged me and asked good questions; parts of the book are making clear points they found obscure.

Easton was a little different. He was the first political scientist I talked to about this. He gave me the references to the idealist position which was sort of the opposite idea. In a way the idealist position was the only coherent defense that I could see in political philosophy. It wasn't a very acceptable position, but it was the only one that had at least a coherent view of why there ought to be a social ordering.

JK. Why did you call it a "Possibility" Theorem?

KA. That was Tjalling's idea. Originally I called it an impossibility theorem, but he thought that was too pessimistic! He was my boss and a very sweet man, so I changed it for him.

JK. There was a meeting of the Econometric Society where you presented these results.

KA. I guess I must have presented it at the December, 1948 meeting.

JK. Who was there and what was the reaction?

KA. I remember Larry Klein was in the chair and Melvin Reder was reading another paper at the same session. My recollection is that there were 30 or 40 people in the room. I distinctly remember that in the audience was this contentious Canadian, David McCord Wright, who objected because among the objectives, I hadn't mentioned freedom as one of the essential values in social choice and apparently he went out of the room saying that Klein and Arrow were communists—this was quoted to me at least by Kenneth May who was also present.

I thought under the circumstances, I got a pretty good reception. I don't think anybody said "We've seen a revolution before our eyes," but it *was* taken as a serious contribution. I wonder why it was accepted so well. There really was no resistance. It made my reputation.

There had been, of course, a fair amount of controversy about the foundations of welfare economics, beginning with papers by Harrod, Hicks, then Kaldor, then the long chapter in Samuelson's *Foundations*, then Scitovsky in 1941 with intersecting community indifference curves (Scitovsky 1942). So unease about the foundations of economic policy were there. So the debate was serious—people were already concerned about these things.

Right after the summer I developed this, on the way back to Chicago, I stopped at Stanford to be interviewed for a job. Girshick had meanwhile moved to Stanford to contribute to starting a Statistics Department there. He was their star and he wanted me to join him. The Economics Department there had already in fact made me an offer a year earlier.

JK. Based on?

KA. What happened was due to Allen Wallis, who was recently Chancellor of the University of Rochester and is now, at the age of 75, Undersecretary of State for Economic Affairs. Wallis had been a Professor at Stanford before the war, he was the first really major appointment they ever made. He didn't return after the war, but he was highly regarded and apparently they asked him for recommendations. He had worked with Wald during the war and Wald had spoken about me based on my work as a student at Columbia. It was very common at Stanford to appoint Assistant Professors who didn't have a Ph.D.; they assumed we would finish. I was appointed without a Ph.D. In fact, I got tenure without a Ph.D.

JK. Really?

KA. Well, I'm being a little technical, but my statement is technically correct. In those days you couldn't get your degree until your dissertation was printed. So I had these theorems and then sent my changed proposal into Albert Hart, who got kind of excited about it. I defended the thesis in January of 1949. Stigler was on the examining committee, Bergson had come to Columbia in the meantime and they put him on the commitee. Of course, Bergson was asking some searching questions but was very fair and did have a high opinion of me. But I had no real interaction—I sent in my dissertation and got it approved.

Hart was immediately enthusiastic. He said, "I don't really understand it fully, but it sounds like you're dealing with very important issues," and I've heard later that around Columbia it was held to be an exciting event. Of course I had been regarded a kind of a star student. In fact, one of the things that had worried me was whether I was just an eternal student.

Stanford had a custom where all initial appointments were for one year. They kept this idea since they frequently hired people without even interviewing them—because of the geography. Even Moses Abramovitz who transformed the Department was hired as a full Professor for just one year just before I came. The jet has ended all this. So I came as an Assistant Professor on a one year appointment. My thesis had been approved, but I couldn't get a degree until it was printed. This was just about the end of that era. I must have been one of the last people to come under that rule. In fact, while the thing was in the process of being printed, I received a notice that if I submitted a typed manuscript, I could get my degree immediately. But that was kind of expensive, to have somebody retype that all up. I really lost about a year on my degree by deciding to go ahead on the printing. But then they gave me tenure on the basis of this unpublished dissertation. You couldn't do it today. It would never be approved today.

Then it is interesting—the reception question. Hart, who didn't work in this line, was very enthusiastic; he had spoken, I gather very well, around the Columbia faculty. The people at Stanford were very impressed; essentially all I had for them to see was this work—I hadn't done anything else except trivial stuff. They were so impressed that by the end of the interview day—within a day they were ready to make me an offer. So it's interesting to get this reception from all sorts of people not logically trained, not mathematically trained. And when the book came out, it made a great success. It is a little puzzling—at the time I took it as that's what happened. But in retrospect I sort of wonder why.

JK. Let's go to Blau's discovery of a mistake in your proof (Blau 1957). *That was eight years later; was that a great surprise?*

KA. Yes, it certainly was. Blau was working that year at Stanford and showed it to me. I was surprised, but I knew right away that a universality assumption would correct things. It had seemed obvious to me that the non-dictatorship property was hereditary, but it wasn't. I still think there is a better correction than that one, but I've never really gone back to work on it.

Blau's was a very, very nice result. It didn't obviously change the basic impact, but it did show my little attempts at generalization didn't work. It's interesting to see how easy it is to make a mistake on things that seem so airtight.

JK. Let's leave the origins now. Over the succeeding 40 years, what were the most important developments in social choice theory?

KA. Some of the work that has cohered around the original question is mathematically interesting but not very relevant to the original field. The literature that depends on small numbers of alternatives is in this category. I think the alternative space must be taken to be very large.

I also have qualms about results like those of Kirman and Sondermann (1972), What do we learn from Kirman-Sondermann exactly? When you have an infinity of voters, then the axioms as I wrote them become consistent and you can produce voting systems. But they are consistent because in some sense the dictator has a different meaning; banning a dictator is no longer enough. As it turns out, if you have a sequence of decisive sets, each of which is properly contained in its predecessor, the intersection of the whole sequence is empty, but everyone in that sequence is then less decisive. In some sense, the spirit of non-dictatorship ought to rule that out.

Incidentally, there's a recent result I haven't had a chance to study by a former student of mine, Alain Lewis, who says if you confine yourself to recursive functions then the voting paradox occurs even with an infinite number of voters; in the strict sense, even with just the ordinary axioms. The examples Kirman and Sondermann use are non-constructive; something with cofinite sets is not something you can actually construct—you just show it exists. But since they are only examples, that's not a proof that there isn't a constructive procedure. But Lewis says he's given a proof and I have to study it. If I can understand it.

JK. What about the Gibbard-Satterthwaite (Gibbard 1973) *results?*

KA. Gibbard's work was a bombshell. That was very exciting. I didn't know about Satterthwaite's work for a couple of years, but it was very much the same thing. I had taken the liberty of abstracting from manipulability in my thesis and I never went back to that issue. What's surprising is not really that there is an impossibility of non-manipulability, but that the issues should be essentially the same. That strikes one as a remarkable coincidence.

I still find it surprising and feel that we might not have the right proof. Somehow you feel that if you had the right proof it would be obvious. But then I thought that about my work, too. My impossibility theorem ought to be totally obvious when looked at the right way. Yet every proof involves a trick. Maybe not a big trick; I don't think it's a mathematically hard theorem. But somehow if you had the right way of approaching it, it should be trivial. Yet, no matter how you present the proof, and they're all pretty close to equivalent, its not yet trivial. For example, when ultrafilters came in, I thought, Aha!, this is a beautiful way of showing it. But it turns out that to prove the decisive sets form an ultrafilter involves essentially all the original calculations.

JK. Still, it's a nice approach conceptually.

KA. I don't know. I'm less convinced than when I first saw it. It has the advantage of referring to a known body of knowledge. But this is a body of knowledge which is somewhat technical. You're bringing in a fair amount of technical apparatus and it ought to pay for itself somewhere, if I may use an economic approach. It ought to pay for itself in making the proof trivial. But in fact you need just about every step in the original proof to show that the issue *is* one of a fixed ultrafilter. So therefore, why bring in all this apparatus? I was a little surprised by how little you get from all that apparatus. I can't help feeling there's some way out of it. In the same way, I always feel the Gibbard-Satterthwaite result should be more transparent than it is. But maybe it can't be done.

JK. What about Sen's Paretian liberal approach (Satterthwaite 1975); *does that interest you?*

KA. I thought that was stunning and penetrating to a very important issue. But …why do we have rights? What I am after all is a kind of utilitarian manqué. That is to say, I'd like to be utilitarian but the only problem is I have nowhere those utilities come from. The problem I have with utilitarianism is not that it is excessively rational, but that the epistemological foundations are weak. My problem is: What are those objects we are adding up? I have no objection to adding them up if there's something to add. But the one thing I retain from utilitarianism is that, basically, judgements are based on consequences. Certainly that's the sort of thing we do in the theory of the single individual under uncertainty; you make sure utility is defined only over the consequences. I view rights as arrangements which may help you in achieving a higher utility level. For example, if you are much better informed about a certain choice, because it's personal to you and not to me, I don't really know anything about it, I should delegate the choice to you.

JK. You don't want to allow preferences over processes?

KA. Well, one of the things I fear is emptiness. You put preferences over enough things, then anything that happens can be defended. It destroys the idea of discourse. Of course, it is a delicate issue, you can always say, of any particular process that it is specially privileged. You could take Nozick's point of view; you can have an absolute preference about certain processes. For example, we have a property system; if you and I make an agreement about anything within our property rights, that just fixes it, period. Now I've got to admit Nozick's courage is good. Suppose somebody invented a cure for cancer and allowed it to be used only at an extremely high price. Nozick says: No problem. Most everybody else would regard that as a fatal counterexample, but Nozick has the courage of his convictions. But that's a strong example of preference over processes. Most of the people who are advocating rights are very different, like Dworkin. They tend to support so to speak left-wing rights rather than right-wing rights, but once you grant that, who settles what rights are legitimate? The consequentialist view—I won't say that fully settles it either, but at least you have something to argue about. So this is why I'm a little unsympathetic to the rights issue—everybody just multiplies the rights all over the place and you get total paralysis.

Consider the consenting adult example—say homosexuality—and think about the concept of externality. Now why do we say intercourse among consenting adults should be allowed? One argues because there's no externality. But if I care, there *is* an externality. I actually allude to this even in the first edition of my book. It's just a rhetorical passage and doesn't enter the logic, but I mention that the concept of preference is just what everybody thinks their preferences are. Different people might have different ideas of externalities. I took the view that all preferences count. From the logical point of view, it doesn't matter; if you purify the preferences by rejecting the nosy preferences, the theorem applies to whatever is left. There is, of course, a technical problem in systematically combing out inadmissible preferences. Transitivity says you can't just look at separate preference pairs, you have to look at the whole system. That's what Gibbard's paper is really devoted to. Gibbard is not totally convincing, because there are some arbitrary choices in his elimination procedure; he doesn't make it compelling that his is the only way of doing it. It's just *a* way. It looked pretty devastating but it eliminated more than was really necessary.

I'm quite puzzled. People really care about consequences as they see them. If I'm really offended because people are seeing obscene material, well, I'm hurt. I really am hurt. I'm hurt just as much as if somebody blew smoke in my eyes—or whatever your favorite form of pollution is. Indeed a lot of people probably care much more. I really find it difficult to decide.

Unless somebody produces a logic of rights in terms of which we can *argue*, I really find the whole issue is unfocused. The reason why it is compelling is that there are at least some cases where we do feel strongly about the rights. It's not clear you can always reduce those to utilitarian considerations like information.

JK. One last question. What outstanding problem in social choice theory would you most like to see solved?

KA. Well, if I had to pick just one, it would be reformulating a weakened form of the independence of irrelevant alternatives which stops short of just dropping it completely. There are a lot of arguments used today, extended sympathy, for example, or the relevance of risk-bearing to social choice as in Harsanyi or Vickery (Sen 1970), that do involve, if you look at them closely, use of irrelevant alternatives. Suppose I'm making a choice in Harsanyi's story among totally certain alternatives. I somehow use preferences among risky alternatives as part of the process of social decision. We use a chain of reasoning that goes through irrelevant alternatives. It seems quite open to acceptance, not at all unreasonable, that these are useful. I would not want to rule out in an argument, a line of reasoning which goes through a chain of transitivity via an irrelevant alternative. And yet I don't want to be in the position of saying, well the whole thing depends on the whole preference ordering. My current feeling is that that is the most central issue—the most likely way of really understanding issues.

JK. Are you anticipating that if you allow chains of transitivity over irrelevant alternatives you will obtain a "*good*" *social choice procedure or are you expecting a deeper impossibility theorem?*

KA. I'm expecting—no, let me put it more cautiously—I'm *hoping* for a possibility result.

References

Arrow, K. J., Girshick, M. A., & Blackwell, D. (1949). Bayes and minimax solutions of sequential decision problems. *Econometrica, 17,* 213–244.

Black, D. (1948). On the rationale of group decision making. *Journal of Political Economy, 56,* 23–34.

Blau, J. (1957). The existence of social welfare functions. *Econometrica, 25,* 302–313.

Foley, D. (1967). Resource allocation and the public sector. *Yale Econ Essays, 7,* 45–98.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica, 41,* 587–601.

Hart, A. G. (1942). Risk, uncertainty and the unprofitability of compounding probabilities. In O. Lange, F. McIntyre, & T. O. Yntema (Eds.), *Mathematical economics and econometrics: In memory of Henry Schultz* (pp. 110–118). Chicago: University of Chicago Press.

Hicks, J. (1939). *Value and capital*. Oxford: Clarendon Press.

Hotelling, H. (1932). Edgeworth's taxation paradox and the nature of demand and supply functions. *Journal of Political Economy, 40,* 571–616.

Hotelling, H. (1935). Demand functions with limited budgets. *Econometrica, 3,* 66–78.

Karman, A., & Sondermann, D. (1972). Arrow's theorem, many agents and invisible dictators. *Journal of Economic Theory, 5,* 267–277.

Kenney, J. F. (1939). *Mathematics of statistics* (2vols.). Wokingham: Van Nostrand.

Lange, O. (1942). The foundations of welfare economics. *Econometrica, 10,* 215–228.

Lange, O., McIntyre, F., & Yntema, T. O. (Eds.). (1942). *Mathematical economics and econometrics: In memory of Henry Schultz.* Chicago: University of Chicago Press.

Modigliani, F., & Miller, M. H. (1958). The cost of capital, corporation finance, and the theory of investment. *American Economic Review, 48,* 261–297.

Popper, K. (1938). A set of independent axioms for probability. *Mind, 47,* 275–277.

Samuelson, P. A. (1941, 1942). The stability of equilibrium. *Econometrica 9*, 97–120; *10*, 1–25.

Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions. *Journal of Economic Theory, 10,* 187–217.

Schroder, E. (1890–1905). *Vorlesungen über die Algebra der Logik* (3 vols). Leipzig: B. G. Teubner.

Scitovsky, T. (1942). A reconsideration of the theory of tarifs. *Review of Economic Studies, 9,* 89–110.

Sen, A. (1970). The impossibility of a Paretian liberal. *Journal of Political Economy, 78,* 152–157.

Tarski, A. (1941). *Introduction to logic and the methodology of the deductive sciences.* New York: Oxford University Press. Appeared originally in Polish, 1936, and was translated into German, 1937.

Wold, H. O. A. (1943, 1944). A synthesis of pure demand analysis, I—III. *Skandinavisk Aktuarietidskrift, 26*, 85–118, 220–263; *27*, 69–120.

John C. Harsanyi

Claude d'Aspremont and Peter J. Hammond

Cd'A: *Reinhard Selten, in his contribution to your* Festschrift, *concluded with some biographical remarks, presumably accurate.*[1] *They mention that in 1947 you received a doctorate in philosophy. What was exactly the topic of your dissertation? Did you use some of it in later work?*

JCH: Well, it's certainly true and the topic of my dissertation was the logical structure of philosophical errors.[2] By that I meant that philosophers have some valid insight into some general principle, but then they go much too far in pressing it, which causes serious errors and that's probably their main error. But it is not a very deep observation. And of course I didn't have much opportunity to use this later.

PJH: *Except in so far as you became interested in philosophy and philosophical work generally.*

JCH: Well I was always interested in logic and the philosophy of mathematics. But I never had the opportunity to study it, although I wrote later a paper and sent it off [to a journal].

PJH: *Is that later paper published?*

[1]See Selten (1992, pp. 430–432).

This chapter was previously published in the journal *Social Choice and Welfare* (2001) 18:389–401.

C. d'Aspremont (✉)
Center for Operations Research and Econometrics, Université Catholique de Louvain, Voie du Roman Pays 34, B-1348 Louvain-la-Neuve, Belgium
e-mail: Claude.dAspremont@uclouvain.be

P. J. Hammond
Department of Economics, Stanford University, Stanford, CA 94305-6072, USA
e-mail: peter.hammond@stanford.edu

[2]See Harsanyi (1947).

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_3

JCH: Yes, it was published in *Erkenntnis*.[3]

PJH: Selten's *biographical remarks also suggest that, once in Australia, you started to study economics in order to enhance your career opportunities, and that some early papers concerned international trade.*[4] *Would you like to elaborate?*

JCH [In Australia] of course they didn't accept my Hungarian degrees. But I could work as a factory worker, and in the evening I took an evening course at the university. As I said, they didn't accept my Hungarian degrees. But they gave me very generous allowances for my previous university work, which was in pharmacy because my father was a pharmacist, and then in philosophy and psychology and sociology, the three subjects of my dissertation. So, I did study economics, but I don't remember ever having written a paper on international trade. I did have to write an essay,[5] which was part of my MA degree in Economics, and it was about the theory of the firm, something about the investment policies of an innovative firm.

PJH: *And this was at the University of Queensland?*

JCH: No, the University of Sydney. I enrolled at the University of Sydney as soon as we arrived there. We arrived in the last few days of 1950 and then I enrolled when the school started, I think probably in March.[6] I had to do two years of studies although the normal curriculum is four years, but they permitted me to skip the first two years. As a result, I never learned some part[s] of economics. So that was what happened and I got my degree at the end of '53 and I got a university job at the University of Queensland early in '54.[7]

Cd'A: *Is it fair to say that you chose to visit Stanford in 1956/57 so that you could work with Ken Arrow, even though by then you were moving on to game theory?*

JCH: Yes, that's very true, and I really did it as a result of a mistake, although it was a very favourable one. I read a semi-popular article of Ken which was published in a collected volume on game theory.[8] He made game theory look so interesting that I decided I would like to work with him and I thought that may be his main line of study. But it turned out of course that this was very much a sideline. But it was very good to work with him. I learned a lot from him. Even in game theory, of course.[9]

---

[3]Presumably as Harsanyi (1983). This paper, however, makes no mention of his dissertation. Also, to judge from the citations and acknowledgements, it must have been largely written in the years immediately prior to publication.

[4]See Selten (1992, p. 431).

[5]Probably his masters' thesis, in fact—see Harsanyi (1953b). One resulting publication would seem to be Harsanyi (1954).

[6]Of 1951.

[7]From March 1951 to the end of 1953 is three years according to the Australian academic calendar. It seems likely that two years were spent completing the requirements for a bachelor's degree, and a third year was spent doing a master's degree, possibly including a thesis as mentioned above.

[8]Kenneth Arrow's very plausible suggestion is that the relevant paper must be Arrow (1951)—see especially Sect. III (pp. 139–147). Actually, in the whole volume, just one section of Arrow's paper is devoted to game theory! So John Harsanyi presumably means that the article, rather than the volume, is on game theory.

[9]The resulting Ph.D. thesis is Harsanyi (1958). Apart from some details of an existence proof, and also with the exception of Appendix II, which contains an economic application in the form of a

PJH: *Are there any significant people you remember having a positive influence on your career at the University of Queensland or later at Stanford, apart from Arrow?*

JCH: At the University of Queensland, I was very lucky because there were two or three young economists among my colleagues and they were very friendly and some of them were extremely bright. Two of them got first class, second division, or whatever it's called,[10] which is not the best degree, but a good degree, and we had many interesting discussions. But I don't think professionally I benefited very much because my interests were different, and in Stanford it so happened I learned a lot from a short course on game theory given by Herb Scarf, who was there for a short time.

PJH: *Do you remember the names of these people in Queensland? Did they do anything interesting later?*

JCH: Well, yes, two names, which I remember most were Austin Holmes and Ron Lane. Both of them were interested, roughly speaking, in public finance, and Austin Holmes later derived a very ingenious scheme of how to make use of banking data in computing monetary circulation and other figures, and he became later a high-ranking official of the Commonwealth Bank, the national bank.[11] But he died very young, and Ron Lane is still alive. He also wrote textbooks, I think, in public finance.[12]

Cd'A: *What attracted you to social choice theory and what motivated you to set out your version of utilitarianism based on choice under uncertainty, in what Rawls later called the "original position"?*

JCH: Well, it's interesting. It was again a result of a chance observation by two economists. I read an article of Friedman and Savage and they pointed out the importance that von Neumann–Morgenstern utilities have in the study of risk-taking behaviour. But then they had a side remark which says[13]: Sure, this is a cardinal utility concept – but it doesn't mean that it can be used in a traditional way in welfare economics or in ethics. I wrote back a rejoinder in which I pointed out how it can be used. And that was my '53 paper, I think two pages in the *Journal of Political Economy,* where I started with this model in which you assume you have the same

---

numerical example involving three firms, this was published in a very slightly amended form as Harsanyi (1959).

[10]This should probably be second class, first division—commonly known as a II.i. Usually, in Commonwealth countries following the British system, the first class is undivided and is the best possible degree, so a second class, first division matches the following observation better.

[11]In 1960, this split into the Reserve Bank of Australia for central banking, and the new Commonwealth Bank of Australia for commercial and savings banking.

[12]We were unable to find records of any such textbooks, even with the very kind assistance of Tanya Ziebell of the Economics Library at the University of Queensland, arranged through Dr. Ghanshyam Mehta. Two relevant works, however, are Lane (1975, 1995).

[13]In footnote 11 of Friedman and Savage (1948, p. 283), they criticize the work of Vickrey (1945) and write that "it is entirely unnecessary to identify the quantity that individuals are to be interpreted as maximizing with a quantity that should be given special importance in public policy."

probability of ending up in anybody else's place.[14] And then I continued a similar paper in 1955 in which I discussed what people now call the aggregation model.[15] So that started me on this subject, but of course these are not strictly speaking game theoretic subjects. I started being interested in game theory because I read the major John Nash papers. They were available in the university library in Sydney. And I read, among other things, *Econometrica*, and there were three or four brilliant articles of John Nash.[16] And I was particularly impressed by the fact that, of course, as I learned a few months earlier in my university studies, there is no unique solution to bargaining problems in economics. You have an upper limit and a lower limit, and if they [i.e., the bargaining parties] agree at all, they must find a price between these two limits. But there was no theory whatever about whether it will be closer to the lower limit or the upper limit and what will be decided. Nash implicitly answered these questions and it interested me very much.

PJH: *If I may come back to the '53 paper, which as you say was a response to Friedman and Savage. As I recall that footnote on the fact that cardinal utility had no ethical significance, in some sense that was probably a response to earlier work by Lerner and Vickrey. Is that right?*[17]

JCH: Could have been, I don't know. It wasn't at least obvious to me and I didn't actually know of Vickrey's work, but when I published this work and went to Stanford, Ken Arrow, of course, pointed out to me that he anticipated some of my ideas.

PJH: *But, in Australia, you weren't aware?*

JCH: Yes.[18]

Cd'A: *So did you have some kind of comment by Savage or Friedman?*

JCH: No, I think that I didn't have any comments but I am sure that Friedman and Savage, at least Friedman, didn't really appreciate this idea. He didn't agree with such abstract models.... [Some words lost in the general laughter.] However, I think—I am sure the editor asked them whether they should publish it—and they said, "Of course, yes". Because it still was about their work.

PJH: Now *you also had an early article on welfare economics with variable tastes, and as I recall it seems to have been motivated by a strange review of Arrow's* Social Choice and Individual Values *written by Schoeffler and published in the* American Economic Review.[19] *That's accurate?*

---

[14] Harsanyi (1953a).

[15] Harsanyi (1955).

[16] Four, in fact, but only two in *Econometrica*, both of which concerned bargaining and the "Nash solution", whereas the others dealt with non-cooperative "Nash" equilibrium. See Nash (1950a, b, 1951, 1953).

[17] See Lerner (1944) and Vickrey (1945). As pointed out above, Friedman and Savage (1948) do make a point of criticizing Vickrey.

[18] Meaning "Yes (I was not aware)" rather than "Yes (I was aware)", presumably.

[19] See Harsanyi (1953–1954), which (p. 206, footnote 1) has three paragraphs discussing Schoeffler (1952). However, this footnote could well have been added at a late stage of the writing process, perhaps even as a result of the interchange with Ursula Hicks mentioned below.

JCH: All I remember is that the editor of the *Review of Economic Studies* was Ursula Hicks, and I had a long correspondence with her because she thought that my ideas that you can compare your utility before your tastes change with the one afterwards.… This is [making] interpersonal comparisons of utility, and that's unorthodox. But I convinced her that it isn't so … [phrase inaudible]. She published it.

PJH: *So like all powerful and novel work, the papers on cardinal welfare, especially the paper in 1955, have drawn criticism. Particularly noticeable, perhaps, were the comment by Peter Diamond in 1967, and the paper in 1968 by Prasanta Pattanaik. We believe our readers would like to know what you think of their criticisms.*[20]

JCH: Well, I did read Diamond's criticism and I think I answered it about at the same time [as] when I wrote my article about the maximin principle, in which I criticized Rawls.[21] I didn't feel that Diamond's criticism was valid and I remember I had a very complicated argument based on the idea: Imagine a society in which it is an old custom that they exchange the babies of people randomly so that every baby ends up in a different family, and I don't remember the detailed argument, but I thought that this refuted some of the contentions of Diamond.[22] But I never read Pattanaik's paper so I haven't answered it.[23]

Cd'A: *Recently it has been suggested that John Rawls would justify his difference principle by the claim that, in the original position, people would be extremely risk averse and so maximize the minimum possible utility level they could experience. If Rawls were to abandon this hypothesis, and propound a theory based instead on expected utility, would most of your disagreements be resolved?*

JCH: No, I don't think so. One disagreement which would remain would be his claim that ethics should be based on giving absolute priority to the interests of the least advantaged social group, which I thought, as I mentioned in my paper the other day,[24] leads to rather extreme conclusions. And so that disagreement would remain. I also would keep on being critical of Rawls's attitude of creating a new absolute priority principle for every particular reason. I don't think this is a sensible thing to do. Again I mentioned this in my paper. And finally, I am very unhappy about his concept of a system, what he calls a "system of democratic equality", in which he essentially argues that people who do good things for society usually don't deserve any moral credit, or certainly not material rewards for it, because their talents are, of course, results of genetic luck. Which is of course true. But I would think that when people use their native talents and develop them further and use them for mutual benefit, that does deserve some social, I mean moral, credit. But I was particularly

[20] See Diamond (1967) and Pattanaik (1968).

[21] See Harsanyi (1975a or b).

[22] See Harsanyi (1975a, pp. 316–317) or (1976, pp. 69–70)

[23] It seems he never responded to Pattanaik and may genuinely not have noticed the paper when it appeared. Approximately 18 months after the interview, he sent one of us (PJH) a fax which made it clear that he had fundamental disagreements with Pattanaik's paper, but that he preferred not to see them elaborated in the edited version of this interview.

[24] See Harsanyi (2008).

surprised … that he argued that apart from … having some talents, you need very often some kind of moral qualities which he describes as good character. And they enable you to work harder and to make greater efforts than people without good character could do, and that doesn't give you moral credit. I thought if anything gives you moral credit, it is exactly that you developed a good character and used it for good purposes. I think this [i.e., this Rawlsian system of democratic equality] comes very close at least, if it doesn't [actually] get into hard determinism, which I think would destroy morality, … and strange that a moral philosopher should use a theory which would really destroy the moral value of our actions. If we can't get any moral credit or discredit for them, this would happen. And of course, his main arguments are based on the implicit assumption that decent people will help him and other people to create a fairer and juster society. But if people might be apparently very fair and just but really have no moral merit in doing so, then this whole thing collapses.

Cd'A: *What about primary goods?*

JCH: I doubt very much that you can have a sensible index. And I doubt very much that under Rawls' own assumptions such … an index, which would be an economic index, should have moral significance.

PJH: *And you really don't think that these objections to Rawls would be somewhat muted at least if he moved to an expected utility position in the original position?*

JCH: Of course. Of course, I find the maximin principle simply absurd as a decision rule. So clearly if he … somebody … gives it up …, that's a benefit. But …

PJH: *But there'd still be serious differences?*

JCH: Yes, yes. Really, I think that this is a technical mistake, if it is a mistake, in decision theory. And I much more, much more object to his basic moral views than to decision-theoretical views.

PJH: *So, all readers will know that you were awarded the Nobel Prize in 1994, jointly with John Nash and Reinhard Selten, for your contribution to game theory. And that a most important contribution to game theory is your treatment of games with incomplete information played by Bayesian players.*[25] *These games have very many applications, of course, but we wonder if you could highlight some relation to social choice and welfare.*

JCH: Well, if you consider the theory of rule utilitarianism, you have to ask yourself the question: How much social utility would be contributed by a particular moral code which society accepts? And that is hard to predict if you don't know the types of the subject of the moral code. Especially in modern societies, … even if utilitarianism became the leading philosophy, surely there would be many people who, for religious reasons or for other reasons would use a different moral code. And there would be some people who don't care about any moral code whatever. Now, such people's behaviour would be hard to predict without trying to estimate the probabilities that they are of one type or another, so I think such questions to be useful.

[25] See J. C. Harsanyi (1967–1968).

PJH: *Do I understand that you are in the process of writing a book in which these themes may be addressed?*

JCH: Well, I was. I was when the Nobel Prize news came out in October '94. I was writing a volume but I didn't have any opportunity to go back to it up to now. I hope I will be able to.

Cd'A: [amid laughter] *A Nobel Prize is very time consuming.*

JCH: Let me mention that—though the Nobel Prize committee mentioned this, and this certainly is one of my contributions, the theory of games of incomplete information—I also did, I think what is worthwhile work on Nash's theory and on … on the generalization of the Shapley value to games with … non-transferable utility.

PJH: *You mean the Nash bargaining solution?*

JCH: Bargaining, yes, and let me just mention as a curiosity, that when I read Nash's work, I then tried to generalize it to *n* persons, and what came out of it was, of course, the Shapley value. And then I sent it to the *American Economic Review* and got the answer: This looks like an interesting paper, the pity is that this solution concept was developed a few years earlier by Shapley.

Cd'A: *Now, if we come back to incomplete information, the so-called "Har-sanyi doctrine"—I think Aumann gives the name—requires consistent beliefs (differences between individual probabilities are explained by differences in information). Now, the impartial observer approach to utilitarianism requires a fundamental (or inter-individual) utility (differences between individual utilities are explained by differences in causal variables). Is this just an analogy? Or, if not, and then you think that incomplete information should be introduced, don't you find [that] the "preference revelation" problem would appear? What do you think?*

JCH: Well, I think there is certainly an important formal similarity. Both are based on the idea that any important general theory should try to explain differences between people, if they are relevant to the theory. So one should explain why people have different priors, and the so-called "Harsanyi doctrine" would try to do that. And then the other question is that people obviously have different preferences and this should be explained somehow, and I could explain it in terms of what I called causal variables. But once you do this, I don't think that the preference revelation problem really disappears because even if you have a general theory which is purely formal,… we of course don't know how specific differences cause specific types of preferences (I mean differences in causal variables called preferences). So this is only a formal theory. But [in] the formal theory, even if we knew the actual rules which connect the causal variables with preferences, it wouldn't follow that we could easily decide what the true preferences of another person are. You still would have, the … what do you call it, revelation problem. And so I don't think that my causal variables theory, whether it's correct or incorrect, would do much. For instance, suppose that we observed that people educated in particular universities pick up a particular ideology, and this affects their preferences very much. What would we do then?… Even [if] you could predict the person's preferences by knowing that he studied at Princeton, to some degree he might be lying about where he studied just as much as he might lie about his preferences.

Cd'A: *And do you think that this poses a problem with respect to the original position and the foundation of morality?*

JCH: Well, I think that you can set up a moral code for society without knowing exactly how many people are just pretending to have certain preferences and how many … [really do have them]. That, of course, does cause a problem.

PJH: *Well, it imposes extra constraints on what is feasible, presumably?*

JCH: Yes, yes, it certainly would.

Cd'A: *Now, you have mentioned your contribution to bargaining, and maybe I would like to return to that, because this is a major part of your general theory of rational behavior, as described in your 1977 book.*[26] *But it is only applied to individuals deciding on practical issues from a personal viewpoint. Did you ever consider applying it from a moral viewpoint on ethical issues?*

JCH: No, I had a feeling that there are some theoretical reasons why one shouldn't do this because bargaining models are meant to be models of real-life bargaining … and bargaining depends on the relative power positions of the parties. So, a good bargaining model won't predict what the fair solution of the bargaining would be, but what the realistically expected solution would be. And, on the other hand, I think in ethics, you shouldn't start from the assumption that some people are strong, some people are weak, and those who are weak should make major concessions to those who are strong because I really [think] fairness should decide this and fairness shouldn't depend, or at least only marginally, on the question how much power they have. I think the English distinction between right and might comes in. You should distinguish clearly between might and right. And those … many, many game theorists did interpret even Nash's bargaining solution as a fairness doctrine, but I think this is a misunderstanding.

Cd'A: *And even the Shapley value was interpreted like that?*

JCH: Yes, surely.

PJH: *Now I suppose that the contractarians tend to use bargaining models in clear opposition to your approach. So, you would say that they are using them inappropriately?*

JCH: Yes, I wrote a book review of a philosopher who uses … a bargaining approach.

PJH: *Gauthier?*

JCH: Gauthier, yes.[27] And I think there are many weaknesses in it, and this is one of them.

PJH: *One of your most celebrated arguments in ethics concerns rule utilitarianism, which advocates choosing general rules that maximize expected total utility, rather than acts. So, one way of distinguishing a rule like voting, even at some personal cost, over an act (like abstention)*[28] *seems to be that a good rule is one*

---

[26]See Harsanyi (1977a).

[27]See Harsanyi (1987).

[28](PJH) This is what I had meant to say. The tape shows I actually said "abstinence", following a rough written draft of the questions we planned to ask. In any case, the question was not misunderstood.

that is beneficial if most people follow it. Is this a correct interpretation? And, since a good rule may also allow exceptions, should it be a requirement for a particular (lazy) person to vote, even if it would be time-consuming?

JCH: Well, I think that first of all, the government can do a good deal to get more votes out than they usually do. And there have been recently some measures adopted to make voting trivially easy, and that … one could hope … will increase voting.

PJH: *Voting, or registering to vote?*

JCH: Registering, maybe registering. But of course in California, for instance, you can always vote without any extra expense. Be an absentee voter if you want to, I mean if you don't want to stand in line.

Cd'A: *In Belgium, voting is compulsory.*

PJH: *And in Australia.*

JCH: Oh, is that so? So, I think it's a good idea to encourage people to vote. Though, of course there are counter arguments because one might say that those who wouldn't vote otherwise,… they don't vote because they are absolutely uninformed about the issues. Maybe to get what they are saying is the "asinine vote" is not a good idea.[29] But in any case, in principle, of course, one should vote. And if the law requires, at least if the social customs require, then one should vote. But of course this is subject to the general rule that if you have some minor duties, if complying with them would cause you hardship, then you are basically exempt.

PJH: *Now, I asked the question about voting because I remember early papers on rule utilitarianism where this is an example you had used.*[30] *Now it seems you prefer to discuss promise keeping. Is that right?*

JCH: Yes. Yes. Well I used it [i.e., the voting example] because it was easy to connect with a game-theoretical equilibrium point. And it has interesting properties. But I'm not sure that for ethical purposes, this is the most informative example.

PJH: *Right, but it does capture the essence?*

JCH: Yes, it does.

PJH: *So, finally, what would Harsanyi the game theorist say about resolving the prisoner's dilemma by having each prisoner follow the utility increasing rule of not confessing?*

JCH: Well,… I would assume that normally a prisoner in that situation has a moral duty to confess if he's guilty. But on the other hand, it's a common experience that our self-interest points one way and our moral duties point the other way. And, of course, depending on the situation, it might be a sort of pardonable sin not to admit that they are guilty, though in many cases they should, because otherwise some other person might be found guilty. So, it's just the typical case that self-interest is one way and moral duty goes the other way.

PJH: *Now, that reply, I think, is taking the prisoner's dilemma literally.*

JCH: Yes.

---

[29] Here the recording is too indistinct for us to be sure that "asinine vote" is John's exact phrase, but it is our best guess and certainly conveys rather forcefully what we believe he meant.

[30] See Harsanyi (1977b), especially Sect. 7 (pp. 38–41). The previous section of this paper discusses promise keeping, in fact.

PJH: *But as you are well aware, so many works use it as an example of conflict between cooperative and non-cooperative behaviour. So, if you thought of one of these more general situations, is the answer similar?*

JCH: Well, no.… Usually,… if there is any moral duty involved, of course, it is to cooperate. And the only problem is that cooperation might be unstable in such situations. As … was very clear during the cold war, there were some pacifists who argued that the United States should disarm if necessary unilaterally, which I could never agree to. But, on the other hand, clearly … we were more lucky than wise to avoid a catastrophe. And, of course, it was very good that we could avoid it. But, … Rapoport always argued that we should unilaterally disarm. And … [on] the whole, he was very hostile to the very concept of Nash equilibrium because he thought that would justify the cold war situation.[31] But,… clearly your moral duty may be to disarm if it's possible safely, possibly only by mutual agreement, but it can be only crazy to disarm unilaterally, given the other side being what is was. So …

PJH: *So you should think carefully whether the other side is really going to follow the rule?*

JCH: Exactly, and really the problem is mutual trust—not [whether] one should trust trustworthy people, but to trust people who are obviously untrustworthy is crazy; it's not moral.

PJH: *So, is there anything else we should have asked you?*

JCH: I don't know. You are kind enough to spend a lot of time with me.

Acknowledgments The following is an edited and slightly abridged version of an interview which John Harsanyi kindly agreed should take place on June 19, 1996. This was during the Conference on "Justice, Political Liberalism and Utilitarianism in Honour of John Harsanyi and John Rawls" at the University of Caen. In particular, the footnotes, references, and text in square brackets have been added by the interviewers, and some hesitations or repetitions removed. Our warm thanks to Maurice Salles for arranging the interview room (as well as the conference), to Lucien Bézier for producing a sound recording, to Tina Corsi and John Weymark for transcribing that recording, and to Kenneth Arrow for some advice concerning the final editing process. Finally, our thanks also to Dr. Ghanshyam Mehta and to Tanya Ziebell for bibliographic research in the Economics Library and the Archives section [UQA S 135 Staff files (1911–)] of the University of Queensland, and especially for passing on information concerning the dates and titles of the two theses—Harsanyi (1947, 1953b).

References

Arrow, K. J. (1951). Mathematical models in the social sciences. In D. Lerner, & H. D. Lasswell (Eds.), *The policy sciences: Recent developments in scope and method* (Chap. VIII, pp. 129–154). Stanford: Stanford University Press.

Diamond, P. A. (1967). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy, 75,* 765–766.

[31] This refers to an old debate that Harsanyi had with Anatol Rapoport. See Harsanyi (1977a, pp. 276–8) and the work cited there—especially Rapoport (1966).

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of Political Economy, 56,* 279–304.
Friedman, M., & Savage, L. J. (1952). The expected utility hypothesis and measurement of utility. *Journal of Political Economy, 60,* 463–474.
Harsanyi, J. C. (1947). *The logical structure of errors in philosophical arguments* (Ph.D. thesis). Budapest University.
Harsanyi, J. C. (1953a). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy, 61*, 434–435 (reprinted in Harsanyi 1976).
Harsanyi, J. C. (1953b). *Inventions and economic growth* (M.A. Thesis). University of Sydney.
Harsanyi, J. C. (1953–1954). Welfare economics of variable tastes. *Review of Economic Studies, 21*, 204–213.
Harsanyi, J. C. (1954). The research policy of the firm. *Economic Record, 30,* 48–60.
Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy, 63*, 309–321 (reprinted in Harsanyi 1976).
Harsanyi, J. C. (1958). *A bargaining model for the cooperative n-person game* (Ph.D. thesis). Stanford University.
Harsanyi, J. C. (1959). A bargaining model for the cooperative *n*-person game. In A. W. Tucker, & R. D. Luce (Eds.), *Contributions to the theory of games* (Vol. IV, Chap. 17, pp. 325–355). Princeton: Princeton University Press.
Harsanyi, J. C. (1967–1968). Games with incomplete information played by 'Bayesian' players, I–III. *Management Science, 14*, 159–182, 320–334, 486–502 (reprinted in Kuhn 1997).
Harsanyi, J. C. (1975a). Nonlinear social welfare functions: Do welfare economists have a special exemption from Bayesian rationality? *Theory Decision 6*, 311–332 (reprinted in Harsanyi 1976).
Harsanyi, J. C. (1975b). Can the maximin principle serve as a basis for morality? A critique of John Rawls's theory. *American Political Science Review, 69*, 594–606 (reprinted in Harsanyi 1976).
Harsanyi, J. C. (1976). *Essays on ethics, social behavior, and scientific explanation.* Dordrecht: D. Reidel.
Harsanyi, J. C. (1977a). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.
Harsanyi, J. C. (1977b). Rule utilitarianism and decision theory. *Erkenntnis, 11,* 25–53.
Harsanyi, J. C. (1978). Bayesian decision theory and utilitarian ethics. *American Economic Review (Pap Proc), 68,* 223–228.
Harsanyi, J. C. (1983). Mathematics, the empirical facts, and logical necessity. *Erkenntnis, 19,* 167–192.
Harsanyi, J. C. (1987). Morals by agreement. *Economics & Philosophy, 3,* 339–373.
Harsanyi, J. C. (2008). John Rawls's theory of justice: Some critical comments. In M. Fleurbaey, M. Salles, & J. A. Weymark (Eds.), *Justice, political liberalism, and utilitarianism: Themes from Harsanyi and Rawls.* Cambridge: Cambridge University Press.
Kuhn, H. W. (1997). *Classics in game theory*. Princeton: Princeton University Press.
Lane, W. R. (1975). *Financial relationships and section 96*. Canberra: Centre for Research on Federal Financial Relations, Australian National University.
Lane, W. R. (Ed.). (1995). *Equality in diversity: History of the Commonwealth grants commission* (2nd ed.). Canberra: Australian Government Publishing Service.
Lerner, A. P. (1944). *The economics of control*. London: Macmillan.
Nash, J. F. (1950a). Equilibrium points in *n*-person games. *Proceedings of the National Academy of Sciences, 36*, 48–49 (reprinted in Kuhn 1997).
Nash, J. F. (1950b). The bargaining problem. *Econometrica, 18*, 155–162 (reprinted in Kuhn 1997).
Nash, J. F. (1951). Non-cooperative games. *The Annals of Mathematics, 54*, 286–295 (reprinted in Kuhn 1997).
Nash, J. F. (1953). Two-person cooperative games. *Econometrica, 21,* 128–140.
Pattanaik, P. K. (1968). Risk, impersonality, and the social welfare function. *Journal of Political Economy, 76,* 1152–1169.

Rapoport, A. (1966). Strategic and non-strategic approaches to problems of security and peace. In K. Archibald (Ed.), *Strategic interaction and conflict*. Berkeley: University of California, Institute of International Studies.

Schoeffler, S. (1952). Note on modern welfare economics. *American Economic Review, 42,* 880–887.

Selten, R. (1992). John C. Harsanyi, system builder and conceptual innovator. In R. Selten (Ed.), *Rational interaction: Essays in honor of John C. Harsanyi* (pp. 419–432). Berlin, Heidelberg, New York: Springer.

Vickrey, W. S. (1945). Measuring marginal utility by reactions to risk. *Econometrica, 13,* 319–333.

Paul A. Samuelson

Kotaro Suzumura

1 Introduction

Social Choice and Welfare has a tradition of interviewing pioneering contributors to welfare economics and social choice theory to keep their recollections on the formative stages of their seminal work, their current views on the past and present states of the art, and their perspectives on the agendas to be pursued in this branch of normative economics officially on record. Professor Paul Samuelson has been on the list of potential scholars to be interviewed for a long time in view of his enormously influential contributions to economics in general, and theoretical welfare economics in particular. Indeed, the purpose of these interviews would not be served unless and until we could interview a scholar "who before 1938 knew all the relevant literature on welfare economics and just could not make coherent sense of it," and is willing "to set the record straight as only a living witness and participant can (Samuelson 1981, p. 223)." In November–December 2000, this long overdue interview with Professor Samuelson finally took place in his office at MIT. It started from the list of preliminary questions I had submitted to him beforehand. Needless to say, he had much more to offer, which colored and enriched this interview. To facilitate the readers' better appreciation of the rich information provided by Professor Samuelson, I added a few footnotes and provided an extensive list of references so as to link Professor Samuelson's recollections with what the readers could usefully learn by reading the existing literature. It is in similar vein that I inserted some relevant passages from Professor Samuelson's and others' past writings into my questions to him so as to place this interview in better perspective. It is hoped that this added material does

This chapter was previously published in the journal *Social Choice and Welfare* (2005) 25:327–356.

K. Suzumura (✉)
Institute of Economic Research, Hitotsubashi University, Naka 2-1, Kunitachi, Tokyo 186-8603, Japan
e-mail: suzumura@ier.hit-u.ac.jp

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_4

not distract the readers' attention from the real and novel gems contained in this interview.

I am most grateful to Professor Paul Samuelson whose kind collaboration and generous sacrifice of his time made this interview possible. Thanks are also due to Professors Kenneth Arrow, Marc Fleurbaey, Peter Hammond, Prasanta Pattanaik, Maurice Salles and Amartya Sen, with whom I had many conversations over the years relating to the subjects discussed in this interview. My deep gratitude goes to Professor Nick Baigent who kindly read several drafts of this paper and gave me detailed comments which led to the improvement of substance as well as exposition of the final draft. Needless to say, nobody other than myself should be held responsible for any remaining deficiency of the eventual outcome.

2 Interview

KS (Kotaro Suzumura): Thank you very much for giving me this opportunity to interview you on behalf of the Society for Social Choice and Welfare. In Chapter 8 of your Foundations of Economic Analysis, you have given a brief, yet fairly comprehensive overview of the whole area of welfare economics at the time of your writing. At the risk of a slight overlap with what you have already explained there, let me begin by asking you about Arthur Pigou and his "old" welfare economics, and the subsequent advent of the "new" welfare economics.

2.1 On Pigou's "Old" Welfare Economics

KS: Several people including your former teacher, Joseph Schumpeter, in his History of Economic Analysis, as well as yourself in Chapter 8 of the Foundations traced back the origin of welfare economics far beyond Arthur Pigou's Economics of Welfare. However, John Hicks was technically right, wasn't he, when he asserted that "[if welfare economics] existed before Pigou, it must… have been called something else (Hicks 1975, p. 307)." What is your current view on the status of Pigou in welfare economics in general, and his "old" welfare economics, so-called, in particular?

PS (Paul Samuelson): Yes, but first, let me say this. Since you referred to Chap. 8 of the Foundations, you should be alerted to the fact that I prepared an enlarged edition of the Foundations in 1983. I did not change the text of the original edition, but I added the Introduction to the Enlarged Edition on the development since the original edition. Mostly, I do not consciously feel changed in my views on welfare economics after the 1938 clarification of the subject by Abram Bergson, but a reader who read Chapter 8 should perhaps also read the corresponding part of the Introduction to the Enlarged Edition, pp. xxi–xxiv, because I remark specifically there on the change in my thinking on welfare economics due to Harsanyi's 1955 article published in the Journal of Political Economy.

Let me now answer your question. I understand why Hicks made that sentence, but I think it is not a very useful or accurate sentence. We take nothing away from Pigou when we remember that he was a culmination of a long tradition called "moral philosophy." It was this long tradition that Pigou first crystallized into the Wealth and Welfare in 1912, and then into the Economics of Welfare in 1920.

I had a great admiration for Pigou. I thought that, in many ways, he was not only a faithful follower of Alfred Marshall, but he was also a more fertile developer of the Marshallian tradition than Marshall himself. He was too faithful to Marshall in his language, and he never disagreed with Marshall. A great philosopher, Alfred North Whitehead, came to Harvard in 1924 after retiring from the University of London. This is long after Russell and Whitehead's Principia Mathematica. Whitehead said to me: "Don't you think that Pigou was an overrated economist? Wasn't Foxwell a better man?" Herbert Foxwell had been the candidate who was expected to succeed "Marshall's chair when Marshall retired. But Marshall manipulated and contrived that the 30-year-old Pigou receive the chair. Since I am an honest man, I said to Whitehead: "No, I think Pigou was a much more important economist than Foxwell.[1]

I think Pigou was a very fertile economist. A sign of this was his assigning Frank Ramsey the task of solving the 1927 problem of second-best optimal excise taxes. He of course worked to a very old age, but I am much older than he was in his old age. I knew the Economics of Welfare well, including a fundamental mistake in it, which was not corrected until about the third or fourth edition. The mistake, which was common to Marshall and Pigou, was that Pigou believed that increasing cost industries should be taxed and the tax revenue collected should be used as a transfer subsidy to constant cost industries. He might have added: "… and to decreasing cost industries." However, decreasing cost industries were never handled properly by Marshall. Indeed, they are incompatible with laissez-faire competition and Marshall knew it. Thus, most of the thoughts which were worked out by my teachers' generation and by my own generation were in Marshall. He actually knew about it in 1890. John Neville Keynes, the logician and the father of John Maynard Keynes, was a friend of Marshall and a kind of an assistant, who warned Marshall: "Your consumers' surplus is wrong, and you will be picked on." But, instead of Marshall's going to work and going beyond his at best approximation under certain conditions, he never did do it properly.

I think Marshall was a great economist, but he was a potentially much greater economist than he actually was. It was not that he was lazy, but his health was not good, and he worked in miniature. Early on, in 1874, when Marshall deduced that alternative multiple equilibria of supply and demand could occur, he noted that this rebutted any notion that laissez-faire markets could be relied on to achieve maximal interpersonal well-being.

Pigou's mistake was pointed out by Allyn Young, then at Cornell, who was the teacher both of Edward Chamberlin and Frank Knight, in his Book Review of the

[1]Those who are interested in Herbert Foxwell's life, work and his relationship with Alfred Marshall are referred to Foxwell (1939), Groenewegen (1995, pp. 622–627 and pp. 670–679) and Keynes (1936).

Wealth and Welfare published in Quarterly Journal of Economics. He pointed out that, in modern language, it is Pareto-optimal for rents to rise in an increasing cost industry, and that should be built into the price that is paid under laissez-faire, because that is the socially optimal way of organizing the allocation of resources. Pigou and Marshall got confused on this, because they brought in the externality argument. Now externality is very important—the whole theory of public goods, I guess, is a case of externalities proper. But, in the absence of any externalities, if you have the law of diminishing returns, let variable labor be applied to fixed land, and when there is expansion of the demand for good vineyard wine, that raises the rent. If the marginal cost is rising, that should be built into the laissez-faire price.

Somewhat redundantly, Frank Knight made essentially the same point in his important article, "Some Fallacies in the Interpretation of Social Cost," published in Quarterly Journal of Economics. Dennis Robertson, a good Cambridge economist, also made essentially the same point independently in 1924. Isn't it interesting that Pigou never had corrected it until maybe the 1932 edition? I looked for Allyn Young's name in the 1932 edition. It is there, but not in this connection, but in connection with the discussion of depreciation, which is irrelevant for our present purpose. Isn't it interesting that this important and world famous scholar did not say: "I made a mistake. I corrected it, but I owe thanks to Allyn Young, and perhaps to Frank Knight and to Dennis Robertson."

Pigou was a much better expositor of Marshall's welfare economics, which was implicit in Marshall, than Marshall himself ever was. Pigou had a mathematical structure in his mind, but following Marshall's instructions, he kept it concealed. Also, Pigou did not attempt to go deeply into solving the troublesome problems of fundamentals.

KS: Could you please give us an example?

PS: For example, he says as a recurring theme that if there are very poor people in a market society who do not have the basic necessities of life, then it is manifestly, obviously desirable to make transfers from the more affluent people to the poor people. He has not, however, provided the kind of argument that Francis Edgeworth would have given. Like most classical economists, Edgeworth, a neoclassical economist, was an environmentalist who did not believe in the Darwinian superiority of certain people over others. John Stuart Mill, who had the highest IQ ever recorded, said in his autobiography: "If you had James Mill for your father-trainer, you would also have a high IQ." Thus, everybody has the same potentiality, and it is only the environment that makes them different. Likewise, Edgeworth would have shamelessly believed he could measure utility by the Benthamite procedure of measuring "minimum sensible" jolts of just-recognizable increments of pleasure. This is the theory of sensation like the Weber–Fechner Law. So, you draw the utility curve for each person, which is concave embodying the law of diminishing marginal utility. Thus, the extra dollar you get when you have 100,000 dollars of income is less important than the extra dollar you get when you have 10,000 dollars of income. I think that there is a layman's tendency to believe something like that. Most of the sharp solutions in classical welfare economics, or moral philosophy, are for special "Santa Claus" cases of symmetry among individuals. Take, for example, Kant's categorical imperative, or the golden rule in the New Testament: "Do unto others as you would have them do unto you." If you do not believe that human beings are the same, you may have to follow George Bernard Shaw and say that it is not right. Instead, you should say: "Don't do unto your neighbors what you would have them do unto you. Their tastes may be different from yours." The moment you do not have the same commensurable utility there is an end to the century-old welfare economics or moral philosophy. Thomas Nixon Carver as an over-age graduate student wrote around 1900 that: "You should equalize the marginal utility of the dollar between rich man and poor man by transfers through progressive taxation." Of course, he said: "I am abstracting from incentive distortions that would take place." Some background like this is, I think, implicit in Pigou. But he keeps it under the carpet rather than arguing it out.

2.2 On Robbins's Criticism of the "Old" Welfare Economics

KS: You have identified in your 1981 Bergson Festschrift article that there exist two distinct schools of the "new" welfare economics. One school is based on the compensation principles developed by Nicholas Kaldor, John Hicks, Tibor Scitovsky, Paul Samuelson and others[2] and the other school is based on the seminal concept of the social welfare function due to Abram Bergson and Paul Samuelson. The evolution of both schools was preceded by a harsh methodological criticism by Lionel Robbins against the epistemological basis of Pigou's "old" welfare economics. Would you please give us your personal recollection of the formative days of the "new" welfare economics?

PS: I think Lionel Robbins's essay in 1932 was not only important for my thinking, but was important for the whole profession. I cannot autobiographically relate the influence of Gunnar Myrdal's book, The Political Element in the Development of Economic Theory, which was originally published in Swedish in 1930. It was not available to us, but I think there were some quasi-nihilistic views in Myrdal about the conventional welfare economics, which were similar to those in Robbins. These views were not just on Pigou's "old" welfare economics, but on moral philosophy which predated Pigou's (1912) work. Henry Sidgwick would be an important example, and, of course, Jeremy Bentham and John Stuart Mill. But to the lay person, it seems natural that the same loaf of bread is less significant when you already have a hundred loafs of bread than when you have ten loaves of bread. You see it in the Old Testament when King David or somebody has been discussed. One of the prophets gives a parable. There was a King who invited a poor shepherd to dinner. They killed

[2][Paul Samuelson's footnote] Long before these writers, J. S. Mill had recognized that the winners from free trade had (transferable) gains larger than the losings of the losers. Implicit in what today we call "Pareto-optimality" is a parallel theme, and two decades before Pareto Edgeworth's 1881 "contract curve" construction shows that he understood when deadweight loss did or did not negate the ability to "make compensation." Already prior to 1930, my teacher Jacob Viner had anticipated the Kaldor–Scitovsky notions.

a lamb and made the meat for the dinner. The poor shepherd had only one lamb, and the King had a superfluously large number of lambs. In the course of the dinner, the King said: "By the way, what we are eating is your lamb." The fact that the story could just be told in that way means that every reader could understand that it was a terrible thing to do. That is what I mean by the "old" welfare economics. It can be utilitarian; it can even be hedonistic; it can be additively utilitarian; but importantly interpersonal commensurability is somehow taken for granted. Robbins was not the first to be critical of this tradition, but he was very important as he wrote beautifully, and the book was short. This is the reason why, I believe, the good element of Robbins's book had a very significant influence.

KS: What precisely do you mean by the "good element of Robbins's book"?

PS: It is that you cannot deduce and test norms by means of science, by measurement of the elasticity of demand, by any other means of the objective observations and model buildings in empirical science. You must put in a normative axiom to get out a normative theorem. This position of Robbins really goes back at least to the philosopher, David Hume. I am separating in Robbins's book a bad element from this good element. A different "bad element" was first edition Robbins's "Austrian-like" belief in a priori "truths."

KS: Abram Bergson and yourself were in basic agreement with this good element of Robbins's book, weren't you?

PS: Yes. But, you see, most economists resisted Robbins, because they thought there was nothing left by way of policy prescription, although Robbins never quite said that. He said: "As a scientist, I cannot tell you this. But, as a voter, I can tell you which way I would go." This view can be traced back to David Hume, who was a great reductionist. I was ripe for that, because when I was an undergraduate student at the University of Chicago and studying sociology, I had to read William Sumner's Folkways. Sumner was a very conservative economist at Yale, but he was a great sociologist. He studied all cultures and showed how what was right in one culture was wrong in another and you could not prove by the methods of science which of them was correct.

KS: Could you please tell us about the "bad element of Robbins's book" in more detail?

PS: The bad element of Robbins's book was that it was more Austrian than Ludwig Mises and Friedrich Hayek. Like Carl Menger and especially Ludwig Mises, Robbins believed in a priori thinking; you could solve all problems of the world in economics by introspection; economics is a deductive science; the deductive laws are much more powerful than any empirical laws and they are independent of almost anything empirical. I was taught something like that Austrian view at the University of Chicago. I was a very young student, but I was a good student. Aaron Director was my first teacher. He is the only man in the world who could truthfully speak of "my radical brother-in-law, Milton Friedman," because Milton's wife, Rose Director Friedman, is Aaron's young sister. Aaron believed that Hayek could reason out the business cycle in his 1931 book, Prices and Production, without any command of any important facts about the business cycle. The first edition of Robbins' essay is full of that view. It was modified a little bit later, but we should always attach importance to the first edition of anything, because in the history of ideas that is pragmatically the simplification which carries the greatest weight.

KS: In your 1981 Bergson Festschrift article, you described the initial thrust of Robbins's criticism as follows: "When Robbins sang out that the emperor had no clothes—that you could not prove or test by any empirical observations of objective science the normative validity of comparisons between different persons' utilities—suddenly all his generation of economists felt themselves to be naked in a cold world. Most of them had come into economics seeking the good. To learn in midlife that theirs was only the craft of a plumber, dentist, or cost accountant was a sad shock."

Could you please cite a few examples of economists who went through this period of turmoil?

PS: Take, for example, Abba Lerner, who was not that mathematical, but a very clear thinker and really very new in economics. He was 30 years old, I think, when he went bankrupt in the hat business. He wanted to know why he went bankrupt, so he went to the London School of Economics, which was a kind of a nightschool at that time mostly. He was a student of John Hicks, and he wanted to learn about Marxism, because he thought he could learn the necessary lesson there. Hicks has told this in some autobiographical writing. Lerner was unconservative in political philosophy, definitely not a libertarian, but, of course, he was not a Marxist. He became very anti-Marxist as soon as he understood Marx. I predicted that he would end up in the arms of Hayek, which proved in a degree true. But, still, he had social sympathies. I don't think John Hicks had any particular social sympathies. He came from the above average class structure in Britain, but not from the elite aristocratic structure as, say, Ian Little did. But, he really talked, like Frank Knight, much more in terms of his own personal economics.

Another example is Simon Kuznets. Interestingly enough, when the Nobel prize was first granted, at MIT we developed an informal custom of having each Nobel prize winner come to lunch and speak personally about his early history, but we were unable to continue the custom. Of course, the first two prize winners were Europeans, who weren't available. After me came Simon Kuznets, who studied economics first in Russia before the revolution, because he was interested in the Jewish problem and he thought economics must have a fundamental answer to it. He thought Marxism might give the needed fundamental answer, which is why he went to a commercial university instead of a classical university. But later he changed his opinion. Kuznets, like his contemporary expatriate Wassily Leontief, when I first knew them, seemed burned out by early experiences and eschewed politics and policy diagnoses. Only in later life did they become more liberal in the American sense. Jacob Marschak, a similar Menshevik, by contrast was uniformly interested in altruistic "good causes."

Likewise, in those days, many scholars started their study of economics in search of the good. For them, Robbins's criticism brought about a sad shock.

2.3 On the Advent of the "New" Welfare Economics

KS: The first step in the attempt to reconstruct welfare economics on the basis of ordinal and interpersonally non-comparable utilities in active response to Robbins's criticism was to develop the concept of "Pareto-optimality" and establish the so-called fundamental theorems of welfare economics. Could you please explain how these crucial steps were taken in the first place?

PS: When I was a student at the University of Chicago, where I was a direct student of Jacob Viner in the classroom, and an indirect student of Frank Knight, I could not learn why price should equal marginal cost. Even when I got to Harvard in 1935, I went around asking everybody: "What is the proof that this is so?" Of course, I did not know the 1892–1893 work of Vilfredo Pareto in which he essentially shows that a perfectly competitive equation system gives you the necessary and sufficient condition, not for ethical optimality—he was always a little slippery on that problem—but for what came to be called Pareto-optimality so that there is no avoidable deadweight loss. I think I had most to learn from Abba Lerner, although I, of course, worked it out for myself. If I had had perfect teachers, they would have known the Pareto work; they would have known Enrico Barone and what you might call the fundamental theorems of welfare economics that the conditions for Pareto-optimality would be exactly realized by competitive arbitrage. Before Bergson, Lerner–Hicks–Hotelling–Kaldor–Scitovsky insufficiently understood that the full set of Pareto-optimality conditions constituted an incomplete set of conditions for ethical maximization. You must ask the right questions and make the right distinctions. All of my teachers believed there was something to Adam Smith's invisible hand—that each person pursuing their self interest would, by some miraculous action of the invisible hand, be led to contrive in some vague sense the best interest of all. However, none of them could explain properly what the truth and falsity was in that position. I would say that if I had been a bright student in 1894 and read Pareto's Italian journal article, I would have understood what I now understand to be the germ of truth in the invisible hand argument. All it refers to is the avoidance of deadweight loss. Here is where my association with Abram Bergson becomes relevant.

KS: How did you come to know Bergson to begin with, and how did you collaborate with him in developing the "new" welfare economics and the concept of social welfare functions?

PS: Bergson was my contemporary in the Harvard Graduate School. He was two years ahead of me. We were both puzzled by Pareto's writings. Bergson would read to me a passage from Pareto and ask: "What do you think is being said there?" What really puzzled us was that he seems to use a singular form for what is generally an infinitely broad class. Indeed, there isn't a Pareto-optimal point; there is a whole continuum of uncountable infinity of Pareto-optimal points which is what makes it a necessary condition and not a complete sufficient condition.

I was not an independent co-author of Bergson's 1938 paper published under his birth name, Abram Burk, which caused some confusion in the literature. I was a helpful midwife in helping to pull the baby out. I felt once the baby was pulled out, I had reached perfect clarification of the so-called new welfare economics.

KS: Who, in your opinion, were the most instrumental scholars in the evolution of the "new" welfare economics?

PS: The process of publishing the "new" welfare economics was not a well-organized, logical, and systematic thing at all. The names of the people who, at the minimum, would be involved include the following: Abba Lerner who, I think, is most important, John Hicks, Nicholas Kaldor, Tibor Scitovsky, Harold Hotelling, Ragnar Frisch, … Lerner never claimed that he was discovering a new principle, but Kaldor, Hicks and others did. We should expand our list by counting in Ian Little. There was also a pupil of Hicks at Manchester, Alexander Henderson, who perceived the following question: "Suppose that there are three necessary conditions for Pareto-optimality. Is it true that satisfying two out of these three conditions and not satisfying the third is always better than satisfying one of them and not satisfying the other two?" Now, if you count three apples, they are greater than two apples, and two apples are greater than one apple. It is also true that, in some sense, all of the three necessary conditions being satisfied is better than only two necessary conditions being satisfied. Yet it is not true in general that the more necessary conditions you satisfy, the better you always are.

KS: That is one of your concluding observations in Chapter 8 of the Foundations.

PS: Could have been, and Ian Little had that also.

KS: To identify the conditions for Pareto-optimality is one thing, and to go beyond Pareto-optimality by introducing the possibility of hypothetical compensation payments between gainers and losers, thereby expanding the reach of the Pareto principle to the situations involving interpersonal conflicts, is a different matter altogether. On reflection, what is your current verdict on the "new" welfare economics of the compensationist school?

PS: I think on the whole the "new" welfare economics of Kaldor, Hicks, Lerner and Scitovsky was overrated. In the first place, you know already you can find it in John Stuart Mill who discusses something like free trade. He in effect says that free trade may help some people, and hurt some other people, but the gainers would be able to compensate the losers. Thus, the "new" welfare economics of the compensationist school is not really that new. In the second place, there is a great ambiguity as to whether the fact that gainers would be capable of compensating the losers, yet do not actually pay compensations, has any significance.

I will give you an example. In 1959, my late wife and I made a trip to Japan at the invitation of the Japan Economic Newspaper (Nihon Keizai Shimbun). It was a wonderful trip—unbelievable. The head of the newspaper, Mr. Jiro Enjoji, took three weeks out of his busy life to travel all over Japan with us. Shigeto Tsuru and his wife were also with us. Shigeto was the tandem translator of my lectures which I gave in Tokyo, Nagoya, Osaka and Fukuoka.

KS: Shigeto Tsuru is your old friend from your Harvard days.

PS: That is right. During the war, Shigeto was evacuated to Japan. When he had to dispose of his books, I was the lucky recipient of his copy of the 1932 edition of Pigou's Economics of Welfare, which I read carefully.

At the time of our travel, Carl Christ, who was a Visiting Professor at the University of Tokyo for a year, told me that he was shocked by the rent controls in Japan. His advice was that they should be abolished. People said: "Well, yes, but it is not appropriate. There are a lot of poor people that will be very much hurt. A lot of old people will be very much hurt too. "Christ said:" No problem. Just compensate them." Now, there was no chance in the world that any Japanese Diet, or any post-MacArthur occupation, would have the ability to compensate, or would have the effective political desire ever to do it. So, you could not pin people down as if there was something important that could be done. But nobody took seriously what could be done. Lerner always taught us about ideal lump-sum taxes. However, there were very grave game-theoretical difficulties with lump-sum taxes, because the reason you ought to give people a lump-sum transfer is that they are poor, but as soon as the poor realize you are giving it to them because they are poor, they incur a blunting in their desire to work. This is a moral hazard problem. If, on the other hand, the potential compensation of the losers by the gainers remains a purely theoretical possibility, those who suffered losses remain unsalvaged. Thus, to say that lump-sum taxes could in principle solve the problem is not to say that they would actually solve it.

KS: Before turning to the core concept of the Bergson–Samuelson social welfare function and the "new" welfare economics based on it, I would like to ask you to clarify one specific point on the concept and nomenclature of Pareto-optimality. From what you have described so far, I understand that Bergson and yourself had a crystal-clear idea about what came to be known as Pareto-optimality. However, neither the 1938 Bergson article, nor the 1947 Foundations of Economic Analysis, made any explicit mention of the name of Pareto-optimality. As a matter of fact, in your 1981 Bergson Festschrift article, you have written that "the necessary condition(s) for an optimum, that such a universal improvement not be possible, Ian Little came in 1950 to call 'Pareto-optimality,' a felicitous and useful coinage." May I take it that the concept of Pareto-optimality was clearly grasped by Abram Bergson and Paul Samuelson, and maybe more vaguely by Abba Lerner and John Hicks, but the nomenclature of Pareto-optimality was first introduced by Ian Little?

PS: I'd guess that the person who put the word in print is indeed Ian Little. Somebody told me that he made a study and could not find the word, Pareto-optimality, in the literature. I was very surprised, because from the beginning that is the way Bergson and myself talked about it.

2.4 On the Concept of the Bergson-Samuelson Social Welfare Function

KS: Let us proceed to the crucial concept of the Bergson–Samuelson social welfare function. It is presumably to go beyond Pareto-optimality and spell out the exact necessary and sufficient conditions for ethical optimality that Bergson and Samuelson introduced the extraneous ethical norm in the form of a social welfare function. Could you please clarify the motivation behind the introduction of a social welfare function? Would you also explain how this concept was conceived in the first place?

PS: You cannot obtain an ethical result without already putting an ethical premise in the proposition from outside. This is already what 1951 Arrow would call an "imposition." Bergson laid out how the different forms of ethical premises could be implemented through the concept of social welfare functions, and how these different norms could reflect themselves in the results you would obtain. Of course, you could immediately understand how Pareto-optimality would fit into the Individualistic Bergson Social Welfare Functions because, if you took the necessary conditions that would survive no matter how you changed the interpersonal weightings, what you would have left would be nothing other than the necessary conditions for Pareto-optimality, which by themselves do fall short of achieving any ethical maximum. That is still true only under certain circumstances; you need to rule out altruism and envy or sadism or masochism. Bergson's Individualistic Social Welfare Function, by definition, must have the mathematical property of "weak separability." Without this, there may indeed exist no meaningful Pareto-optimal conditions. Let me give some partial examples. Start with Crusoe and Friday. Let neither have transitive preferences that satisfy any integrability conditions. Then never can you assert that letting them trade freely and spontaneously will "end both of them better off." "Better off-ness" is undefined and undefinable even for a one-individual universe! Add the further complication that corn and cloth both can change over from being "good goods" to being "bad goods." Then most of Lerner-like production-efficiency conditions cease to be capable of meaningful applications. All the 1954 alleged Pareto-optimal conditions that I derived for the "Wicksell–Lindahl public good problem" evaporate into thin air. Buddha or Saint Francis or Aristotle or Bergson can still impose on every state of the world an ethical transitive ordering. But of course theirs could be four contradictory ethical norming.

I recall that, at the NYU Sidney Hook conference on Philosophy and Economics, Kenneth Arrow startled the philosophers present (and me, too) when he declared something like: "Surely when all the individuals agree that situation A is better than situation B, any admissible ethical system must not second guess their desires." I don't recall Bergson as ever going to that extreme, even though to make sense of well known Pareto-optimality conditions he did include in his admissible Social Welfare Functions the weakly separable species in which those conditions did make sense. But never did he make the following common error: If situation α is Pareto-optimal and β is not, then always society should prefer α to β. And when asked to also contemplate situation γ which like α is Pareto-optimal, never did he pronounce on how one could deduce which of α and γ was the better ethically.

KS: In the provocative 1976 article devoted to the Paretian heritage, Chipman (1976, pp. 66–67 and pp. 109–110) claimed that Vilfredo Pareto had already "essentially developed the concept of a social welfare function" prior to its inception by Bergson and Samuelson. What do you think of this claim?

PS: I think that Chipman attributes the concept of a social welfare function to the 1913 article of Pareto. I also seem to remember that Kenneth Arrow may have had a similar viewpoint.

KS: Yes, Arrow is in fact of that belief, which I had an opportunity to confirm.

PS: I don't want to be definite in my reaction to that query. However, I should say that as a person with great but guarded admiration for Pareto, I think he was often, at least momentarily, confused, and he was simultaneously at different levels of his stages of thinking. You must remember that Pareto never had any students really. He lectured to lawyers. He had disciples, but he didn't have the advantage of people like us today, where you try out your ideas on 20 different equals. He had no equals; that made him uneven and a little eccentric. But, just like Joseph Schumpeter, Pareto professed great self-confidence, sureness, and disconcernment toward everybody else's ideas. Chipman argues that when Pareto introduced the word, ophelimity, he did it partly to get rid of various hedonistic and other connotations. But, Chipman believes when all is said and done, he did have a notion of preferred cardinal utility and believed that everybody had that. Well, if that is so, it is a kind of confusion, because he gives no rational grounds for preferring one cardinal-numbering over another. Pareto's discussion on complementarity was uneven. Mathematicians must be very exact, but late in the day he was using the sign of the cross-derivatives of cardinal utility. You know, the moment you transform cardinal utility, you can change the sign of the cross-derivatives.

Let me connect this up with the social welfare function. I had to read Pareto in the Italian original, and my command of Italian was very poor. Nevertheless, I had a feeling when I read the 1913 article—I say this with diffidence—that he may momentarily have had the notion of an imposed-from-outside social welfare function which itself would not be different from Bergson's one. I don't think that subtracts anything from Bergson's originality. But I thought I detected in it also a positivistic real political function of certain elites in any society. Each one of these elites has different power, like the powers of father and mother, oldest son, younger sons in a family. If you try to get a demand function for the family, you must combine these different influences. Generally speaking, when you do that, you don't get an integrable function. To me, that was what Pareto was talking about in the 1913 article.

The same puzzle comes about in my 1956 Quarterly Journal of Economics article on social indifference curves. A key concept in this article is that of the "just" society. It is the society in which, somehow in the background, lump-sum payments have been made so as to keep maximizing a collective (Bergson) social welfare function that is not too distinct from weak separability. Of course, it is just a thought experiment. It would be extremely hard in any experimental situation to get information and to do it. When Gary Becker tried to write on the economics of families, he kind of took over that notion. He somehow thought that there really exists conceptually such an archetypical family of social indifference curves. I think it is extremely unrealistic; I am not sure that Pareto, who by 1913 was deeply in sociology, would have agreed with Becker. He regarded sociology as everything that included more than economics, including very contradictory items and with emphasis on irrationalities and non-integrable preferences.

Thus, it could be that I could see places in which Pareto had a concept very much like the Bergson social welfare function. But I think there are other logically distinguishable notions in his discussion. The problems have been made more complicated by the fact that Pareto liked to use little deltas and equate them, which I never liked. You can't be sure what Pareto meant by his infinitesimals. I don't believe that he was above reproach with respect to confusing and even being himself confused as if he knew what he was saying.

That is all I can say on the problem which you posed.[3]

2.5 On the Concept of the Arrow Social Welfare Function

KS: Soon after the publication of Kenneth Arrow's Social Choice and Individual Values, Ian Little, James Buchanan, and Abram Bergson, respectively, published a harsh conceptual and substantive criticism against Arrow's use of the concept of a social welfare function and his general impossibility theorem. To the best of my knowledge, your own published criticism on Arrow's social welfare function and general impossibility theorem appeared in the 1967 article entitled "Arrow's Mathematical Politics." In this article, you exported Arrow from economics to politics with a remark that Arrow's general impossibility theorem is a seminal contribution to the infant discipline of mathematical politics, but it has nothing to do with welfare economics. Would you please recollect what was your first response to Arrow's social welfare function and his general impossibility theorem?

PS: From the beginning, I thought it unfortunate that Arrow used the terminology of welfare economics when he was in fact making a path-breaking contribution to the emerging discipline of mathematical politics. I read your interview with him with great interest. I am a great admirer of Kenneth Arrow. I consider him as one of the greatest economists of our time. I think that one of the biggest mistakes that Stockholm ever made was to give him a half of the Nobel Prize. There were two mistakes at the same time. They gave Hicks only a half of the prize and they should have given him a full prize. Maybe they should have given Arrow two prizes, one for his contributions to social choice theory, and another for his work in probability and information, which is quite different.

It is interesting to read Arrow's recollection of how he went about the problem of social choice, which agrees a little with my impressions, my imperfect memories. In the summer of 1948, Olaf Helmer, a logician at the Rand Corporation, was trying to develop game theory as a tool for the analysis of international relations and military conflict. He told Arrow that he was troubled by the foundations of economists' application of game theory. When applied to international relations, the players were

---

[3]In an early response to Chipman (1976); Samuelson (1977b, p. 177) made an almost sarcastic remark on Chipman's assertion to the following effect: "This, I believe, involves an act of sympathetic charity since Pareto's many writings are often obscure on what we now call Pareto-optimality, and since expressions such as $\theta_1(\delta U^1) + \theta_2(\delta U^2) + \cdots$ are sometimes used by Pareto as positivistic-politics constructs and sometimes as vague Lagrange multiplier expressions relevant to the first-order conditions for being on the ("Pareto-optimal" points of the) utility-possibility frontier." Subsequently, Bergson (1983, p. 44) basically concurred with this verdict when he concluded that "it still seems difficult to quarrel with Samuelson's ⋯ assessment of Chipman's perception."

countries, not individuals. In what sense, Helmer asked, could collectivities be said to have utility functions? Arrow immediately replied that this question had been answered by Bergson's notion of social welfare functions and he tried to give Helmer a brief exposition. It resulted in his discovery of the general impossibility theorem. Now, I think he went into mud, looking for a small pearl, and came out with a big diamond. It was a very important finding in political science as it showed that the failure of specific voting functions is not due to any lack of cleverness, but is a reflection of general impossibility. However, it had nothing to do with ethics and welfare economics. Arrow's use of "social welfare function" for his "voting function" was unfortunate. Arrow wanted to "impose" nothing, which in my book removed him already from the issue of ethics.

KS: Were you in general agreement with Ian Little and Abram Bergson in their criticisms?

PS: The moment Arrow's book came out, and maybe the moment his article came out in the Journal of Political Economy earlier, independently and at least in three different minds—Ian Little's mind, my mind, and Abram Bergson's mind—, there came a realization that Arrow was not talking about the same thing.

KS: By the "same thing" you mean the historic economists' social welfare function ….

PS: That's right. Arrow's general impossibility theorem does not disprove the existence of the Bergsonian social welfare function, neither does it disprove the existence of the Benthamite hedonistic function. As I said, I am a great admirer of Kenneth Arrow, and there are only two things I have ever disagreed with in his writings. One, not very important, difference is that, in axiomatizing the von Neumann–Savage utility system for gambling, he believes that you ought to make utility bounded. This is to avoid the St. Petersburg paradox. I beg to differ, because I think that the St. Petersburg paradox is only a classroom paradox. It is a purely contrived infinity. I don't think I have ever succeeded in pursuading Arrow on this. Another, this time important, difference is his usage of the concept of a social welfare function. When he brought out his new edition, he must have known the objection of Bergson; he must have known the objection of Little; and I think he certainly knew of the objection by me. As far as I know, however, he just paid no attention to them. I have never heard of Arrow saying that it was a linguistically unfortunate abuse of those three words—the same three words. I think he was sort of reaffirming his right to have done it.[4]

[4]For the sake of setting the record straight, two lengthy remarks on the literature may be in order at this juncture.

In the first place, it seems fair to cite two of Arrow's actual writings on the concept of a social welfare function. On the one hand, in "Notes on the Theory of Social Choice, 1963," which Arrow appended to the second edition of Social Choice and Individual Values, he wrote as follows: "It would perhaps have been better for me to use a different term from 'social welfare function' for the process of determining a social ordering or choice function from individual orderings, although the difference between Bergson's definition and my own was pretty carefully spelled out…. I will therefore now use the term 'constitution,' as suggested by Kemp and Asimakoplos. The difference, however, is largely terminological; to have a social welfare function in Bergson's sense, there must be a constitution [Arrow (1963, pp. 104–105)]." On the other hand, in his contribution to the book

KS: What did you think about Buchanan's criticism of Arrow to the effect that the Arrovian social welfare function, or constitution, which hinges squarely on the concept of collective rationality, is nothing other than a category mistake?

PS: Would you remind me of Buchanan's criticism of Arrow? If you spell it out simply, I will generate a reaction to it.

KS: Let me try. Arrow's general impossibility theorem depends on the assumption of collective rationality to the effect that the social choice is made in accordance with the maximization of an underlying social preference ordering, which is constructed on the basis of the profile of individual preference orderings through some process or rule, within the given social opportunity set. In his 1954 Journal of Political Economy article, James Buchanan criticized Arrow for his use of the assumption of collective rationality in the above sense by asserting that it was an illegitimate transplantation of a property of individuals only: "The mere introduction of the idea of social rationality suggests the fundamental philosophical issues involved. Rationality or irrationality as an attribute of the social group implies the imputation to the group of an organic existence apart from that of its individual components. …

---

edited in honor of Samuelson, Paul Samuelson and Modern Economic Theory, Arrow referred to a passage from Samuelson's 1981 Bergson Festschrift article, "Bergsonian Welfare Economics," and firmly asserted as follows: "If there are 'rumors that Kenneth Arrow's Impossibility Theorem rendered Bergson's "social welfare function" somehow non-existent or self-contradictory,' they are indeed 'quite confused' (Arrow 1983, p. 21)." To substantiate this statement, Arrow observed that the Pareto quasi-ordering corresponding to each and every profile of individual preference orderings can be extended into a complete ordering by virtue of Szpilrajn's classical extension theorem. Thus, it seems fair to say that the conceptual difference and interrelationship between the Bergson social welfare function and the Arrow social welfare function are by now well recognized by Arrow and whole profession. It may also be asserted that a wide recognition exists by now that Arrow's general impossibility theorem does not disprove the existence of the Bergson social welfare function; it is a theorem on the non-existence of the Arrow social welfare function, or constitution, and not on the non-existence of the Bergson social welfare function.

In the second place, although Bergson and Samuelson are in complete agreement on the conceptual distinction between the Bergson–Samuelson social welfare function and the Arrow social welfare function, as well as on the irrelevance of the Arrow impossibility theorem to welfare economics, there are at least two junctures where they seem to have chosen somewhat different directions. On the one hand, there is no room for compromise whatsoever in Samuelson's purge of the Arrow impossibility theorem from the territory of welfare economics. In contrast, Bergson seems to have taken a somewhat more flexible stance in this arena. It is true that Bergson (1954, p. 240) began his examination of "Arrow's Theorem in Relation to Welfare Economics" by declaring that "[i]n my opinion, Arrow's theorem is unrelated to welfare economics." However, he was careful enough to note that there is a conception of the concern of welfare economics which allows a different interpretation of the Arrow impossibility theorem: "According to this view, the problem is to counsel not citizens generally but public officials. Furthermore, the values to be taken as data are not those which might guide the official if he were a private citizen. The official is envisaged instead as more or less neutral ethically. His one aim in life is to implement the values of other citizens as given by some rule of collective decision-making. Arrow's theorem apparently contributes to this sort of welfare economics… (Bergson 1954, p. 242)." It is worthwhile to point out that Arrow (1963, p. 107) fully endorsed this view of welfare economics which Bergson aptly identified. On the other hand, Samuelson (1947, p. 221) admits no reason whatsoever to be concerned with the origin and/or nature of the values captured by the social welfare function: "Without inquiring into its origins, we take as a starting point for our discussion a function of all the economic magnitudes of a system which is supposed to characterize some ethical belief—that of a benevolent despot, or

We may adopt the philosophical bases of individualism in which the individual is the only entity possessing ends or values. In this case no question of social or collective rationality may be raised. A social value scale simply does not exist. Alternatively, we may adopt some variant of the organic philosophical assumption in which the collectivity is an independent entity possessing its own value ordering. It is legitimate to test the rationality or irrationality of this entity only against this value ordering."

PS: My own views about ethics are, generally speaking, against a narrow and special view. Hearing your summary of Buchanan's criticism, I don't at all agree with his position. It boils down to the claim that, if it is a social choice in an individualistic society that is being analyzed, then you should not be interested in any degree of rationality, consistency, or transitivity at the social level. This would be like an answer from fallacy. It seems to be a Humpty-Dumptyism. Humpty-Dumpty says: "If I say a thing twice, then it is true." I see no reason to think that there is any cogent force in Buchanan's argument. What he says boils down to the statement: I, Buchanan, have no interest in that." He gives no reason why other reasonable men should go along with him. I think those were blinders of his own creation. If readers recall why Harsanyi in 1955 converted me into accepting some role for strongly additive interpersonal BSWF's, then we'll recognize that I had respect for an Ethics Giver who wants to obey the Marschak–Savage Independence Axiom of Laplacian rationality. Buchanan is interested only in living human beings—sober or drunk, young or old, …. Dogs or chimps or Alzheimer sufferers need not apply.

KS: I understand that you firmly retain your previous verdict that Arrow's contribution to social choice theory is not relevant to ethics and welfare economics. What, then, is your current opinion on the scientific status of social choice theory in general, and Arrow's general impossibility theorem in particular?

PS: I regard social choice theory in the narrow sense as orthogonal to welfare economics. It can be a part of positivistic study of voting systems. I like the title of your journal, Social Choice and Welfare, but by connecting social choice theory with welfare economics, Arrow seems to have created much of the unfortunate confusions. Indeed, social welfare can be completely congruent with the pre-Arrow literature on welfare economics and moral philosophy. Arrow wanted to find out how an individualistic Bergson–Samuelson social welfare function could be generated democratically. But I should register a difference in opinion here.

---

a complete egoist, or 'all men of good will,' a misanthrope, the state, race, or group mind, God, etc. Any possible opinion is admissible, including my own, although it is best in the first instance, in view of human frailty where one's own beliefs are involved, to omit the latter. We only require that the belief be such as to admit of an unequivocal answer as to whether one configuration of the economic system is 'better' or 'worse' than any other or 'indifferent,' and that these relationships are transitive.…" In contrast, Bergson (1976, p. 186) is ready to be concerned with the nature of the values to be captured by the social welfare function: "The practitioner of welfare economics is in principle free to take any values as a point of departure, but the resulting counsel as to economic policy is not apt to be too relevant unless the values in question are held by, or can plausibly be imputed to, one or more officials concerned with the policies in question. Should the practitioner for any reason disapprove of those values, he may, of course, refrain from offering the officials any counsel at all."

Arrow has said more than once that any theory of ethics boils down to how the individuals involved feel about ethics. I strongly disagree. I think every one of us as individuals knows that our orderings are imperfect. They are inconsistent; they are changeable; they come back. We go out at night and we leave our wallet at home, because we don't trust ourselves, and we are right not to trust ourselves. We do things and say, "I am going to hate myself in the morning" and, in the morning, we do hate ourselves. There are no ideal individuals who, as adults, suddenly become these perfect individuals. People talk about paternalism as if we were bowing down to a dictator, but it is wrong in ethics to rule out imposition, and even dictatorship, because that is the essence of ethics. Take, for example, the simple axiom of unanimity and suppose that people are self destructive ethically. The notion that every ethical system will have to recognize a unanimous agreement by people is like encouraging bad children to be bad children. I am serious in my belief that difference between a child and an adult is only a difference of degree. In the old paintings, the children are little adults; in modern paintings, if you did them right, the adults are only badder or older children. We are all imperfect. This is not a doctrine of the original sin; it is a doctrine of the imperfectability of mankind. It is too presumptuous to suppose that individuals are consistent, transitive and meaningfully unchangeable in their views. By the way, Piero Sraffa never believed in modern demand theory at all and tried to do everything with cost and technology alone, because he believed people are changing all the time. In this he does not earn my blessing. Piero, like Margaret Fuller, "accept the universe!"

I would say that the ruling theme among economists since 1750 goes something like this. There is a vague notion, which could not be written up for a classroom examination, that there is something optimal about laissez-faire pricing. Among the most sophisticated lay people, it is realized that laissez-faire pricing systematically makes some people better off and some other people worse off, and this pattern quickly changes. There is a chivalrous rule of thumb: "Don't interfere with it." In the first place, if you do interfere with it, you probably do as much ethical harm as good because of imperfect government. But, more than that, there is the law of large numbers operating. One invention helps A, another invention helps B; by James Bernoulli's theorem of large numbers, it evens out. Perhaps. The trickle down theory from inequality is bred by the Schumpeterian dynamic process of innovation. The total pie is improved; on the whole and over time, it evenly lifts up everybody. The same tide raises all ships. That is dogmatic faith, but I think it is in the background of intelligent conservatives. John Hicks certainly. His implicit faith is that it will even out upward. In terms of economic history, there is a lot of truth in that faith. This is a kind of common sense ethics, and most people don't want to go into the complicated questions, I think. I don't know whether most people should.

KS: You have been consistently asserting that the informational basis of the Bergson–Samuelson social welfare function of the individualistic type is the profile of individual preference orderings which are ordinal and interpersonally non-comparable. However, if we require that the social welfare judgments are complete and quasi-transitive with unrestricted domain, and the Pareto principle and the anonymity principle should be respected together with the Arrovian axiom of independence of irrelevant alternatives, then the social welfare judgments should be such that all the Pareto non-comparable social alternatives are judged to be socially indifferent. This simple theorem is due to Amartya Sen, and it tells us that the exclusive reliance on the ordinal and interpersonally non-comparable preference information may be inappropriate, as it excludes distributional equity judgments in the situation of interpersonal conflict altogether. Could you please comment on this concern and clarify your stance once again?

PS: Here is a (singular) quintessential Bergson Individualistic SWF. Jane lacks altruism or envy. So does Tom. Each consumes his apples and his oranges; or her apples and her oranges. Present the Ethicist with a total of 100 apples and 100 oranges, which can be allocated 50–50, 10–90, 100–0, ... between them. Suppose Tom's and Jane's choices could be described by any one of the infinity of following cardinal utility functions:

$$\text{Tom} : \Phi = \log\left(\text{apples}^{\text{T}}\right) + \log\left(\text{oranges}^{\text{T}}\right) \text{ or } F\{\Phi\} \text{ with } F' > 0 \geq F''$$
$$\text{Jane} : \phi = \left(\text{apples}^{\text{J}}\right)^2\left(\text{oranges}^{\text{J}}\right) \text{ or } f\{\phi\} \text{ with } f' > 0 \geq f''.$$

Then Pareto-optimality conditions would have the same content whatever was the IBSWF of the form

$$\Theta(\Phi, \phi),\ \partial\Theta/\partial\Phi > 0,\ \partial\Theta/\partial\phi > 0.$$

But let Jane and Tom each have algebraic sympathy. And perhaps introduce a public good that both consume at once. Then the general BSWF might be of the form

$$\psi(\text{Oranges}^{\text{T}},\ \text{Oranges}^{\text{J}},\ \text{apples}^{\text{T}},\ \text{apples}^{\text{J}},\ \text{public good})$$

or

$$G\{\psi()\},\ G' > 0 \lesseqqgtr G''.$$

Given $\sum$oranges $= 100 = \sum$apples and public good $= 1$, what Pareto efficiency condition(s) could you ever deduce? Often none.

As bad is when 100 chocolates are to be divided between atom Tom and atom Jane. Every allocation is (emptily?) Pareto-optimal. Room is left for any imposed ethics. I hope a Sen would not say that no non-trivial BSWF's exist. Room is left for indefinitely many.

The most general B-S SWF can make judgments like: "Five biscuits to Tom, other things equal, is better than four biscuits to Jane" without having to mean that some utility of Tom is being compared to some utility of Jane. Maybe Tom has no transitive ordinal ordering that the ethicist must "respect." And what Tom chooses could be deemed to be ethically wrong and ignorable.

2.6 On the Single-Profile Impossibility Theorems

KS: Ian Little and yourself emphasized that the Bergson–Samuelson social welfare function is defined for any fixed profile of individual preference orderings characterizing the given society. In contrast, Arrow's social welfare function, or constitution, is a "process or rule" assigning a social welfare ordering to each and every logically possible profile of individual preference orderings. Your 1967 article on "Arrow's Mathematical Politics" has identified this sharp contrast between the Bergson–Samuelson single-profile framework and the Arrow multiple-profile framework to be the primary logical culprit for the Arrow impossibility theorem. As a response to your charge against the multiple-profile framework of the Arrovian social choice theory, many single-profile counterparts of the Arrow impossibility theorem have been presented by Kemp and Ng (1976), Parks (1976), Pollak (1979), Roberts (1980), Sen (1993), and many others. Would you please recapitulate your verdicts on the status of Arrovian impossibility theorems in view of these single-profile general impossibility theorems?

PS: There are 999 such single-profile impossibility theorems but none that I know of cogently exclude interesting possible single profiles. I don't know of any important single-profile impossibility theorem. I can generate lots of Kemp–Ng theorems, none of which are cogently relevant. Ian Little has recently published his collected papers, Collections and Recollections, in which he commented on his famous criticism of Arrow published in the 1952 Journal of Political Economy as follows: "One of the main points made was that Arrow's famous book… had no bearing on traditional welfare economics. One of the reasons given was that the conditions required of a satisfactory 'social welfare function' (SWF) were stated in terms of changes in individual ordering, whereas the locus classicus of a SWF (viz. Bergson 1938) stated it for a given set of orderings. It seems that too much was made of this, in that it has been subsequently shown [in Kemp and Ng (1976)] that a very similar impossibility theorem can be proved for a given set of orderings [Little (1999, pp. 17–18)]." I wrote to Little and said: "You weren't wrong earlier, but maybe just confused."[5]

---

[5]When I had corresponded with Professor Ian Little about this interview, he kindly made the relevant passages of Professor Samuelson's letter available to me. Since it is of some interest, I am hereby citing it after receiving permission to do so from Professors Samuelson and Little: Little to Suzumura: 29 March 2005

Dear Suzumura,

The relevant part of Paul Samuelson's letter of 3 November 1999 is as follows:

'Belatedly I have learned about the existence of your *Collection and Recollections*. Now that I have got as far as page 18, I wish to present you with a gift. On page 18, in your first complete sentence you seem to be lowering your flag— which is also my flag. This is because of the Kemp and Ng (1976) *Economica* paper.

I suggest you rewrite that sentence in all the subsequent editions of your Memoirs to read as follows: "I was quite right in my original position, even though Kemp and Ng in a 1976 *Economica* article purposed to prove the opposite. Professor Samuelson in a pre-humous letter has supplied me with a reprint of his cogent 1977 *Economica* refutation of the Kemp-Ng contention, which serves as a confirmation of my critique of Arrow."

Take care of yourself. They are not making many of our kind any more.'

Take, for example, a single-profile impossibility theorem à la Kemp and Ng (1976), which hinges squarely on their Axiom 3. I wrote in my published response to them: "I must regard Axiom 3 of Kemp–Ng as anything but 'reasonable' to impose on a Bergson–Samuelson Individualistic Social Welfare Function…. As Oscar Wilde might put it, "For any ethical observer to understand Axiom 3 is to reject it (Samuelson 1977a, p. 81)."[6]

KS: What about Amartya Sen's Impossibility of a Paretian Liberal, which is another example of a single-profile impossibility theorem without taking any recourse to the Kemp–Ng Axiom 3, or anything like that?

PS: Earlier on, I discussed when Pareto-optimality conditions might evaporate away. If Sen agrees with that, well and good. You might want to clarify to me how interesting and important is Sen's case against a Paretian liberal.

KS: Let me try. One of the axioms in Arrovian social choice theory, which has been left almost unchallenged in the literature, is the Pareto principle to the effect that unanimous preferences among individuals for a social state x against another social state y is to be faithfully embodied in the social preference for x against y. Sen posed a serious criticism against this ubiquitous acceptance of the Pareto principle. He did this in terms of an intuitive example involving an individual's libertarian right to read a book in his/her private room or not without outside interference, which Sen elaborated into a simple yet powerful impossibility theorem on the Paretian

---

I had regrettably not read Samuelson's 1977 *Economica* article. If I had, I would not have "lowered our flag." I promised him I would include his amendment in any future edition of *Collection and Recollections*. However, I fear that the probability of there being another edition is extremely close to zero. Do with this what you like. I am very happy to know that Paul is still pre-humous. With best wishes.

Yours sincerely,

Ian Little

[6]Let us recapitulate Samuelson's criticism on Axiom 3 of Kemp and Ng more in detail.

Suppose society has a fixed total number of chocolates that could be partitioned between two specified selfish hedonists: say, 80 and 20, 50 and 50, 20 and 80, or more generally as any of two non-negative real variables (X chocolates to Person 1 or x chocolates to Person 2), where $X + x = 100$ and neither is negative…. What is the meaning of the new Axiom 3 in this context? It says, "If it is ethically better to take something (say one chocolate or, alternatively, say 50 chocolates) from Person 1 who had all the chocolates in order to give to Person 2 who had none, then it must be ethically preferable to give all the chocolates to Person 2." One need not be a doctrinaire egalitarian to be speechless at this requirement. Is it "reasonable" to put on an ethical system such a straightjacket? Few will agree that it is [Samuelson (1977a, p. 83)].

It seems to me that the forcefulness of this criticism originates in the fact that we are informed of the material background of the following preference orderings of Person 1 and Person 2:

Person 1: $(100, 0), (100 - \varepsilon, \varepsilon), (0, 100)$

Person 2: $(0, 100), (100 - \varepsilon, \varepsilon), (100, 0)$

where ε is a small positive number. If the informational basis of social welfare judgments is limited only to the profile of (ordinal) individual utilities and we are deprived of whatever non-welfare information about the social alternatives, the Kemp–Ng Axiom 3 may not be that easy to shoot down. It is in this sense that the Kemp–Ng Axiom 3 is said to be a counterpart in their single-profile framework of Arrow's Axiom of "Independence of Irrelevant Alternatives" in his multiple-profile framework. Therefore, what is to blame may not be the Kemp–Ng Axiom 3 per se, but the narrow informational basis of ordinal welfarism.

liberal.[7] In conspicuous contrast with Arrovian impossibility theorems, which hinge squarely on the multiple-profile framework a la Arrow, Sen's impossibility theorem invoked only a single profile of individual preference orderings. Another contrast to be noticed is that Sen's impossibility theorem depend neither on the axiom of collective rationality, nor on the axiom of independence of irrelevant alternatives, which is another important constituent of Arrow's impossibility theorem.

PS: Protestant ministers objected around 1600 to bear baiting, not because bears should not be made to suffer but rather because spectators should not be allowed this obscene pleasure. If that's the kind of thing Sen has in mind, I or Bergson might say: "Who are we to tell those ethical prescribers that they are being silly or acting inadmissively?"

2.7 On Consequentialism and Welfarism

KS: Our discussion on Sen's impossibility theorem is a convenient step toward further examination of the informational basis of social choice theory and welfare economics. As Arrow (1987, p. 124) has aptly observed, "it has been taken for granted in virtually all economic policy discussions since the time of Adam Smith, if not before, that alternative policies should be judged on the basis of their consequences for individuals." As a matter of fact, most of the contemporary welfare economics is based not just on consequentialism in this sense; it is based on welfarist-consequentialism, or welfarism for short, in the sense that consequences are evaluated solely on the basis of utilities entertained by individuals from these consequences.[8] To the best of my knowledge, it was Hicks (1959) who first declared in "Preface—and a Manifest" in his Essays in World Economics that welfarism is too narrow as the informational basis of welfare economics for it to serve the enhancement of human well-being. It was in his farewell to the traditional informational basis of welfare economics that he coined the term, economic welfarism:

> The view which, now, I do not hold I propose (with every apology) to call 'Economic Welfarism': for it is one of the tendencies which has taken its origin from that great and immensely influential work, the Economics of Welfare of Pigou .... The line between Economic Welfarism and its opposite is not concerned with what economists call utilities; it is concerned with the transition from Utility to the more general good, Welfare (if we like) itself [Hicks (1959, pp. viii–ix)].

[7]There is a debate in the literature concerning the legitimate articulation of individual rights in the conceptual framework of social choice theory. As this debate has very little to do with the present issue of the ubiquitous applicability of the Pareto principle, we have only to refer those who are interested in this debate to Gaertner et al. (1992), Gärdenfors (1981), Sen (1992), Sugden (1985) and Suzumura (1996).

[8]According to Sen (1979, p. 538), welfarism just represents an informational constraint to the following effect: "Social welfare is a function of personal utility levels, so that any two social states must be ranked entirely on the basis of personal utilities in the respective states (irrespective of the non-utility features of the states)."

Hicks was led to dissociate himself from Economic Welfarism, because he came to believe that "it is impossible to make 'economic' proposals that do not have 'non-economic aspects,' as the Welfarist would call them; when the economist makes a recommendation, he is responsible for it in the round; all aspects of that recommendation, whether he chooses to label them economic or not, are his concern [Hicks (1959, pp. x–xi)]." However, Hicks was surely not ready to jump to the other polar extreme:

> I have … no intention, in abandoning Economic Welfarism, of falling into the 'fiat libertas, ruat caelum' which some later-day liberals seem to see as the only alternative. What I do maintain is that the liberal goods are goods; that they are values which, however, must be weighed up against other values. The freedom and the justice that are possible of attainment are not the same in all societies, at all times, and in all places; they are themselves conditioned by external environment, and (in the short period at least) by what has occurred in the past. Yet we can recognize these limitations, and still feel that these ends are worthier ends than those which are represented in a production index. It is better to think of economic activity as means to these ends, than as means to different ends, which are entirely its own [Hicks (1959, p. xiv)].

What is your current response to Hicks's manifest against economic welfarism? Do you feel sympathetic to his conversion?

PS: As a reporter on the philosophy of ethics, how would I want to react to J. S. Mill's disagreement with Bentham's dictim: The pleasure of the game push-pin is as important as Shakespeare's poetry? Understanding Mill's reaction I would still have to say: "Each has a right to his opinion. After all it is his (Bentham's) opinion." If Hicks is newly converted to being able to admit judgments like this I see nothing revolutionary in that. Why is it a rejection of something called "welfarism?" I call it a welfarism that differs from regarding each individual as an atom who values algebraically only his vector of his goods and who is put in a strongly separable normative function that insists on the equality of

$$(\partial \text{ Tom's apples}/\partial \text{ Tom's oranges})_{\text{Tom}}$$

to

$$(\partial \text{ Tom's apples}/\partial \text{ Tom's oranges})_{\text{Bergson ethicist}}.$$

Maybe, like a character in a Moliere play, J. R. H. was becoming in 1959 more Bergsonianly eclectic without realizing it.

KS: Both Abram Bergson and yourself were careful enough to avoid premature commitment to welfarism in your initial exposition of the concept of a social welfare function. However, your famous Chap. 8 in the Foundations on welfare economics has a passage where an explicitly welfaristic formulation of social welfare function is presented. To be more specific, in p. 228 of the Foundations, we encounter the expression for social welfare W as a function of the profile of individual utilities: $W = F(U^1, U^2, \ldots)$. It is this formulation which is often cited, e.g., by Sen (1979), as a sure-fire proof that a social welfare function à la Bergson and Samuelson is unambiguously welfaristic in nature.[9] Would you please tell me whether you regard yourself as a welfarist in your own social welfare function?

PS: I named as an extreme atomistic type the case where each person cares only for his goods and bads and where the ethical prescriber gives some measurable weight to each of their own private rankings. To declare that elements of envy and sympathy and sadism and altruism bring us into or out of "welfarism" is mere prattle. My view would be as wide as possible. In the sense that Hicks and Sen used the term, I am not exclusively a welfarist; the expression cited by Sen is just an example of the possible class of social welfare functions, which happens to be welfaristic. Consistent with Hicks's manifest, my own social welfare function will have a large room to accommodate freedom, liberty and rights. My own ethical value must not dictate my analyses of ethics.

By the way, Milton Friedman is not a consequentialist, who neither wants there to be more bread, nor particularly cares whether that bread is equally distributed among people. He wants there to be liberty. He would be disappointed if, by non-liberty, a rational collective state could create prosperity. He does not believe it would be possible, but he would be disappointed if it should happen to be the case. I am not that extreme non-consequentialist; if people do want liberty, I would ask how much they are willing to pay in terms of sacrifice with bread. Milton Friedman thinks that liberty is something that belongs to him, and somebody else is taking it away from him. He thinks that liberty is something that can be treated algebraically, and scaled to get more quantities. You can get more of that good stuff. I am a more cautious libertarian. There is an old saying. A man is walking down St. James Street in London, swinging a cane in a wide curve. An old passerby speaks to him: "Hey, white bear, you are swinging your cane." The guy replies: "It's a free country, isn't it?" The old passerby retorts: "Your freedom ends where my nose begins." But the old man is wrong; the white bear's freedom ends long before his neighbor's nose begins. One man's right to privacy is another man's condemnation to loneliness. I don't say that in order to make ethics of liberty simple; I say that in order to make it realistic, because it is not simple.

Let me also tell you my personal experiment at the University of Chicago as a little hobby. I was curious. Are economic libertarians, who are against exchange controls, against price controls, and against rationing, also zealous Voltairean believers in freedom of opinion, free expression of opinion like John Stuart Mill's irreducible civil liberties? Therefore, I observed (covertly) the behavior of my friends who might be thought to be strong economic libertarians to see whether they were also strong political libertarians. Quite to the contrary. I asked Milton Friedman, in a quiet non-confrontational way so that he had not known I was studying his behavior. The

---

[9]Likewise, we find a passage in Bergson (1948, p. 418), which reads as follows:

If the decision is in favor of consumers' sovereignty, the welfare function may be expressed in the form, (1) $W = F(U^1, U^2, U^3, \ldots)$. Here U^1, U^2, U^3, etc., represent the utilities of the individual households as they see them and W, the welfare of the community, is understood to be an increasing function of these utilities. The welfare of the community, then, is constant, increases or decreases, according to whether the utilities of the individual households are constant, increase or decrease.

If U^1 has arguments about corn^2 in it, the weak separability is spurious (Added by P. Samuelson).

question was about Paul Sweezy, who was invited by a leftist philosophy teacher to the University of New Hampshire to talk to his class back in the Joe McCarthy witch-hunt days. He was subsequently brought up before the New Hampshire legislature to testify on what he had talked about. He refused to do both. I asked Milton Friedman: "Do you think he should have been required to do that?" Friedman replied: "Of course! Public money is running the University." I asked further: "You mean, it would be different if it was Dartmouth College, a private school?" He said: "Well, a wise and honest man should be willing to admit what he said." I said: "You don't understand. Everybody knows what he said. The meaning of this is not to learn new information. It is to bring out the despicable fact that he spoke, let us say, in favor of the Soviet Union." Milton Friedman had no sympathy for Paul Sweezy. The only exception I found in my wide sample was Fritz Machlup. I mentioned this to my late colleague, Evsey Domar, who was a colleague of Fritz Machlup at Johns Hopkins University. He said: "Oh, that is nothing. He is in love with professors." I said: "I don't care. I just want to get at the barebone fact by whatever reasons."

I think those who were the most derogatory in what they think of the narrow welfarism exaggerate what most people feel. If you want to find out who are the happiest people in the world, it is very difficult to do because of the way you ask the question, and the way your question is answered. I heard at a private dining club the philosopher, Sissela Bok, who is the daughter of Gunnar Myrdal but has a very different personality from her father. She was making a study of what people say about their own happiness. It turns out that it is people in Finland, Sweden and Holland, not people in Africa, not people in Indonesia or France or the USA who report most happiness. We used to think of those Northern countries as having a lot of suicide; they kept more honest records than Catholic countries. Why are they happy? They are happy because they have three good meals and a good medical care. They are falling behind us a little bit in the sweepstakes of growth from 1970 to the present time, but for a hundred years they evolved up from a very slow, very cold and unproductive society mostly through education. I think the lip services people give to the non-economic objectives turn out to mean very little to them when they cost a lot in economic terms. Along with liberty come the unintended consequences of liberty. Spain after Franco's day is a very nice free country, but I was told by an accompanying government official that "in Franco's day, we could take the subway to the office where we are going, but I really don't advise two middle class people in the middle of the day to use the Madrid subway." The Soviet Union freed from Stalin's tyranny has a lot of chaos including mafia chaos.

2.8 On the Resurgence of Consumers' Surplus

KS: In a famous section, "Why Consumer's Surplus is Superfluous," in the Foundations, you raised a famous and devastating criticism against the Marshallian concept of consumer's surplus, which started as follows (Samuelson 1947, p. 195): "Any judgment as to the usefulness or lack of usefulness of consumer's surplus has nothing to do with the problem of the admissibility of welfare economics as a significant part of economic theory since nobody has ever argued that the latter subject presupposes the validity of consumer's surplus. Can it then be said that consumer's surplus if not necessary, is nevertheless a useful construct?" Answering this question of your own strongly in the negative, you concluded as follows: "It is for these reasons that my ideal Principles would not include consumer's surplus in the chapter on welfare economics except possibly in a footnote, although in my perfect Primer the concept might have a limited place, provided its antidote and alternatives were included close at hand (Samuelson 1947, p. 195)." Not many economists were bold enough to challenge your sweeping and definitive criticism, yet we may find in Max Corden's Theory of Protection the following passage:

> The reader might recall the story of consumers' surplus. Here was a simple intuitively appealing idea, discovered by Dupuit, rediscovered and developed by Marshall, revived by Hicks, and obviously useful. Upon careful examination it turned out to require many assumptions for its validity, and to have several possible meanings. The purists convinced themselves it was unnecessary for dealing with any relevant problem. It was a 'totally useless theoretical toy'. Officially, one might say, it died. And yet it would not stay in the grave. It has such a strong intuitive appeal, and there is nothing better available, so people keep on measuring it. … One suspects that the perfectionist theorists gave up too quickly [Corden (1971, pp. 242–243)].

We should also mention the frequent use made in recent years of the concept of consumer's surplus in the theory of international trade as well as theoretical industrial organization. Would you please comment on these rebuttals, and elaborate your verdict on the use and usefulness of the concept of consumer's surplus?

PS: My thought is very nuanced. In the 2004 debates about protectionism, I've published Ricardo–Mill models where a uniform money-metric utility gives better measures of gains and losses than concave surplus triangles. But I'd never use uniform homothetic axioms to redistribute incomes ethically. Even if the rich and poor did partition their incomes identically—and they don't—I'd be the last one to maximize any sum of "money-metric utility." Corden–Harberg consumer surplus triangles are a very treacherous concept. The correct thing is to look at the indifference curves. It is much clearer in the indifference curve space, and it is even clearer when you do it for Peter alone than when you do it for Peter and Paul together. When you merge Peter and Paul in an aggregate demand curve, and you start taking areas under the aggregate demand curve, in the first place, it is technically wrong—these triangles do not measure anything you want to measure when the marginal utility of money is an endogenous variable. From the very beginning, this was the criticism of Marshall by many different people. There was a 1889 letter from John Neville Keynes to Marshall, in which he wrote: "You are going to be in a trouble on this and you know it is not right. What you pay for the first unit if you are buying only one unit is different from what happens if you are buying others." John Hicks wrote articles on the compensated demand curve and Milton Friedman argued, in a particularly silly article published in the Journal of Political Economy, that Marshall's dd curve was a compensated demand curve. He just did not understand the language which had been used in those days. Arnold Harberger, Chicago's leading applied economist during his time, tried to measure the consumers' surplus triangle. His dogma was that a square inch of area is a square inch of area; you don't have to worry about poor people or rich people; you can aggregate the jelly of Peter with the jelly of Paul, and you have got jelly. Now, what Marshall says is something a little more careful. He says: "Most things affect all classes equally." In other words, they all even out. That goes back to what I said is the underlying principle of most economists of all ages. If you do the thing that increases the size of the pie, it will trickle down, which is a vague law of large numbers. One time it will hurt one group, and another time it will hurt another group. I am sure that Joseph Schumpeter believed in something like that, and the widespread use of the Marshallian consumers' surplus hinges squarely on such a belief to be widely shared.[10] Ricardo famously recanted on his earlier belief that every invention must raise the real wage. Wicksell, Kaldor, Schumpeter and Stigler all believed that he goofed—until I proved that he had not. I did wonder why Ricardo never favored slowing down such inventions. My best guess was that he too relied on the guess that in the long run chance would favor wage growth. A comfortable wishful guess.

3 Welfare Economics and Economic Policy

KS: In your 1981 Bergson Festschrift article, you wrote on the role of competition as follows: "The Pareto-optimality property of competitive equilibrium is no theoretical argument for laissez-faire and is in many situations no cogent practical argument for favoring the use of competition." This interesting observation leads us to a series of questions. In the first place, what, in your opinion, is the main message of the basic theorems of welfare economics? In the second place, what, in your opinion, is the theoretical basis for favoring the use of competition in the allocation of resources? To put it slightly differently, what, in your opinion, is the theoretical foundations of competition policy?

PS: All of the glories of competition are only appropriate when you have constant returns to scale, or when you have replicability so that the lumpiness involving fixed cost gets replicated innumerable times and you have what I wrote in the Chamberlin Festschrift article [Samuelson (1967b)], capitalizing on Joseph's article in 1933.[11] It is true that people frequently refer to the fundamental theorem of welfare economics as a support for promoting competition, but that is a mistake. Someone like Milton

[10] Among many post-Marshallian literature on the concept of consumer's surplus, those who are interested should start their reading with Willig (1976) and Hausman (1981), which present rather contrasting messages with each other.

[11] It was Joseph's (1933) pioneering work that showed how U-shaped cost curves, belonging to replicable plants or to replicable firms under free entry, leads asymptotically to a horizontal unit cost curve for the industry and multiplant firms. Capitalizing on this seminal result, Samuelson (1967b) showed that the possibility of replication leads to "asymptotic-first-degree homogeneity" of the production function.

Friedman does not understand that it is only under a very special institutional condition that you can play the game of competition and get the beneficent results from playing the game. As soon as you have fixed cost, you have a public good problem. Free competition for buying voters does not deliver what free competition for growing corn may.

KS: What is your view on the practical use of welfare economics? To what extent can welfare economics serve as the theoretical foundations of economic policy?

PS: What a lot of welfare economics of my own writing and my own time had been trying is to rule out certain situations as almost universally conceivable as Pareto sub-optimal. Most of my generation have believed that it is better to be on a contract curve than to be off the contract curve. Who won't be for that belief? If your only choice is a point off the contract curve, and you are offered a point on the contract curve which is inside the lense-shaped area enclosed by the two indifference curves passing through the initial point, then you would agree to accept that offer. However, you don't know when you come into negotiation if you are going to end up inside the area enclosed by the two indifference curves in question. This is what is essentially wrong about Ronald Coase. The Chicago School was just delighted when Coase came along and told them: "All you have to do is to set property rights; then no deadweight loss whatever occurs." What they never asked was: "Why should anyone agree to a new situation with property rights, unless they knew their own possession would be as good as, or better than, the status quo?" Under general property rights, the people having property rights end up better off, but people who are excluded end up worse off. The best argument that could be made would be that there should be enough extra gain, and the gainers could bribe the losers. But this argument involves begging the question, as it presupposes that something is correct about the status quo.

My 1974 article entitled "Is the Rent-Collector Worthy of His Full Hire?" put forward an interesting theorem, which was also proved independently by Weitzman (1974), and Cohen and Weitzman (1975), to the following effect. Consider the famous "problem of the common" under general diminishing returns and static conditions, where the free access equilibrium is inefficient whereas private ownership equilibrium is efficient. However, the variable factor (labor) will always be better off with inefficient free access rights rather than under efficient private property rights. If somebody says that there is no content of welfare economics with policy relevance, here is an example of something that is not obvious before you actually analyze it. As I concluded my 1974 article, "Pareto-optimality is never enough [Samuelson (1974, p. 10)]."

Another case in point is my article in the Festschrift for Margaret Hall, which proved that things often get worse before they get better. It is the same set-up as the first case, but now you take half of the common and you make it enclosed, which means that you have private property. On that half, the marginal productivities of variable factors are equalized; on the other half, the average productivities of variable factors are equalized. The naive pre-1935 writer would say: "Surely, it is better to get half of the Pareto-optimality conditions in real life." But it is not. Among the three possible situations, viz. the pre-enclosed common (the situation A), the completely enclosed common (the situation B) and something in between (the situation C), the middle is worse than the either end. There are a lot of suggestions in this little theorem of policy relevance. It suggests why a lot of good improvements don't get done. The same thing applies to the Darwinian evolution. If you can make the big leap, and have feet, then you can get out of the ocean and occupy the land. How can you, by little changes, ever make it worthwhile to make the big change? Of course, the true evolutionist knows there is no mind involved, and it is just a process. There is no selfish gene which is consciously doing this or doing that.

I should also mention Joseph Schumpeter in this context. He spoke repeatedly of the Ricardian vice to the effect that the trouble with Ricardo was that he had too much interest in policy; poor Keynes would have been a better economist than he was if he had been free from the Ricardian vice. I should say people who live in glass houses should not throw stones. Schumpeter, who professed not to give advice, gave me advice all the time. His political thought was very close to Pareto's view, although it was arrived at independently of Pareto. Schumpeter had contempt for the middle classes, because they didn't stand up for their Victorian liberties. He himself wasn't free of the Ricardian vice, which affected his otherwise good work. He was terrible on the Great Depression. He said that it was a good thing when 25% of the population was unemployed, a million homes were in foreclosure, and 10–15 thousand banks shut their doors with no payments to the depositors. One of the uses of welfare economics is to teach you to be alert to study how our ethical beliefs interact with it, and how they contaminate our analytical writings and viewpoints. Pareto who was contemptuous of political viewpoints interfering with economics was the most opinionated man possible. In fact, in the last part of his argument, he analyzed all those irrational things but chose to call them sociology, not economics.

KS: In a recent article entitled "The Strange Disappearance of Welfare Economics," Anthony Atkinson expressed his strong concern about the conspicuous tendency among modern economics in general, and the standard graduate curriculum of economics in particular, to do without what has been the major concern of welfare economics, viz., the foundations of social welfare judgments. Would you please comment on this concern and give your own perspective on the future of welfare economics?

PS: As economists forget about the 1929–1935 Great Depression and the 1939–1945 World War II, they become more tolerant of inequality and own-wallet minded. Voters too display a similar trend in most advanced countries. Like that or hate that, it is a fact.

KS: In your opinion, what are the useful directions toward which welfare economics and social choice theory should be promoted in the future?

PS: I would be remiss not to make the point that, almost as important as being clear about one's BSWF's, in order to be useful in giving policy advice one needs to be sensible about the realistic feasibility constraints that will be binding on every SWF. A true anecdote may be explanatory.

Back before or after 1970, when US College students were everywhere restive about the Vietnam War, MIT's economics students asked me to debate Noam Chomsky, the great linguist and powerful critic of the modern order. I did not feel I could refuse, and to a large audience we two did debate. That day Chomsky was gentle and we were on the homeground of the MIT economics and business departments. What I did not enjoy was to hear some MIT students with almost half a dozen years of economic study utter some surprisingly stupid remarks. Driving home pondering over the afternoon's discourse, I asked myself: "What seemed to be the structure of Chomsky's beliefs?" Rightly or wrongly, I came up with the following hypotheses.

> Chomsky hoped that national and world societies could be economically organized along the model of a harmonious Israeli kibbutz. Families involving several scores of human beings would distribute to each according to sensible needs, while from each would be expected what was proportional to their natural and acquired abilities.

Forget Chomsky and focus on the above old familiar nomination. Based both on recent 1917–2000 experiences in the USSR, Mao's China, Castro's Cuba, East Germany and North Korea, and going back over 5000+ years of economic history, why have large populations that eschewed considerable reliance on quasi-markets been so pitifully unable to attain feasibly high standards of living and comparable rates of growth in longevity and consumption potential? In the end, though I understand the likelihood of inadmissible inequalities and macro instabilities from any laissez-faire market system, I do in the end want to tolerate considerable deadweight inefficiencies inseparable from public interferences with private markets and at the same time want to put limits on those public interferences. If, say, John Rawls were to differ with me and plump for more activism in the Chomsky direction, I think correct analysis of the Rawls and Samuelson BSWF's would find them to be similar (with exceptions)—and yet, differences in how realistic we are might explain most of our policy differences which differences in our transitive norms could not.

4 Concluding Remarks

Paul Samuelson is an almost inexhaustible source of first-hand information on the historical evolution of normative economics. Thanks to his generosity, in this interview we were able to cover many aspects of welfare economics, "old" and "new" as well as social choice theory, with many fresh testimonies which would prove revealing especially to those who are relatively new in the field. Yet there are many aspects of Samuelson's contributions to normative economics to which this interview could not do full justice, including, among others, his monumental work on gains from free trade, his path-breaking work on intertemporal efficiency and turnpike theorems, his pioneering work on overlapping generations economies, to say nothing of his vastly influential work on public goods. A further interview with Paul Samuelson focussing on these aspects of his work seems warranted in order to shed further light on his legacy in the whole area of normative economics. I wish him continued good health, and I am looking forward to learning further from him for many years to come.

Postscript. It is my sad duty to write that Professor Abram Bergson passed away on April 23, 2003. Professor Paul Samuelson dedicated a biographical essay (Samuelson 2004) to honor Professor Bergson, where he wrote as follows: "When Bergson died at age 89, he was the last survivor of Harvard's age of Frank Taussig, and had been a young star in the new age of Joseph Schumpeter, youthful Wassily Leontief, eclectic Gottfried Haberler, and after 1937 Alvin Hansen, the 'American Keynesian.' As Leontief's second protégé I am proud to have been preceded by Abram Bergson, his first protégé, for much of my own work in welfare economics owes virtually everything to his classic 1938 Quarterly Journal of Economics article that for the first time clarified this subject."

References

R1: Books and Articles by Paul Samuelson

Samuelson, P. A. (1947/1983). *Foundations of economic analysis.* Cambridge, Mass.: Harvard University Press, (Enlarged edn., 1983).

Samuelson, P. A. (1950). Evaluation of real national income. *Oxford Economic Papers, 2,* 1–29.

Samuelson, P. A. (1954). The pure theory of public expenditure. *Review of Economics and Statistics, 36,* 387–389.

Samuelson, P. A. (1956). Social indifference curves. *The Quarterly Journal of Economics, 70,* 1–22.

Samuelson, P. A. (1964). A. P. Lerner at sixty. *The Review of Economic Studies, 31*, 169–178.

Samuelson, P. A. (1967a). Arrow's mathematical politics. In S. Hook (Ed.), *Human values and economic policy* (pp. 41–51). New York: New York University Press.

Samuelson, P. A. (1967b). The monopolistic competition revolution. In R. E. Kuenne (Ed.), *Monopolistic competition theory: Studies in impact. Essays in honor of Edward H. Chamberlin* (pp. 105–138). New York: Wiley.

Samuelson, P. A. (1974). Is the rent-collector worthy of his full hire? *Eastern Economic Journal, 1,* 7–10.

Samuelson, P. A. (1977a). Reaffirming the existence of 'reasonable' Bergson-Samuelson social welfare functions. *Economica, 44,* 81–88.

Samuelson, P. A. (1977b). When is it ethically optimal to allocate money income in stipulated fractional shares? In A. Blinder & P. Friedman (Eds.), *Natural resources, uncertainty and general equilibrium systems: essays in memory of Rafael Lusky* (pp. 175–195). New York: Academic Press.

Samuelson, P. A. (1977c). Reminiscences of Shigeto Tsuru. In H. Nagatani & K. Crowley (Eds.), *The collected scientific papers of Paul A. Samuelson* (vol. IV, pp. 897–902). Cambridge, Mass.: The MIT Press.

Samuelson, P. A. (1981). Bergsonian welfare economics. In S. Rosefielde (Ed.), *Economic welfare and the economics of Soviet Socialism* (pp. 223–266). Cambridge: Cambridge University Press.

Samuelson, P. A. (1987). Sparks from Arrow's Anvil. In G. R. Feiwel (Ed.), *Arrow and the foundations of the theory of economic policy* (pp. 154–178). London: Macmillan.

Samuelson, P. A. (1990). When deregulation makes things worse before they get better. In C. Moir & J. Dawson (Eds.), *Competition and markets: Essays in honour of Margaret Hall* (pp. 11–20). London: Macmillan.

Samuelson, P. A. (2004). Abram Bergson, 1914–2003: A biographical memoir. In *Biographical memoirs* (vol. 84). Washington, DC: The National Academies Press.

R2: Books and Articles by Other Scholars

Arrow, K. J. (1951/1963). *Social choice and individual values*, 2nd edn. New York: Wiley.
Arrow, K. J. (1983). Contributions to welfare economics. In E. C. Brown & R. M. Solow (Eds.), *Paul Samuelson and modern economic theory* (pp. 15–30). New York: McGraw-Hill.
Arrow, K. J. (1987). Arrow's theorem. In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new palgrave: A dictionary of economics* (Vol. I, pp. 124–126). London: Macmillan.
Atkinson, A. B. (2001). The strange disappearance of welfare economics. *Kyklos, 54,* 193–206.
Barone, E. (1908/1935). Il ministro della produzione nello stato collectivista. *Giornale degli economisti, 37*, 267–293 and 391–414; English translation, The Ministry of Production in the Collectivist State. In Hayek (1935), pp. 245–290.
Bergson, A. (1938). A reformulation of certain aspects of welfare economics. *The Quarterly Journal of Economics, 53,* 310–334.
Bergson, A. (1948). Socialist economics. In H. S. Ellis (Ed.), *A survey of contemporary economics* (pp. 412–448). Homewood, Illinois: Irwin.
Bergson, A. (1954). On the concept of social welfare. *The Quarterly Journal of Economics, 68,* 233–252.
Bergson, A. (1976). Social choice and welfare economics under representative government. *Journal of Public Economics, 6,* 171–190.
Bergson, A. (1982). Paul A. Samuelson: The Harvard days. In G. R. Feiwel (Ed.), *Samuelson and neoclassical economics* (pp. 331–335). Boston: Kluwer-Nijhoff.
Bergson, A. (1983). Pareto on social welfare. *Journal of Economic Literature, 21,* 40–46.
Buchanan, J. M. (1954). Social choice, democracy, and free markets. *Journal of Political Economy, 62,* 114–123.
Chipman, J. C. (1976). The Paretian heritage. *Revue européenne des sciences sociales et Cahiers Vilfredo Pareto, 14,* 65–171.
Cohen, J. S., & Weitzman, M. L. (1975). A marxian model of enclosures. *Journal of Development Economics, 1,* 287–336.
Corden, W. M. (1971). *The theory of protection.* Oxford: Oxford University Press.
Foxwell, A. G. D. (1939). Herbert Somerton Foxwell, a portrait. In *The Kress library of business and economics, Publication No. 1* (pp. 3–30). Cambridge, MA: Harvard University Press.
Friedman, M. (1949). The Marshallian demand curve. *Journal of Political Economy, 57,* 463–495.
Gaertner, W., & Pattanaik, P. K. (1988). An interview with Amartya Sen. *Social Choice and Welfare, 5,* 69–79.
Gaertner, W., Pattanaik, P. K., & Suzumura, K. (1992). Individual rights revisited. *Economica, 59,* 161–177.
Gärdenfors, P. (1981). Rights, games and social choice. *Noûs, 15,* 341–356.
Gorman, W. M. (1955). The intransitivity of certain criteria used in welfare economics. *Oxford Economic Papers, 7,* 25–35.
Groenewegen, P. (1995). *A soaring eagle: Alfred Marshall 1842–1924.* Aldershot: Edward Elgar.
Harberger, A. C. (1971). Three basic postulates for applied welfare economics: an interpretive essay. *Journal of Economic Literature, 9,* 785–797.
Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics and interpersonal comparisons of utility. *Journal of Political Economy, 63,* 309–321.
Hausman, J. A. (1981). Exact consumer's surplus and deadweight loss. *American Economic Review, 71,* 662–676.
Hayek, F. A. (1931). *Prices and production.* London: George Routledge.
Hayek, F. A. (Ed.). (1935). *Collectivist economic planning*. London: George Routledge & Sons.
Henderson, A. M. (1948). The case for indirect taxation. *The Economic Journal, 58*, 538–553.
Hicks, J. R. (1939). The foundations of welfare economics. *The Economic Journal, 49,* 696–712.
Hicks, J. R. (1940a). The valuation of social income. *Economica, 7,* 105–124.
Hicks, J. R. (1940–41). The rehabilitation of consumers' surplus. *The Review of Economic Studies, 8*, 108–116.

Hicks, J. R. (1943–44). The four consumer's surplus. *The Review of Economic Studies, 11*, 31–41.
Hicks, J. R. (1945–46). The generalized theory of consumer's surplus. *The Review of Economic Studies, 13*, 68–74.
Hicks, J. R. (1959). *Essays in world economics*. Oxford: Clarendon.
Hicks, J. R. (1975). The scope and status of welfare economics. *Oxford Economic Papers, 27*, 307–326.
Joseph, M. F. W. (1933). A discontinuous cost curve and the tendency to increasing returns. *The Economic Journal, 43,* 390–398.
Kaldor, N. (1939). Welfare propositions in economics and interpersonal comparisons of utility. *The Economic Journal, 49,* 549–552.
Kelly, J. S. (1987). An interview with Kenneth J. Arrow. *Social Choice and Welfare, 4,* 43–62.
Kemp, M. C., & Asimakopulos, A. (1952). A note on 'social welfare functions' and cardinal utility. *The Canadian Journal of Economics and Political Science, 18,* 195–200.
Kemp, M. C., & Ng, Y.-K. (1976). On the existence of social welfare functions, social orderings and social decision functions. *Economica, 43,* 59–66.
Kemp, M. C., & Ng, Y.-K. (1977). More on social welfare functions: The incompatibility of individualism and ordinalism. *Economica, 44,* 89–90.
Keynes, J. M. (1936). Herbert Somerton Foxwell. *The Economic Journal, 46*, 589–614. Reprinted in The Collected Writings of John Maynard Keynes, Vol. X, Essays in Biography, London: Macmillan, 1972, pp. 267–296.
Knight, F. H. (1924). Some fallacies in the interpretation of social costs. *The Quarterly Journal of Economics, 38,* 582–606.
Lerner, A. P. (1934). The concept of monopoly and the measurement of monopoly power. *The Review of Economic Studies, 1,* 157–175.
Little, I. M. D. (1950/1957). *A critique of welfare economics*, 2nd edn. Clarendon: Oxford.
Little, I. M. D. (1952). Social choice and individual values. *Journal of Political Economy, 60,* 422–432.
Little, I. M. D. (1999). *Collections and recollections: Economic papers and their provenance*. Oxford: Clarendon.
Mishan, E. J. (1960). A survey of welfare economics, 1939–59. *The Economic Journal, 70,* 197–265.
Myrdal, G. (1953). *The political element in the development of economic theory.* London: Routledge & Kegan Paul. Swedish original edition published in 1930. Translated from the German edition of 1932 by Paul Streeten.
Pareto, V. (1896–1897). Cours d'économie politique, Lausanne: Rouge, (Tome 1, 1896, Tome 2, 1897).
Pareto, V. (1913). Il massimo di utilità per una collettività in Sociologia. *Giornale degli economisti e rivista di statistica, 46,* 337–341.
Pareto, V. (1927). *Manuel d'économie politique*. Genève: Librairie Droz.
Parks, R. W. (1976). An impossibility theorem for fixed preferences: A dictatorial Bergson-Samuelson welfare function. *The Review of Economic Studies, 43,* 447–450.
Peacock, A. (1987). Alexander Henderson (1914–1954). In J. Eatwell, M. Milgate, & P. Newman (Eds.), *The new palgrave: A dictionary of economics* (Vol. 2, p. 638). London: Macmillan.
Pigou, A. C. (1912). *Wealth and welfare*. London: Macmillan.
Pigou, A. C. (1920/1932). *The economics of welfare*, 4th edn. London: Macmillan.
Pollak, R. A. (1979). Bergson-Samuelson social welfare functions and the theory of social choice. *The Quarterly Journal of Economics, 93,* 73–90.
Robbins, L. (1932/1935). *An essay on the nature and significance of economic science*, 1st edn. London: Macmillan (2nd edn., 1935)
Roberts, K. W. S. (1980). Social choice theory: Single-profile and multi-profile approaches. *The Review of Economic Studies, 47,* 441–450.
Robertson, D. (1924). Those empty boxes. *The Economic Journal, 34,* 16–31.
Scitovsky, T. (1941). A note on welfare propositions in economics. *The Review of Economic Studies, 9,* 77–88.

Sen, A. K. (1970/1979). *Collective choice and social welfare.* Holden-Day: San Francisco (Republished, North-Holland, Amsterdam, 1979).

Sen, A. K. (1979). Personal utilities and public judgements: or what's wrong with welfare economics? *The Economic Journal, 89,* 537–558.

Sen, A. K. (1992). Minimal liberty. *Economica, 59,* 139–160.

Sen, A. K. (1993). Internal consistency of choice. *Econometrica, 61,* 495–521.

Sen, A. K. (1995). Rationality and social choice. *American Economic Review, 85,* 1–24.

Sen, A. K. (1999). The possibility of social choice. *American Economic Review, 89,* 349–378.

Sugden, R. (1985). Liberty, preference and choice. *Economics & Philosophy, 1,* 213–229.

Sumner, W. G. (1906). *Folkways: A study of the sociological importance of usages, manners, customs, mores, and morals.* Boston: Ginn.

Suzumura, K. (1987). Social welfare function. In Eatwell, J., Milgate, M., & Newman, P. (Eds.), *The new palgrave: A dictionary of economics* (vol. 4, pp. 418–420). London: Macmillan.

Suzumura, K. (1996). Welfare, rights, and social choice procedure: A perspective. *Analyse und Kritik, 18,* 20–37.

Szpilrajn, E. (1930). Sur l'extension de l'ordre partiel. *Fundamenta Mathematicae, 16,* 386–389.

Weitzman, M. L. (1974). Free access vs private ownership as alternative systems for managing common property. *Journal of Economic Theory, 8,* 225–234.

Willig, R. D. (1976). Consumer's surplus without apology. *American Economic Review, 66,* 589–597.

Young, A. (1912–1913). Pigou's wealth and welfare. *The Quarterly Journal of Economics, 27*, 672–686.

Amartya Sen

W. Gaertner and P. K. Pattanaik

The following is the edited version of an interview conducted on July 4, 1987, with Professor Sen in London.

WG & PKP. When did you first get interested in welfare economics and the theory of social choice? Please tell us how your interest in this area started.

AKS. I have been interested in welfare economics ever since I have been interested in economics. In 1951 I became an economics undergraduate at Presidency College, Calcutta. My original plan was to study physics, but I changed my mind and one of the factors that influenced my decision was my interest in ethics and political philosophy, in addition to the attractions of economics itself. We had some remarkable teachers at Presidency College, which has produced many good economists. Among our teachers, Bhabatosh Datta and Tapas Majumdar took a deep interest in welfare economics.

As far as social choice theory is concerned, I remember that some time in 1952, or maybe early 1953, my friend and classmate Sukhamoy Chakravarty—a truly remarkable intellect—mentioned to me that he was just looking at Kenneth Arrow's *Social Choice and Individual Values,* and he explained to me the nature of Arrow's results. While I found the idea of the "impossibility theorem" to be analytically and politically exciting, I did not see clearly enough its importance to the kind of welfare economic problems I was then interested in, e.g., poverty, unemployment, and exploitation.

This chapter was previously published in the journal *Social Choice and Welfare* (1988) 5:69–79.

W. Gaertner (✉)
Fachbereich Wirtschaftswissenschaften, Universität Osnabrück, Postfach 4469, 4500 Osnabrück, Germany

P. K. Pattanaik
Department of Economics, Faculty of Commerce and Social Science, The University of Birmingham, P.O. Box 363, Birmingham B15 2TT, UK
e-mail: prasanta.pattanaik@ucr.edu

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_5

In 1953 I moved to Trinity College, Cambridge, continuing my undergraduate education. I had the good fortune of having three rather outstanding teachers for whom I did tutorials, viz., Maurice Dobb, Joan Robinson and Piero Sraffa. In late 1954 I worked through Arrow's book with an excitement that I had rarely experienced. I was, in fact, in a state of "high" for weeks afterwards, especially since it was by then clear to me why the propositions established by Arrow were so central to my other—more palpable—concerns in economics and ethics. I tried to get my teachers to discuss with me my new "discovery"—it did appear very much like a discovery since nobody in Cambridge had any great interest in this kind of issue, nor knew anything much about it.

Two out of my three teachers (viz., Maurice Dobb and Piero Sraffa) did, in fact, talk with me about the analytical problems involved in the impossibility result, and one of the three (viz., Maurice Dobb) did encourage me to write an essay on the subject and to pursue the problem some more. Joan Robinson, on the other hand, thought this to be a "waste of time", which would deflect me from my other—allegedly more important—pursuits, and she was convinced that my "weakness of will" made me particularly prone to falling prey to purely analytical temptations. Piero Sraffa also did not believe that I could learn much about the world by exploring the "impossibility theorem" and related problems, but he was sceptical enough of most of economic theory anyway to think that to be a characteristic shared by whatever else I might be spending my time on ("so there is no harm in it, anyway!"). I valued Dobb's encouragement very much indeed. Looking back at that time, I think I was very fortunate in being able to discuss all this with someone who had an altogether different economic and political approach from Arrow's (Dobb was, of course, firmly rooted in the Marxian tradition—indeed he was clearly one of the finest Marxist economists ever produced), and who was nevertheless willing and eager to talk extensively about the relevance, importance and implications of the results presented by Arrow.

I remember also spending a good bit of time talking with Maurice Dobb as to how Arrow's results in this area might influence the theory of resource allocation in general and socialist planning in particular. I had done an essay for Dobb on Lange-Lerner pricing systems not long before that, and we had discussed that problem also in a similarly leisure manner (involving, not surprisingly, Arrow's work on optimality of general equilibrium in that context). Maurice Dobb had written a paper in the thirties criticizing Lange's model among other things for trivialising the problem of distribution, and he was sceptical of the possibility of dealing with this problem either through appropriate redistribution of initial resources (even in a socialist economy, he thought differences in skill and educational achievements to be tremendously influential), or through lump sum transfers of products or incomes (both because of problems of political feasibility, and because of the exacting informational requirements at the centre for determining the distributionally appropriate lump sum transfers—going rather against the motivation underlying Lange-Lerner type decentralisation models). In this respect, I did think that Maurice Dobb was entirely right in his critique.

On the other hand, as far as the *basis* for "social welfare judgements" is concerned, it was clear that Lange and Lerner had a consistent structure, which was not undermined by Arrow's impossibility theorem, since both Lange and Lerner had explicitly based their social welfare function on interpersonally comparable utilities, and the demands of equality and social justice were based on that informational foundation. These were some of the problems I had the opportunity of discussing with Maurice Dobb (and with some undergraduate contemporaries, notably Sam Brittan and Michael Nicholson) at the time I got first exposed to social choice theory (still, by the way, in its neo-natal phase).

As a matter of fact, I did not pursue social choice theory very much after that for many years, and went on to write a Ph.D. thesis on "choice of techniques", discussing the theory underlying some very "practical" problems (such as how labour-intensive techniques to choose in developing economies), with empirical appendices on cotton spinning and cloth weaving in the Indian economy. While social choice theoretic problems implicitly figured in many things I was involved in, I did not get back to formal social choice theory again until the sixties. It was during my visit to the University of California at Berkeley in 1964–65—an interesting year made more exciting by the Free Speech Movement—that I started doing some work of my own in this field. My first two papers on social choice were on the method of majority decision, and my political interests were rather closely mixed with my welfare-economic ones at that stage. Then for quite a few years I was very involved with social choice theory, before moving on to other things. I was then teaching in Delhi, and had marvellous students there, including, of course, you, Prasanta.

WG & PKP. Development economics has been another major area of your interest. Do you feel that the theory of social choice has any significance for development economics?

AKS. Yes, indeed, I do believe that social choice theory has great relevance to development economics. One of the problems with development economics has been the arbitrariness of many criteria that have been typically used in order to assess success. When the subject was first formally initiated during and after the Second World War, the focus was firmly on economic growth, i.e., on the GNP as the measure of success. As an index the GNP suffers from two different types of inadequacies. First, individual income is a very inadequate guide to individual wellbeing or advantage. Second, the method of interpersonal aggregation involved in the calculation of GNP takes the form of simply *summing* individual incomes, without paying any attention to considerations of distributional equity and justice. Nor is any importance attached by GNP to the type of criteria that may be based on recognizing freedoms and rights—either of the "negative" kind involved in the standard theory of liberty, or of the "positive" kind involved in ideas related to guaranteeing to all the minimal capabilities to lead decent lives. The problem of the informational basis of development criteria was systematically neglected in the literature, and here social choice theory did indeed have—and still has—much to contribute.

As it happens, problems of interpersonal comparisons of well-being and those of integrating rights in social welfare judgements did receive substantial attention in formal social choice theory in the seventies, and a lot is happening in these areas even now. These discussions are quite central to development economics.

Also, the political problem involved in making sure that public decisions respond adequately to political actions and activities of citizens—a central aspect of social choice theory—is, in fact, a significant issue in development economics as well. For example, some of us have tried to argue recently that the absence of famines in post-independent India (the last famine was in 1943: The Bengal Famine which killed 3 million people) has much less to do with the rise in food output per head in India (this has happened only to a relatively small extent), and much more to do with the political process in post-Independence India by which the government has to respond to journalistic reports of threatening famines in the form of early cases of starvation, and also to political pressure generated by opposition parties inside and outside the parliament and state assemblies. In this respect India's success in famine prevention can be contrasted with continuing famines in Africa and also the disastrous Chinese famines during 1958–61 (in which estimates of mortality vary between 15 million and 30 million)—even though the Chinese record in eliminating *chronic* hunger (as opposed to famine) is very, very much better than India's. In the prevention of famine, this "social choice" process, it can be argued, is no less important than the exact size of food production and the exact channels of distribution. Even the failure of the Indian economy in eliminating regular and chronic undernourishment can be seen as a *political* failure, and calls for investigation as to why political pressures and pluralities are less effective in dealing with chronic hunger than with more visible famines and starvation.

There are indeed many central "social choice" problems that call for deeper investigation in development economics. The measurement problems involved in evaluating poverty, or living standard, also do include clear "social choice" issues of aggregation. There have recently been many social-choice-theoretic contributions on these measurement problems.

WG & PKP. Impossibility results have received much attention in the theory of social choice. What do you feel about the significance of these impossibility results?

AKS. I believe Arrow's "impossibility theorem" was one of the truly outstanding intellectual achievements ever produced. Now that the mechanism that yields the impossibility is better understood (after many decades of discussion and debate), it is, of course, easy to see how it goes through and also to provide a nearly complete formal proof on the back of a postcard (I have tried to do one in my paper in the festschrift volume for Arrow; Sen 1986), but simplicity is, of course, the characteristic of many of the most profound intellectual contributions.

The theorem can be interpreted in many different ways, depending on the context. In aggregating individual interests in social welfare criteria for economic assessment and in the choice between policies (much discussed before Arrow by such authors as Edgeworth, Marshall, Pigou, Robbins, Kaldor, Hicks, Scitovsky, Bergson, Samuelson, Little, Graaff, and others), Arrow's impossibility theorem established the unviability of the informational base used in contemporary welfare economics at the time when Arrow was writing. After the rejection of interpersonal comparisons of utility (following the effective attack by Lionel Robbins), there were several attempts at arriving at systematic social welfare judgements based on rather poor utility information (no interpersonal comparisons, no cardinality), sticking nevertheless to utility information only (no effective use of non-utility information as in, say, theories of rights or liberties or freedoms, or in theories of justice based on "objective" criteria of individual advantage or freedom). Arrow's impossibility theorem showed that in that informational framework, even moderate demands of consistency and apparent reasonableness (in the form of unrestricted domain, independence, the weak Pareto principle and non-dictatorship) could be mutually inconsistent. While each of these conditions ended up being thoroughly investigated—as indeed they should have been—yielding a rich variety of results related to Arrow's principal theorem (e.g., extending the impossibility to non-transitive social preference and also to non-binary social choice), the informational foundation itself also got critically evaluated. I got personally much involved in developing the structure of informationally more inclusive social welfare *functionals* (as opposed to Arrovian social welfare functions), in my *Collective Choice and Social Welfare* (1970).

In recent years, both in welfare economics and in social choice theory much richer informational bases have been used (e.g., various types of interpersonal comparisons), in contributions by Hammond, Strasnick, d'Aspremont, Gevers, Deschamps, Maskin, Roberts, Kelly, Blackorby, Donaldson and Weymark, Dasgupta, Gaertner, Basu, Dutta, Suzumura, and others. These uses relate to judgements of interpersonal inequalities (e.g., in the works of Tony Atkinson and others), assessment of right fulfilments and violations, examination of demands of justice and so on. John Harsanyi and Patrick Suppes were among the pioneers in this development, but there were others too, such as Franklin Fisher, Serge Kolm, and Tony Atkinson. It may appear agreeably odd that these expansions of informational basis have been tolerated despite the impatient over-confidence that characterizes the positivism underlying so much of the traditional methodological beliefs in economics. Perhaps, this is at least partly because of the recognition that impossibility results are generated by the more austere informational formats investigated by Arrow (as Peter Hammond has rightly emphasized). In this way, a negative contribution has come to play a very creative and positive part in rescuing economics from the narrow box into which welfare economics had got confined.

The impossibility theorem has quite a different type of relevance and significance in the interpretation of political processes, and here the analogy with the voting paradox, which is quite misleading for welfare economics (whoever would want to make distributional judgements on the basis of majority votes?) becomes quite relevant. The fact that political processes of great versatility would nevertheless fail to satisfy these minimal regularity criteria (whether applied to elections, or parliamentary votes, or referendums, or committee decisions) has led both to searches for more radical departures, and to reassessments of the nature and limits of social aggregation. One of the points emphasized by Kenneth Arrow in the latter context, viz., the need for social cohesion in successful aggregation (it is in this context that Arrow had referred to Rousseau's idea of the "general will"), relates to other developments of political philosophy (outside formal social choice theory), which have received attention recently (see Elster and Hylland 1986). These philosophical discussions as well as the logistic issues investigated by voting theories and related studies have great importance of their own, and they too have to be taken side by side with the lessons learned for welfare economics from Arrow's impossibility theorem.

WG & PKP. Your paradox of the Paretian liberal first introduced the notion of individual rights into the formal theory of social choice. Would you please comment on the implications of such rights for actual policy prescriptions of the type with which economists are concerned.

AKS. In fact, the notion of individual rights had been used earlier by Pigou in discussing people's claims to "national minimum standard of real income" in his *Economics of Welfare.* He had characterized them in rather similar ways to what are now called "basic needs". Underlying all this was, of course, Pigou's firm belief that such rights could be justified on utilitarian grounds (in this respect Pigou was in the Benthamite tradition of seeing rights as intrinsically non-important but instrumentally crucial), but much of the discussion in *Economics of Welfare* on this takes place without going much into the *basis* or *justification* of these rights. Indeed, it is not even clear how consistent these Pigovian claims are with his general use of utility criteria to which he was totally loyal—in other sections of his book.

The conflict does, of course, arise from the fact that two principles based on different informational requirements can easily conflict (unless the respective informational bases are completely in line with each other in some systematic way). The impossibility of the Paretian liberal shows the inconsistency that can be produced by combining in a social decision procedure the requirement of the weak Pareto principle, which is based entirely on individual utility rankings (irrespective of the nature of the social states thus ranked), and of minimal liberty, which is based on giving priority to individual preference over a pair of social states specifically in a person's private domain. The notion of the private domain (and the pairs in it) inescapably involves the nature of the choice faced, i.e., the differences between the social states in the pairs. Whether I read some book or not which I happen to own—other things given—belongs to my personal domain, but the choice as to whether Britain should have unilateral disarmament or not does not belong to it, even though I may feel more strongly about the latter. Of course, if preferences are identified with utility rankings, the preferences over personal domains will have *some* relation with utility rankings over other pairs that may be brought in (in order to precipitate the impossibility), and the point of the theorem was to show that nevertheless the ordering properties of individual preferences (and this would extend to cardinal and interpersonally comparable utilities as well) do not prevent such impossibilities from being generated. Giving priority to even two persons' preferences over one pair each in their respective personal domains can generate a cycle when combined with the Pareto Principle being applied to another pair of states.

The fact that this impossibility result requires neither transitivity nor independence of irrelevant alternatives came to me initially as a bit of a surprise, since I was used to thinking of these requirements as basic ingredients of Arrow-type impossibility results. However, this is not really surprising, since the impossibility is generated by different rules of pair-rankings m social choice even tor a given profile (i.e., a given set of individual preferences).

Oddly enough, in a lot of the discussion that has followed this result, it had been implicitly or explicitly assumed that the result applies only to *binary* social choice. However, as I did discuss in *Collective Choice and Social Welfare,* this is not so (pp 81–82), since both the Pareto principle and the demand of minimal liberty can be made directly in choice-functional terms. For example, the Pareto principle may be seen as demanding that if everyone prefers x to y, then y should not be chosen from any set of social states of which x is an element. Similarly, if a person prefers x to y when that pair belongs to his personal domain, then again y should not be chosen from any set of social states that contains x. Together these two principles may leave nothing that can be chosen. Contrary to many claims made in the literature (I have seen three even in the last two months), binariness is completely unnecessary for his result, as you have yourself discussed elsewhere, Prasanta (Batra and Pattanaik 1972).

The significance of this impossibility lies partly in the fact that it is very tempting to combine utility-based criteria (not only the minimal requirement of the Pareto principle but also more extensive demands that entail the Pareto principle and some more) with other requirements intrinsically involving non-utility information (not only minimal liberty but also more extensive demands of rights of negative or positive kinds). By showing the inconsistency of the two types of principles even in their fairly minimal—but unconditional—forms, the theorem shows the fundamental tension between these two approaches. This causes no problem for, say, Bentham (since he saw no merit in intrinsically valuing rights, which had to be justified only on grounds of being good instruments for the pursuit of utility), or Nozick (since he is opposed to any "patterning" of outcomes, whether based on utility aggregation or not), but insofar as many economists, political theorists and social philosophers have more plural sympathies, there is a genuine issue to be faced here.

In my own work, the impossibility of the Paretian liberal has served as one of the elements of a critique of "welfarism" as an approach to social evaluation. If even the Pareto principle cannot be reconciled with minimal acceptance of certain rights, the possibility of constructing an adequate theory of social assessment on utility information only would seem to be remote. The need to move to broader informational bases (discussed, for example, by Rawls, Williams, Scanlon, Mackie, Dworkin, Dasgupta, and others) is indeed strong. The significance of the theorem has sometimes been misunderstood by trying to see it as a problem of logistics. We do know, of course, that in the presence of externalities Pareto optimality may be hard to *realize*. That is not, however, the subject of the theorem. It is concerned with the *acceptability* of the Pareto principle as a criterion of social evaluation.

A number of authors seem to have tried to "resolve" the conflict by suggesting that the parties may "get together" to have a Pareto improving contract. This does not, of course, do anything at all to deal with the problem, since the issue is whether the persons involved have adequately good reason to offer or accept Pareto-improving contracts. In the extensively discussed—I fear much *over*-discussed—example of Prude and Lewd, some people have seen the solution of the conflict in the form of Prude offering to read the book (*Lady Chatterley's Lover*) he would hate to read on condition that Lewd does *not* read that book, which he would love to read. The problem with the contract is not that *others* may have any good reason to object to such a contract being made, but lies in the more elementary question as to whether Prude really has a good reason for offering a contract which would make him affect his own life in a peculiar way (constantly reading a book he hates) in exchange for guaranteeing that somebody else's personal life gets similarly diverted (in his not being able to read a book he would love to read). A similar question applies to Lewd. When the problem concerns the nature of *right action,* it can hardly be resolved by simply assuming that they *must* offer such a contract and accept it. This is the main question. The difficulty in *enforcing* this contract—a policeman making sure that Lewd does not read the book and Prude does not lift his eyes it—is a subsidiary (though not unimportant) problem, with chilling implications for individual liberty in a free society, as I have discussed elsewhere (Sen 1983).

The basic issue is, of course, the inadequacy of the informational base of utility, whether defined in terms of happiness, or desirefulfilment, or choice. If utility is defined in terms of choice only, the ethical force of utility must depend on the *motivation* underlying choice and also the reflective basis of choice. If, on the other hand, utility is defined in terms of happiness only, or as desire fulfilment, that raises problems of its own. The fact that Prude sees that he will be "more unhappy" if Lewd read the book rather than he himself reading it (even though that too would cause him much unhappiness since he hates the book) may or may not be accepted as an adequate reason for choosing to read the book himself in exchange for Lewd's guarantee that he will not read it. The comparison in the space of happiness or desire fulfilment evades the issue of the importance of the respective events. In a somewhat similar context, John Stuart Mill had pointed out that "there is no parity between the feeling of a person for his own opinion, and the feeling of another who is offended at his holding it" (*On Liberty*, 1859).

Since a great deal of traditional welfare economics is based on utilitarianism in particular and welfarism in general, the type of issues raised by the impossibility of the Paretian liberal appears to me to be quite relevant to a critique of welfare economics. Sometimes this relevance has been missed because of misinterpretation of the necessary assumption for this result. Had binariness been central, a way out could have been found through non-binary social choice. But binariness is not needed, and the impossibility cannot be avoided in a non-binary framework either. Similarly, had the results been based on ruling out the rights that individuals have to offer or accept contracts, it would have been appropriate to see the theorem as being based on a misinterpretation of individual liberty. But no right of contract is denied. The fact that either or both may plausibly decide that they do not have adequate reason for offering or accepting such a contract must not be confused with their not having the right to offer or accept such a contract.

While these confusions present in some of the contributions to the subject have added a great deal of fog to the discussion, other writings in the field have clarified the exact scope, reach and significance of the result to welfare economics in particular and social and political philosophy in general. A number of contributions have gone very deeply into these questions. I am thinking, among others, of contributions by Batra and Pattanaik, Nozick, Bernholz, Gibbard, Blau, Fine, Seidl, Campbell, Farrell, Kelly, Ramachandran, Breyer, Ferejohn, Kami, Stevens and Foster, Suzumura, Austin-Smith, Mueller, Barnes, Gardner, Fountain, Green, McLean, Weale, Gaertner and Krüger, Hammond, Schwartz, Levi, Sugden, Basu, Wriglesworth, Kelsey, Riley, Elster, Hylland, Mackie, Coughlin, Webster, Mezetti, Subramanian, Gärdenfors and Pettit, indeed many others. The literature is now, of course, quite extensive (there are, however, excellent critical surveys, presented by Suzumura and Wriglesworth in their books, Suzumura 1983; Wriglesworth 1985). It is helpful that the basic issues have got so thoroughly explored.

WG & PKP. There has been the argument in the literature that the problem of individual rights should not be analysed within the social choice theoretical structure but within the structure of game form. What is your attitude and your reaction to the approach?

AKS. Social choice can, of course, itself incorporate game forms. Consider, for example, Allan Gibbard's demonstration of the inescapability of manipulability of social choice procedures of a certain class. This social choice result is essentially one of the impossibility of finding a game form with the required characteristics. I think the point that is made is not so much a contrast between game forms and social choice theoretic structures as such, but formulating rights in terms of game forms, on the one hand, and in Arrow-type axiomatic social choice formats, involving a normative assessment of outcomes (what Arrow calls social states).

Here, I think the answer must be that *both* game forms and normative social choice theory are relevant to understanding rights and their uses. When I am trying to decide how I should exercise my rights—what strategic considerations I must take into account—it would be obviously important for me to try to understand the game-theoretic characteristics of the situation. In calculating what would be the effect of choosing one thing rather than another, game considerations must clearly come in, and an understanding of the game form implicit in the situation of social interaction would certainly help. On the other hand, when I decide what objectives I should pursue, what kind of outcomes I should want, I have to face the normative question of outcome assessment, i.e., the evaluation of social states. So even at the level of individual action choice, there is no way of avoiding issues that normative social choice theory has focused on. There is, of course, also the prior question of the determination of what rights people should have, and whether our rights can be supported socially even when the result of that support may well be a violation of the Pareto principle. It may be tempting to think that that question does not have to be faced, since the conflict is an unreal one and cannot really arise. One of the points that the "impossibility of the Paretian liberal" tried to make was to assert the necessity to face this question, by showing that the conflict can indeed arise. If it is sensible for us to accept the result of the exercise of individual rights, it would not be sensible at the same time to insist on unvarying Pareto optimality as an essential requirement of an acceptable social state. In assessing social states there is a genuine conflict involved in trying to combine the Pareto principle as a criterion of assessing realised outcomes and the fulfilment of individual liberties in the realised outcomes. In deciding how a person may, in fact, exercise his or her rights, the game-theoretic considerations, if relevant, should enter this analysis. But that does not do away with the need for normatively assessing the outcomes. The relevance of normative social choice theory is in no way compromised by the importance of game-theoretic considerations in the choice of actions, or by the cogency of seeing individual interrelations in terms of appropriate game forms.

Perhaps what those whom you are quoting really wish to say is that the exact content, relevance or rationale of social choice axioms, relating individual interests or judgements or choices to optimal (or acceptable) social states, cannot be properly understood without bringing in the game forms involved in the decision systems. That claim is sensible enough even though sometimes the game forms involved are so rudimentary that it would be peculiar to make heavy weather of it. But in some contexts the game-theoretic structures may indeed be important to clarify, and as it happens, this aspect of the "impossibility of the Paretian liberal" has indeed been fairly extensively explored by such authors as Hammond, Gardner, Breyer, Aldrich, Basu, Levi, Suzumura, Wriglesworth, Coughlin, and others. The important thing is not to lose sight of the normative assessment issues as the game-theoretic aspects of the problem are better explored. The respective relevance of two different—but interrelated—aspects of the problem must not be confounded.

WG & PKP. What are the directions in which you would like to see the theory of social choice develop further in the future?

AKS. This is hard to say, since social choice theory is such a vast field. As far as the welfare-economic aspects of social choice theory are concerned, the need for greater integration with ethics and moral philosophy is, I believe, quite strong. Some of the issues I have already commented on in this interview relate closely to discussions in philosophy. The question of interpersonal comparisons, for example, is one of them (the central issues have been discussed very illuminatingly in a recent contribution by Davidson 1986). The reliance of ethics on utility-based considerations has been critically rejected by such philosophers as John Rawls, Bernard Williams, Tim Scanlon, Ronald Dworklin, and others. Even those who have defended utilitarianism or welfarism in some more sophisticated form (e.g., Richard Hare, John Harsanyi, Derek Parfit, Allan Gibbard, Jim Mirrlees, Peter Hammond, Jonathan Riley, Dale Jorgenson and Daniel Slesnick), have raised very substantial and interesting questions in the context of their reformulation of utility and the informational basis of social evaluation and choice. Similarly, recent discussions on issues of equity and justice have opened up other areas. Certainly, one direction in which social choice theory can fruitfully go is that of greater integration with moral philosophy. Unfortunately, a great deal of the philosophy that occurs in traditional economics—usually in an implicit form—is very rudimentary and sometimes confused (crude positivism is one—but not the only—problematic influence on traditional economics), and there is much to be said for taking the best and the most exacting of the contemporary philosophical traditions and expanding the reach and strength of social choice theory.

Some problems in social choice theory are of course primarily logistic. These include complicated decision problems, which have been unearthed by social choice theory itself in the process of dealing with impossibility results, or in search of positive possibility characterizations. Here again, there is scope for many advances based on sophisticated and substantial decision analysis. Contributions by Howard

Raiffa, Isaac Levi, Richard Jeffrey, Kahneman and Tversky, Mark Machina, and others are indeed extremely interesting, and they are also of immediate and far-reaching relevance to social choice theory, as indeed Levi has clearly shown (Levi 1986).

As far as discussions of voting problems are concerned, much can be gained from a better understanding of the literature on political organization. Michael Dummett in his important book on *Voting Procedures* (Dummett 1984) notes with some amazement that between the bibliography of Vernon Bogdanor (*The People and the Party System*) and the bibliography in my chapter in the *Handbook of Mathematical Economics* on "Social Choice Theory", there is "only a single writer common to the two", viz., J. S. Mill! I should perhaps add that quite a few of the writings in politics, particularly in mathematical politics (e.g., by Kramer and others) figure in the social choice theory literature surveyed in my paper, but not in the other "political" literature that Dummett cites, based on Bogdanor's survey. The line of division has often been in terms of mathematical techniques and analytical logistics rather than the subject matter. That is, of course, an unnatural division, and Dummett is certainly right that a good deal can be gained by putting the two traditions together, as indeed he has tried to do.

There are many other developments one can think of, but these are some of the lacunae that seem to me to be obvious and glaring, as matters stand at the moment. There is, I believe, much future in social choice theory. The need for departures is particularly important to emphasize, since social choice theory in recent years seems to have entered a baroque phase, and detailed ornamentation of relatively narrow problems has often replaced more probing queries. This is a temptation that is present in any subject as it matures, and especially in a subject in which the scope for technical virtuosity is quite considerable. But the subject has much to gain from remaining innovative and progressive, rather than inward-looking and obsessive. Perhaps it may not be inappropriate to mention here that now that there is a specialist journal in social choice theory, viz., this one, it is particularly important to make sure that the tradition of the subject remains open and dynamic rather than closed and stuck. Much will depend on it.

References

Batra, R., & Pattanaik, P. K. (1972). On some suggestions for having non-binary social choice functions. *Theory and Decision, 3,* 1–11.

Davidson, D. (1986). Judging interpersonal interests. In J. Elster & A. Hylland (Eds.), *Foundations of social choice theory*. Cambridge: Cambridge University Press.

Dummett, M. (1984). *Voting procedures*. Oxford: Clarendon Press.

Elster, J., & Hylland, A. (Eds.). (1986). *Foundations of social choice theory*. Cambridge: Cambridge University Press.

Levi, I. (1986). *Hard choices*. Cambridge: Cambridge University Press.

Sen, A. (1983). Liberty and social choice. *Journal of Philosophy, 80,* 5–28.

Sen, A. (1986). Information and invariance in normative choice. In W. Heller, D. Starrett, & R. Starr (Eds.), *Social choice and public decision making*. Cambridge: Cambridge University Press.

Suzumura, K. (1983). *Rational choice, collective decision and social welfare*. Cambridge: Cambridge University Press.

Wriglesworth, J. (1985). *Libertarian conflicts in social choice*. Cambridge: Cambridge University Press.

Developments

Salvador Barberà

Carmen Beviá and Jordi Massó

It was a nice spring afternoon in Barcelona on April 6, 2017. We met at 5:00 pm at Salvador's place to interview him. We chose this day because Carmen was in Barcelona; the day before she attended the thesis defense of her last Ph.D. student at the UAB. When Salvador received us, Cesca (Salvador's wife) was not in yet, but she joined us later for dinner. We sat in a beautiful living room. In one of its walls hangs Perico Pastor's portrait of Salvador, the present given to him by his friends on the occasion of his 65th birthday celebration, back in June 2011. We tried to follow our prepared questionnaire, but often he moved quickly to his (and quite abstract) work with Dolors Berga and Bernardo Moreno. Salvador wanted to know our reactions about his present and future work. But we finally were able to bring him back to the past. What follows are his answers to our questions.

Q. Incentives and Northwestern in the early 70s: how important was this environment in shaping your interests?

A. The intellectual atmosphere at Northwestern, in the early seventies, was very exciting, due to the presence of different, attractive persons.

Stanley Reiter was a crucial person in the Economics Department and the Business School. He had worked in Purdue before arriving at Northwestern, along with

We thank Eduard Talamàs for many helpful suggestions.

C. Beviá
Universidad de Alicante, 03690 San Vicente del Raspeig, Alicante, Spain
e-mail: carmen.bevia@ua.es

J. Massó (✉)
Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès), Barcelona, Spain

Barcelona GSE, Ramon Trias Fargas, 25-27, 08005 Barcelona, Spain
e-mail: jordi.masso@uab.as

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_6

other very renowned people like Vernon Smith and Charlie Plott and had produced there a generation of Ph.D.'s who eventually ended up at Northwestern along the years: Hugo Sonnenschein, Morton Kemion, Nancy Schwartz, John Ledyard had been his students. He created the Math Center and had a magnetic influence on many of us, through a legendary course on Debreu's *Theory of Value* (Debreu 1959) that he taught by asking—from the back seat of the classroom—the most penetrating and educational questions to the students who presented this book. In addition, Stan was one of the promoters—along with Leo Hurwicz, Tom Marshak and Roy Radner—of annual meetings on decentralization, where a small but intense group of people presented work that gradually built the basis for what was to become the large industry of present day mechanism design. These pioneers looked at specific allocation methods as points in a multi-dimensional space of mechanisms, within which the different methods could be compared and classified according to their performance on several dimensions. Decentralization was a major concern of the times, along with efficiency and coverage, as measured by the amplitude of the set of economies for which each mechanism was well defined. And it was one of these people, Leonid Hurwicz, an inspiring personality who visited Reiter very often, who explicitly added a new criterion of performance to that list of dimensions: that of incentives, on which I would end up working.

Along with Reiter and Hurwicz, I had the privilege of meeting many other people whose work is at the origin of incentive theory and mechanism design. John Ledyard was my first micro teacher, and I could see him developing different approaches to study incentives, and using a variety of game-theoretical concepts. And mind you, at this time, Game Theory was not taught in the Economics Department, nor at the Business School, although I was lucky enough to take a course on the subject from a recent arrival at the School of Industrial Engineering: a very young Bob Rosenthal. Ted Groves arrived with his new ideas regarding incentives in teams, which later on he developed into his much richer class of mechanisms. I remember very well how he rehearsed for the Minnesota meeting where he presented his ideas through joint work with Martin Loeb (Groves and Loeb 1975), who was one of my classmates. He certainly understood that he had gold in his hands, and he managed to convey his message so well that I have never seen such an enthusiastic reaction to a single paper as the one Ted obtained in that meeting, which I attended as a student. I remember Hurwicz jumping to the blackboard and trying to generalize the paper on the spot, even before Ted had concluded. This was a thrilling moment that gave me an overview of what is best in our research profession. I had seen a paper grow, an author working hard to present his discoveries, and a great idea to succeed in front of a demanding audience. And more people were working in the area: faculty members like Elisha Pazner, Jean Marie Blin and John Roberts (who just arrived to Northwestern from Minnesota), and students, who also made early contributions to the theory of mechanism design, like Andy Postlewaite [who produced a beautiful paper with John Roberts (Postlewaite and Roberts 1976)], Jim Jordan (my classmate), and my friend Xavier Calsamiglia (at Minnesota).

An important arrival to Northwestern was that of Mark Satterthwaite, just after he had proven, independently, the fundamental result that we all know by now as the Gibbard–Satterthwaite theorem (Satterthwaite 1975).

Antonio Camacho taught me a course on social choice, based on Amartya Sen's recently published book (Sen 1970). Antonio was a Spaniard who made all his career in the USA, a former student of Hurwicz, also interested in mechanism design, a very original and independent thinker, always a bit critical of the fads that are so influential in Economics research, where some topics are deemed essential today and then forgotten or even declared dead the day after.

And then, Hugo Sonnenschein joined Northwestern. Hugo was already a star professor at 34, when he arrived. And I knew him indirectly because he had co-authored the first published paper of my Catalan friend Andreu Mas-Colell, then a second-year student at Minnesota (Mas-Colell and Sonnenschein 1972). The paper was an important contribution to social choice, a subject on which Hugo has written great papers, and that he has always defended in many ways. Hugo became my advisor during my fourth year, and his advice was absolutely decisive in my career. I was his third Ph.D., which places me among the first of a long list of people who enjoyed his guidance, many of whom populate the best universities in the world. John Roberts—who was his second Ph.D.—and myself proudly call these other students of Hugo our younger brothers, because we came well before Hugo's spectacular success at Princeton.

Q. Why strategy-proofness?

A. Because I had grown in an atmosphere marked by the idea of mechanism design and found myself immersed in the excitement regarding the new developments about incentives. I always wanted to write on that subject and from that perspective. I am talking about the first half of the seventies, when the Gibbard–Satterthwaite theorem has just been proven, the Groves mechanisms were being gradually understood, and Hurwicz was adding the incentives dimension to his grand vision of mechanism design. All of this happened around my university, if you add the seminars I attended there by Gibbard, and by a very young Jean-Jacques Laffont.

The ambition to write about incentives was well into my mind, but my abilities failed me on several attempts. One of these involved trying to show that manipulating the competitive mechanism would require a costly effort to refrain from consuming the preferred affordable bundle of goods, and that this cost might greatly reduce the danger of manipulation, as exhibited in Hurwicz's work (Hurwicz 1972). That led me to try models of adjustment where people could influence the dynamics of prices: The project eventually ended after taking a course on differential games in the math department, that proved to be well above my capacity!

Finally, one day, Hugo met me in the corridor and asked me a simple question. What happens to the Gibbard–Satterthwaite theorem if you have multi-valued social choice functions? And that changed my fate as a graduate student. Here was a question that still related to incentives, but changed my focus from exchange economies to social choice functions. I had learned with Camacho about this framework but had not tried to work on it. I soon realized that I felt very comfortable when working from this new perspective. Indeed, this framework turned out to be very friendly to me. From then on, a thesis emerged in a few months, whose articles were very well published.

Q. Indeed very well: ***Journal of Economic Theory*** **(1977,** Barberà 1977a**) and** ***Econometrica*** **(1977,** Barberà 1977b**). What would you say is the leitmotiv of your research?**

A. The importance of simple questions is something that has always fascinated me. A simple question does not always have a simple answer, and when this is the case, you have a beautiful example of what Reiter used as metaphor of research. He described the researcher as someone who walks a path full of pebbles and chooses which ones to look under. Some of them may hide the entrance to fascinating caves, while others may just have a bit of dust under them. The manipulability of multi-valued social decision functions did not only ask for a wider definition of manipulability, but also raised questions about ranking sets, on which I wrote later as an independent topic, and about the introduction of lotteries in social choice. That topic was the object of my first paper co-authored with Hugo (Barberà and Sonnenschein 1978) and came from a second simple question that he asked me while waiting at a traffic light: what happens when we enlarge the framework of Arrow's theorem to allow for lotteries over preferences to be the result of aggregation? And again, the consequences of writing that paper went beyond the starting quest and led to my interest for stochastic preferences, on which I worked with Prasanta Pattanaik (Barberà and Pattanaik 1984; Barberà et al. 1984; Barberà and Pattanaik 1986) and for probabilistic judgments, a subject I developed with Federico Valenciano (Barberà and Valenciano 1983). Let me just add another simple question. Again, while walking, Arunava Sen asked me: what if agents, instead of always choosing their best available alternative, would choose either that or their second best one? It took me many efforts and those of Alejandro Neme, until we found a way to start from that simple question and to find a characterization of agents' behavior that follows this or similar patterns, where full optimization is not attained but people are still guided by a referential preference order (Barberà and Neme 2015).

Q. Your large network of scientific collaborations has been important for you, hasn't?

A. Yes, it has. Scientific collaborations are important in general, and they have been particularly in my case. Since my first World Congress of the Econometric Society (Toronto, 1975) I was lucky to meet people who were to become co-authors, friends and sources of inspiration for my work. This was the first congress where I presented a paper (Barberà 1977b), and I was lucky to meet there respected seniors like Amartya Sen and Peter Fishburn, to connect with Prasanta Pattanaik—whose papers had been crucial for my thesis—and to start long lasting connections with Claude d'Asprémont, Louis Gevers and Louis-André Gérard-Varet. Soon after, I made additional friends in the profession through early meetings, like Jean-Jacques Laffont and Hervé Moulin, and I got to know Eric Maskin and Roger Myerson during heated discussions about the revelation principle. A second wave of very important contacts came thanks to

Maurice Salles' initiative to organize a Social Choice and Welfare congress in Caen, in the summer of 1980, from which a book emerged. Along with other people I already knew, I first met Bhaskar Dutta there, and strong connections were established, eventually leading to the creation of our journal, first, and of the Society for Social Choice and Welfare later, after a new interesting meeting that I jointly organized in València with Hervé and Antonio Villar. Another extraordinary opportunity to meet people and renew ideas came during a visit to Stanford in 1984–85. Hugo was also visiting, and along with him a large group of his Princeton students. They were many, now important names in the profession, but let me just mention two of them, Arunava Sen and Faruk Gul, who became my co-authors later on (Barberà et al. 2001a, 1993). Also among the recent arrivals from Princeton was a first-year graduate student, Matt Jackson. I was about to become his micro teacher but luckily he got exempted from the course: instead, we wrote our first joint paper (Barberà and Jackson 1988). After so many years, we still write together, and I keep learning from him, as I did even at that early stage! I could go on and tell you about more people with whom I shared the joy of discovery: Michael Maschler, Federico Valenciano, Anke Gerber, Jordi Massó, Alejandro Neme, Lin Zhou, Shige Serizawa, Anna Bogomolnaia, Ennio Stacchetti, Lars Ehlers, Danilo Coelho, José Alcalde, Dolors Berga, Bernardo Moreno, Clara Ponsatí, Carmen Beviá, Antonio Nicolò, Walter Bossert. At least one special story hides behind each of these collaborations, maybe more. And if I keep an active research agenda after becoming emeritus, I owe this privilege to a collection of co-authors who are also friends, and with whom I share, among other experiences and interests, the thrill of looking for a good and challenging question.

Q. Why restricted domains?

A. Much of my work is associated with the idea of strategy-proofness. Indeed, this was already the topic of my dissertation, and I have kept working on the subject. I already told you that initially I got there through a lucky combination between my interest in incentives and my comparative advantage when using the techniques of Social Choice theory. But my approach to the topic evolved with the passage of time. At first, I proved impossibility theorems. Then, for a while, I moved slightly away from dominant strategies without plunging too much into the wave of works that used a Bayesian perspective. Still fascinated by the simplicity of certain solution concepts, I explored the concept of protective equilibrium, with Bhaskar Dutta, and its consequences (Barberà and Dutta 1982). I also continued to work with the idea of dominant strategies in contexts where social outcomes are lotteries. Assuming that voters have von Neumann–Morgenstern preferences in that context was already a first way to introduce restrictions on their preferences. And I was happy to produce a number of direct proofs of the Gibbard–Satterthwaite theorem and of Arrow's as well, using the notion of pivotal voters and of option sets (Barberà 1980b, 1983), as a result of my repeated visits to these seminal sources. It was only later, into the nineties, that I started looking more systematically at the possibility that, although not universally, there might be many instances where non-trivial dominant-strategy mechanisms could be defined. The fundamental and beautiful work of Hervé Moulin on strategy-proof rules under single-peaked preferences was a tremendous stimulus

(Moulin 1980). Thanks to a series of papers with Hugo and Lin (Barberà et al. 1991), Faruk and Ennio (Barberà et al. 1993), Jordi, Alejandro and Shige (Barberà et al. 1997, 1998a, 1999, 2005) I gradually understood the possibility and the difficulties associated with an extension of domains to cover the case of multi-dimensional alternatives. And, more recently, I have worked hard with Bernardo and Dolors in order to provide additional results regarding domains where it might be possible not only to promote individual incentives, but also to avoid group manipulations, and to attain efficiency (Barberà et al. 2010, 2012a, b, 2016).

Of course, looking at domain restrictions and sticking to dominant strategies is only one of the many avenues of research suggested by the Gibbard–Satterthwaite theorem. A major line that was also opened by that result is the invitation to accept that voters play games and go to a full-fledged game-theoretical analysis of their interactions through different mechanisms. The literature on implementation flourished for a while. More persistent is all the literature that models economic agents as the players of Bayesian games, fit to model a large variety of voting and economic situations. That literature was and still is mainstream, but I see an increased interest in going back to examine the simple idea that, although hard to attain, strategy-proofness and its close relative expost incentive compatibility are very interesting properties to demand, and ones that may sometimes be satisfied. Part of my fascination with strategy-proofness and dominant strategies, or the equilibrium properties of truth-telling, undoubtedly come from my limitations as a game theorist. But I also like to think that my life-long commitment with these simple topics comes from the strength that I have also appreciated in them. The idea that more and more people explore them from different angles is comforting.

Q. What research environment did you find in Spain on your return?

A. When I finished my Ph.D., I returned to Spain immediately, as did most of the few other people who had pioneered the quest for a doctoral degree in Economics in England or the USA. A notable exception was Andreu Mas-Colell, who developed most of his brilliant career in the USA. But he always kept well connected with the Spanish reality, and he played a very helpful role for Spanish economics from his American base. Other than him, though, most of us returned to Spain and found that there was a very stimulating task ahead. Not an easy one, but a stimulating task. I returned in the summer of 75, started to work in Madrid in October, and Franco died one month later. He had been in power since years before my birth. We hardly knew what was ahead for our country, but we were full of hope. I was 29 and felt part of a generation that could improve the country at large. And changing the university was part of the task, one about which I had great hopes and was prepared to work on. Academic research in Economics was practically non-existing. Academic journals published translations of articles, and original contributions were rare, often summarizing the readings of those people who read anything at all. I attended the 1975 World Meetings of the Econometric Society in Toronto, along with half a dozen Spaniards, none of us tenured. For many years, any time I met with Andreu in a congress we would count how many more we were, and felt proud of seeing how numbers increased steadily. This increase was partly the result of a social trend toward open relationships with the world that was experienced across the whole of our society. But it was also the result of very deliberate efforts to convince younger generations of Spanish students that investing in graduate studies abroad and then doing research in our country was feasible and exciting. We were in small numbers, but very committed to the task. Joaquim Silvestre, Joan Esteban, Xavier Calsamiglia and Isabel Fradera in the Autònoma de Barcelona, Juan Urrutia, Federico Grafe, Carmen and Inmaculada Gallastegui in Bilbao, Javier Ruiz Castillo and Carlos Escribano in Madrid, myself in Madrid and then in Bilbao, and finally in Barcelona. We preached the good news of economic research, starting from very modest beginnings. In 1977, I published in *Econometrica* and *Journal of Economic Theory* (Barberà 1977b, a), and so did Calsamiglia (*Journal of Economic Theory*, Calsamiglia 1977) and Silvestre (*Econometrica*, Silvestre 1978). These were all-time firsts: No one had published at this level from Spain, and this, along with Andreu's successes in the USA, were proof that there was a future ahead for research of competitive quality. We started sending students abroad, and they tended to come back, because the country was getting more attractive, universities were growing, and governments became more supportive. All remained somewhat local for a while, until new universities and research centers started emerging that shared the same values of excellence and joined forces with the pioneers. But it was no easy task, as we started from very low. I directed the *Revista Española de Economía* for about 15 years, and thanks to a large number of young collaborators it became the first Economics journal ever refereed in Spain, gradually turned into an English language publication, and eventually became the *Spanish Economic Review*. Other journals followed suit, in particular *Investigaciones Económicas*, directed by Rafa Repullo. These journals provided many young Spanish economists their first chances to learn and practice the principles of anonymity and rigor that are required from a professional researcher and that had been missing before in our culture. We also communicated with other emerging groups that were also rowing against the current in their own countries. Collaboration with Jean-Jacques Laffont's group at Toulouse and Louis-André Gérard-Varet in Marseille led to the creation of a Southern European Association (ASSET) which was later to expand, but provided strength to our somewhat isolated groups, and got voice through a joint series of working papers called SEEDS. The group of researchers at Bellaterra started a yearly congress, the Symposium, that was to become twenty-five years later one of the pillars, along with the *Spanish Economic Review*, on which a Spanish Economic Association emerged. I cannot mention by name all the people who added their efforts to the continued task of building a research profession in an incredibly short time, between 1975 and 1990. By then, when I chaired the local organizing committee of the Econometric Society World Congress, we had attained a size that allowed projecting the image of a research community on the go. That image is now much stronger, but it was very rewarding to be part of the initial push, and Barcelona 1990 was certainly a turning point.

Q. But your first job in Spain after the return from the USA was in the Universidad Autónoma de Madrid, right? How was it?

A. My first two years after the doctorate, at the Autónoma de Madrid, were very momentous. It was a time of change in Spain, and democracy needed some years to consolidate, to change regime. I was an interested observer of this change, but concentrated strictly on getting my work published, and was very lucky at these first stages of my career. Less than two years had passed when I got a permanent position in a rather disputed contest. Rules for these promotions to tenure were simply crazy, including six different exercises and the writing of a book on methodology, with a minimum of 250 pages. Luckily, I had two acceptances from *Econometrica* (Barberà 1977b, 1979a), two from *Journal of Economic Theory* (Barberà 1977a; Barberà and Sonnenschein 1978), one from *The Review of Economic Studies* (Barberà 1979b), a contribution to a book (Barberà 1978) and a couple of articles in Spanish (Barberà 1979c, d), all in twenty months.

Q. Not bad! And all from Spain.

A. I often wondered how all this could happen in such a short time. I get quite mad at the idea that our profession is becoming more and more cruel with young people, due to extreme and unnecessary delays in publication times. My new position was in Bilbao, and even if I would have preferred to join the Autònoma de Barcelona, I moved to that city, where I had gotten my undergraduate degree. I was lucky to meet there with young colleagues who had got degrees in England, and with a young professor, Juan Urrutia, who became a lifetime friend with whom I have developed a great deal of projects. Juan has been a government official, a founder of Universidad Carlos III de Madrid, a banker, a philanthropist. We created an Institute of Public Economics, founded ASSET, started a Master's program that sent many Basque students to study abroad and is even now giving us support at Bellaterra through a small research foundation, MOVE, that was created to circumvent some of the endemic problems that our large, rather sclerotic universities, impose on any innovative project for change. Bilbao provided an exciting opportunity for individual and collective work, and we ended up spending nine years of our life there. My children grew up there, my wife wrote her Ph.D. dissertation and started her successful career as an economic historian there. It is a second home, full of former students and colleagues. And it has become a beautiful, vibrant city.

In 1986, we moved to Barcelona, where we were born, after a twenty-years detour. I was forty, full of projects, had started to have some recognition and had acquired some organizational experience in Bilbao. I arrived at a place that already had developed some of the more ambitious projects and was ready for more. Many students abroad were getting ready to return, and facilitating their arrival was one of the first and greater pleasures I've had in academic life. And as more and more well prepared people joined our ranks, we were ready for more ambitious endeavors: the doctoral program, the extension of our basic principles to other universities and research centers, the creation of a conscience that we were part of a revival of economic research at the European level.

Q. Yes, and then you promoted the creation of IDEA, the International Doctorate in Economic Analysis, which was the first program in Spain to be entirely taught in English and open to students from all countries.

A. I had been very reluctant till the early nineties to retain Spanish students in Spain for their doctoral studies. I thought that the best we could do for our students is to send them abroad. But we had an increasing potential to teach a great graduate program, based on the arrival to Bellaterra of excellent people with degrees from abroad. People like Jordi Massó, Jordi Caballé, Inés Macho-Stadler, David Pérez-Castrillo, Xavier Martínez-Giralt, Pau Olivella, Clara Ponsatí in my department, and Xavier Vives, Joan Maria Esteban, Carmen Matutes or Joszef Sakovicz at the Institute of Economic Analysis, along with multiple visitors, were the basis for a new project, not oriented toward our own undergraduates, but open to people from other parts of Spain, Europe and the world. It was hard to break with many of the conventions that did plague our universities, too much used to think of graduate students as cheap labor force in the lab or servants of their master for years, and to create an open program where people could identify their interests, choose their advisors, select topics and measure their own strengths. A program that would lead people to mature as integral, independent researchers and help them integrate into the international research community, open to all subjects and interests. That was our ambition, which we shared with other leading universities and allowed us to form ENTER, a consortium that involved from the start Toulouse, Tilburg, Mannheim and University College London. It always was and still is hard to sustain such program because of many factors. Our universities have not evolved much, the philosophy underlying Ph.D. programs in Economics is not the same as in that of natural Science departments, and many fellowship programs were severely cut during the recent economic crisis. Yet, the program gave my department a unity of purpose and I am proud of having promoted it.

Q. But let's start from the beginning. You studied Economics at the Universitat de Barcelona, but your Bachelor degree is from the Universidad del País Vasco at Bilbao. What happened?

A. I had been a very good student prior to the University years, at the French Lycée of Barcelona and had always enjoyed learning. I was fascinated when introduced to the axiomatization of integer numbers and also enjoyed the rudiments of formal logic. Just before starting my Economics courses, I was a finalist of the first Spanish Mathematical Olympics, but in fact I now understand that it was only because I could solve a combinatorial problem that you could easily reduce to the characterization of majority voting, even if it was not worded like that. All other problems involving differential calculus escaped me, but this was a premonition. Later on, at Northwestern, I also took several courses in math, but eventually I made my living with a few combinatorial calculations and a little logic. I think it is, ex post, another of the nice facts of research: With a little luck and some work, you may discover where, in the large menu of technical options, lays your relative advantage.

Unfortunately, the university I found in 1964, after leaving high school, was not attractive at all. With very few exceptions, the quality of courses ranged from mediocre to very low. Students were a crowd: My class had hundreds of people enrolled in a single classroom. And there were more interesting things to do. In 1964–65, a student revolt was cooking against Franco's fascist-like trade union system, and I quickly engaged in activities that were more exciting than attending rambling classes by teachers I didn't find too exciting. Now I understand that this involved a lot of arrogance, and that some of them, if properly approached, could have taught me quite a bit. But arrogance at 18 is quite natural, and I ended up essentially wasting my four years at the University, from an academic point of view. But it was not a waste in other dimensions. I felt part of the trend that was starting toward a democratic society, met many of the people that would end up being partners in my US adventure, and Cesca, my lifetime companion. Yet, things were not good for learning, I got expelled from the Universitat de Barcelona and sent to Bilbao, where I finished my five-years program in four years, just in time to join the Spanish army as a conscript, for an extra fifteen wasted months. What remained from that period was an unsatisfied hunger for knowledge and the certainty that I had to catch up after years without learning Economics, even if I had learned some things about life. I still remember the thrill I got when Alfred Pastor, one of my friends who went to MIT two years before I went to the USA, sent me a letter that I read while stationed in Ceuta, where I was a soldier. He had just taken a ride on the same elevator than Samuelson, and that made my day: I already felt close to knowledge, through this connection. Just imagine how little we knew, and how large our hope that we could make up for lost time in the USA. I was not even thinking about research: just about learning and catching up!

Q. Although you are best known for your work on strategy-proofness, you have been interested in other topics in Social Choice and also on other subjects relating to individual and collective behavior.

A. I have been fascinated for years by the idea of stochastic choice, which I discovered through work with Hugo (Barberà and Sonnenschein 1978) and developed with Prasanta Pattanaik (Barberà and Pattanaik 1984; Barberà et al. 1984; Barberà and Pattanaik 1986). Some of the projects on which I have spent more unproductive time and effort lie in this area. Even if I could not solve them, I am glad to see that stochastic choice becomes more and more attractive from a theoretical point of view, and that young colleagues like Miguel Angel Ballester and José Apesteguia are among those who push the frontiers on the topic.

Another area that I worked quite a bit on seems rather abstract: how to extend orders from a set to its power set. The initial needs that I felt when writing my dissertation led me to realize that many researchers attribute to economic agents the ability to rank sets of objects and require these rankings to be "consistently" connected with the way in which the same people rank singletons. However, "consistency" can take many forms, and this is because sets of objects admit many different interpretations. I enjoyed writing a survey of the subject with my good friends Prasanta Pattanaik and Walter Bossert (Barberà et al. 2004), which gave us an opportunity to touch upon a variety of models that use apparently similar ideas but reflect of very different issues.

I have also enjoyed working on the subject of coalition formation, from different angles, in works with Michael Mashler (Barberà et al. 2001b), Anke Gerber (Barberà and Gerber 2003, 2007), and Carmen Beviá and Clara Ponsatí (Barberà et al. 2015). I have learned the hard way that even the most apparently simple and easy to describe models may become devilishly hard to deal with. Hedonic games are an example. But I continue to be very interested in understanding coalitional structures, and even if again some other people have taken paths that I could not follow, I am always looking for the next challenge that I can address in this area.

As for Social Choice theory, I have enjoyed the stimulus of different people and been able to produce papers with them, a double reason for pleasure. Working with Matt Jackson is a permanent source of challenges and satisfaction. One of the hardest papers we wrote, on strategy-proof exchange (Barberà and Jackson 1995), was also one of the most rewarding. But that was still on incentives. Later on, we have worked on papers like "Choosing how to Choose" (Barberà and Jackson 2004), "On the Weights of Nations" (Barberà and Jackson 2006) or our recent piece on protests and revolutions (Barberà and Jackson 2016), topics that, thanks to Matt, have led me somewhat away from classical topics in Social Choice and at the same time express the vitality of this research area, broadly understood. This also applies to my interest in matching, rationing models and location models, which I include into a broad view of Social Choice as a particular way to look at resource allocation, as I was taught by my Northwestern teachers. My present projects with Antonio Nicolò (Barberà and Nicolò 2016) and with Dolors Berga and Bernardo Moreno (in preparation), on disclosure of information, are also mind openers for me. And I have also enjoyed the possibility of investigating the properties of specific voting rules in work with Ana Bogomolnaia and Hans van der Stel (Barberà et al. 1998b), with Bhaskar Dutta and Arunava Sen (Barberà et al. 2001a), and more recently with Anke Gerber (Barberà and Gerber 2017b, a) and with Danilo Coelho (Barberà and Coelho 2017).

Q. You have had responsibilities related to research policy, both in Catalonia and Spain. How were they?

A. The chance of a long life has also given me the opportunity to participate in two enriching experiences outside of my activity as a professor. One was to birth of the Catalan Institution for Research and Advanced Studies (ICREA). Andreu Mas-Colell, who was minister of Universities and Research of the Catalan government, commissioned me to create and direct a new sort of institution, whose purpose was to attract talent in all fields of research to Catalonia. ICREA was conceived as a means to facilitate the arrival of highly qualified researchers, who can work in any institution that operates in Catalonia. It was created as a foundation and given the flexibility to act by the standards of the international market for scientists. Under the supervision of highly regarded experts in different fields, we were able to give permanent contracts to people who would never otherwise thought of moving to the rigid Spanish system. The task was exciting in many ways. Creating a new institution is an inspiring task, which requires understanding the many legal, political and financial aspects of actual institutional design. Also fascinating was the opportunity to deal with scientists of all sorts, to discover the rich variety of people involved in archeological research, in photonics, in astronomy or in so many types of medical topics. ICREA provided me with a glimpse at science at large, through the contact with the experts that gave us support and advice and the people that I had the chance to interview, whether or not they ended up accepting offers. When I left, after only three years, ninety people in all fields had already agreed to join, from all parts of the world. It was a pleasure to work with an experienced politician like Andreu, who gave me freedom, resources and advice and also enjoyed the project a lot. And since it was a novel initiative, it also needed the support of the Catalan parliamentary groups, which led me, however slightly, to get in touch with a variety of political leaders. Luckily, I have come to realize that science is the least partisan of the issues, and that, at least in my times, there was a wide consensus regarding research. Unfortunately, this consensus is not so big when it comes to the organization of universities.

The ICREA experience paved the way for a surprising twist in my professional life. My former student at Bilbao and Princeton Ph.D. Maria Jesús San Segundo became Minister of Science and Education of the Spanish socialist government in 2004 and asked me to join as Secretary General of Scientific and Technological Policies. To make a long story short, I found myself in less than twenty four hours in charge of all financing of science in Spain, including all large scientific facilities, all research projects, all fellowships. Luckily, these were times of abundance, and my task was more that of stimulating the arrival of deserving projects than that of cutting budgets, as people in the same job after 2008 had to do. These were two years where I did not have one second for anything else. Research came to a halt, and other concerns took over. It was a very intense period, on a completely different job. I took a close look at many different things: negotiations in Brussels on the ITER fusion project, agreements about Spain's entry to the European Southern Observatory organization, the creation of a new Supercomputing Center and of another for the Study of the Origins of Humans, the supervision of works for the building of a synchrotron and a new telescope, the development of an innovative program to create jobs for high quality individual researchers in universities (I3) and of another program to support large research teams (CONSOLIDER). And at the same time, learning how to report to Congress, how to push for a better budget, how to relate with other ministries. Maria Jesús, my minister, was extremely supportive, I had a very good team, and I enjoyed the job. After two intense years, and when I was developing a project to create a national research agency that might circumvent many administrative problems and give more power to scientists, my Minister was sacked on grounds that had nothing to do with her research policies, and so was I. Two years later, and having seen a different world, I returned where I belonged to, and after a short adaptation period, I started research again. Luckily, I had only been away for two years, and I could go back on track.

All and all, I learned a lot about the importance of science, its rich variety, and the need for continued support to it. About how foolish it is to treat investments in science as current expenditure that oscillates with yearly budgets. I also understood how difficult it is to be a good politician and learned that I am not, although I am proud of having been, I think, a good administrator, attentive to the needs of scientists.

Q. You directed a Master in Economics of Science and Innovation at the Barcelona GSE, didn't you? How was it?

A. Very large amounts of resources are used to promote scientific and technological advances. The money comes from a rich variety of sources, both public and private, at the international, European, state and regional level. It is essential for institutions to understand how to operate in this universe, which is in a way very competitive but also very hard to grasp in all its complexity. During my period in the Ministry of Science and Education, I realized that there was a shortage of well-prepared professionals with a broad understanding of the overall system and able to advise governments, firms and research institutions.

When the Barcelona GSE started its new Master's programs, I thought that there was room for a line of teaching that would endow students from different intellectual origins with general skills in that direction and prepare them to become useful professionals, as advisors to public and private institutions, as fund-raisers or promoters of science and technology-based firms. The program ran for several years. Its graduates came from different scientific fields and acquired an economic perspective, while also being presented with an overview of the scientific, technological and institutional challenges they would face if they chose to find jobs of this sort. Most of them did. The program was eventually discontinued, partly because it was not theoretical enough. I am somewhat surprised in this respect. While very theoretical in my research, the approach I have always taken when it comes to institutional issues has been quite pragmatic and based in my on-the-job learning.

Q. How do you see the research in Social Choice now?

A. I see many interesting avenues for research in Social Choice, broadly understood. My main concern is about the negative role of fads in shaping the junior market. Social Choice is not a popular subject at this point, for reasons that I think have little to do with the intrinsic interest of the questions it addresses or the quality of its major results. This has an impact in the job market and keeps many good students away from presenting themselves with that banner. But this unfortunate initial shock will not erase the main concerns of Social Choice from the intellectual horizon of talented individuals. I am optimistic that, as all fads, this one will also be reversed. Moreover, academic lives are long, and our topics will find their way into the agendas of first class researchers whatever their starting point may have been, as it has already been the case in the past. If you look back, you will easily identify great economists, political scientists, philosophers and mathematicians who have never been perceived as social choice theorists and yet have made important contributions to this field, because at some point in their careers, they have been intellectually challenged by the exciting issues that it deals with.

Q. Social Choice is not a subject offered in our undergraduate degrees except in very few cases. What arguments would you give in favor of introducing it in the curricula? What should be the emphasis in this subject?

A. Let's not be slaves of labels, but look at contents. I think that undergraduate students can appreciate the importance of normative, axiomatic analysis if it is properly presented. But Social Choice issues are usually introduced in the middle of Micro courses, through a hasty presentation of Arrow's theorem that has then very little continuity in the rest of the course. This is a double waste, because students are usually neither ready to understand the depth of that result, nor to appreciate the importance of the axiomatic method as a tool for better thinking. In my experience, it would be better to introduce the axiomatic method through the analysis of other, may be simpler issues. For example, as a way to better understand what is the meaning of fairness, using results from cooperative game theory. Or by showing how the problems of measuring inequality, poverty or human development are better understood through axiomatic analysis.

What I miss, then, is an appropriate exposure of students to the methods of reasoning of Social Choice theory, and a positive valuation of the important issues to whose understanding it can contribute. Whether or not this is part of a special course is not so important.

And, by the way, I believe that Political Economy is a rather different discipline than Social Choice theory, in purpose and methods. Mixing one with the other is, in my opinion, not a good idea.

Q. Will Social Choice and theory in general have any prominent role in ten or twenty years from now?

A. As I said before, my impression is that the topics we study are intellectually challenging and our method of analysis is useful: Therefore, they will stick around. But of course other points of view will be added, as they have continually been added in the past: just think of the recent rise of Computational Social Choice theory.

As for theory in general, I am convinced it will have an essential role at different points in the future, with the ups and downs that has always had. Think of the enormous amounts of work that was made to study the data that became available at the turn between the nineteenth and the twentieth century. Much of it involved no theory, until the then young Econometric Society came to rescue and vindicated the right mix of evidence and theory. I believe that the excitement about the new possibilities opened by new techniques and data in the last decades will eventually bring about a renewed need for unifying theoretical ideas.

Q. In what research projects, and with whom, are you working at the moment?

A. For the moment, I am engaged in many research projects. Of course, they are driven by my intellectual curiosity, but they also respond to something I cherish more and more: the opportunity to work with colleagues who are friends, and whose personal contact I enjoy. Because of that, I am overcommitted, because I am lucky to be friends with many excellent researchers of different ages and varied interests. I have projects with Anke Gerber, Dolors Berga, Bernardo Moreno, Matt Jackson, Alejandro Neme, Antonio Nicolò, Danilo Coelho, Yves Sprumont, Walter Bossert and Kotaro Suzumura, and I am still looking forward to have new opportunities to work with you, Carmen and Jordi, as we did in the past.

Q. What other things are you doing besides research? Enjoying the grandchildren?

A. Cesca and I have been together for 52 years, and we are enjoying the luck of these first years of retirement in good health. She has always had an artistic vein and is now spending a lot of time painting. I still use much of my time to do research, but I am trying to learn a bit about writing fiction. We both enjoy the opportunities that these activities provide to enlarge our understanding of the richness of life, and its many facets, beyond the professional activities that have already given us so much. And then, of course, our two children and our three grandchildren are major players in our life projects. Being parents was a fascinating challenge; being grandparents is an opportunity to rediscover purity, beauty and the marvels of childhood.

Q. Last but not least, a tip for young generations of researchers?

A. Make friends with the colleagues that deserve your friendship: They are a treasury for the whole of your academic life. I was lucky to meet, early in my career, many of the people I mentioned before, and I found inspiration and support from them, as colleagues and friends. I understand research as a collective endeavor, and I think that cooperating with people that you like is the better part of it. Of course, competition is out there and has to be met, but it is better to enjoy our activity as a joint venture. There are many questions to share, many avenues to pursue. Science is a collective effort. Enjoy the individual excitement that discovery provides, but look further and also enjoy being part of the whole adventure of knowledge.

References

Barberà, S. (1977a). Manipulation of social decision functions. *Journal of Economic Theory*, *15*, 266–278.

Barberà, S. (1977b). The manipulation of social choice mechanisms that do not leave 'too much' to chance. *Econometrica*, *45*, 1573–1588.

Barberà, S. (1978). Nice decision schemes. In H. W. Gottinger & W. Leinfellner (Eds.), *Decision theory and social ethics: Issues in social choice* (pp. 101–117). Dordrecht: D. Reidel Publishing Company.

Barberà, S. (1979a). A note on group strategy-proof decision schemes. *Econometrica*, *47*, 637–640.

Barberà, S. (1979b). Majority and positional voting in a probabilistic framework. *The Review of Economic Studies*, *46*, 379–389.

Barberà, S. (1979c). El papel del Estado en la asignación de recursos: algunas recomendaciones desde la Teoría Económica. In J. M. Buchanan, et al. (Eds.), *El Sector Público en las Economías de Mercado* (pp. 139–161). Madrid: Espasa-Calpe.

Barberà, S. (1979d). Teoría económica y diseño de instituciones: el caso de los bienes públicos. In J. M. Buchanan, et al. (Eds.), *El Sector Público en una Economía de Mercado* (pp. 101–109). Madrid: Espasa-Calpe.

Barberà, S. (1980a). Stable voting schemes. *Journal of Economic Theory*, *23*, 267–274.

Barberà, S. (1980b). Pivotal voters: A new proof of Arrow's theorem. *Economics Letters*, *6*, 13–16.

Barberà, S. (1983). Strategy-proofness and pivotal voters: a direct proof of the Gibbard-Satterthwaite theorem. *International Economic Review*, *24*, 413–417.

Barberà, S., & Coelho, D. (2017). Balancing the power to appoint officers. *Games and Economic Behavior*, *101*, 189–203.

Barberà, S., & Dutta, B. (1982). Implementability via protective equilibria. *Journal of Mathematical Economics*, *10*, 49–65.

Barberà, S., & Gerber, A. (2003). On coalition formation: Durable coalition structures. *Mathematical Social Sciences*, *45*, 185–203.

Barberà, S., & Gerber, A. (2007). A note on the impossibility of a satisfactory concept of stability for coalition formation games. *Economic Letters*, *95*, 85–90.

Barberà, S., & Gerber, A. (2017a). A shut mouth catches no flies: Consideration of issues and voting. Mimeo.

Barberà, S., & Gerber, A. (2017b). Sequential voting and agenda manipulation. *Theoretical Economics*, *12*, 211–247.

Barberà, S., & Jackson, M. O. (1988). Maximin, leximin and the protective criterion: Characterizations and comparisons. *Journal of Economic Theory*, *46*, 34–44.

Barberà, S., & Jackson, M. O. (1995). Strategy-proof exchange. *Econometrica*, *63*, 51–87.

Barberà, S., & Jackson, M. O. (2004). Choosing how to choose: Self-stable majority rules and constitutions. *The Quarterly Journal of Economics*, *119*, 1011–1048.

Barberà, S., & Jackson, M. O. (2006). On the weights of nations: Assigning voting weights in a heterogeneous union. *The Journal of Political Economy*, *114*, 317–339.

Barberà, S., & Jackson, M. O. (2016). A model of protests, revolution, and information. WP SSRN-id2732864.

Barberà, S., & Neme, A. (2015). Ordinal relative satisficing behavior. WP SSRN-id2497947 and WP790 Barcelona GSE.

Barberà, S., & Nicolò, A. (2016). Information disclosure under strategy-proof social choice rules. Mimeo

Barberà, S., & Pattanaik, P. K. (1984). Extending an order on a set to the power set: Some remarks on Kannai and Peleg's approach. *Journal of Economic Theory*, *32*, 185–191.

Barberà, S., & Pattanaik, P. K. (1986). Falmagne and the rationalizability of stochastic choices in terms of random orderings. *Econometrica*, *54*, 707–715.

Barberà, S., & Sonnenschein, H. (1978). Preference aggregation with randomized social orderings. *Journal of Economic Theory*, *18*, 244–254.

Barberà, S., & Valenciano, F. (1983). Collective probabilistic judgements. *Econometrica*, *51*, 1033–1046.

Barberà, S., Barret, C. R., & Pattanaik, P. K. (1984). On some axioms for ranking sets of alternatives. *Journal of Economic Theory*, *33*, 301–308.

Barberà, S., Sonnenschein, H., & Zhou, L. (1991). Voting by committees. *Econometrica*, *59*, 595–609.

Barberà, S., Gul, F., & Stacchetti, E. (1993). Generalized median voter schemes and committees. *Journal of Economic Theory*, *61*, 262–289.

Barberà, S., Massó, J., & Neme, A. (1997). Voting under constraints. *Journal of Economic Theory*, *76*, 298–321.

Barberà, S., Massó, J., & Serizawa, S. (1998a). Strategy-proof voting on compact ranges. *Games and Economic Behavior*, *25*, 272–291.

Barberà, S., Bogomolnaia, A., & van der Stel, H. (1998b). Strategy-proof probabilistic rules for expected utility maximizers. *Mathematical Social Sciences*, *35*, 89–103.

Barberà, S., Massó, J., & Neme, A. (1999). Maximal domains of preferences preserving strategy-proofness for generalized median voters schemes. *Social Choice and Welfare*, *16*, 321–336.

Barberà, S., Dutta, B., & Sen, A. (2001a). Strategy-proof social choice correspondences. *Journal of Economic Theory*, *101*, 374–394.

Barberà, S., Maschler, M., & Shalev, J. (2001b). Voting for voters: A model of electoral evolution. *Games and Economic Behavior*, *37*, 40–78.

Barberà, S., Bossert, W., & Pattanaik, P. K. (2004). Ranking sets of objects. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory* (Vol. II, pp. 893–977)., Extensions Dordrecht: Kluwer Academic Publishers.

Barberà, S., Massó, J., & Neme, A. (2005). Voting by committees under constraints. *Journal of Economic Theory*, *122*, 185–205.

Barberà, S., Berga, D., & Moreno, B. (2010). Individual versus group strategy-proofness: When do they coincide? *Journal of Economic Theory*, *145*, 1648–1674.

Barberà, S., Berga, D., & Moreno, B. (2012a). Domains, ranges and strategy-proofness: The case of single-dipped preferences. *Social Choice and Welfare*, *39*, 335–352.

Barberà, S., Berga, D., & Moreno, B. (2012b). Group strategy-proof social choice functions with binary ranges and arbitrary domains: Characterization results. *International Journal of Game Theory*, *41*, 791–808.

Barberà, S., Beviá, C., & Ponsatí, C. (2015). Meritocracy, egalitarianism and the stability of majoritarian organizations. *Games and Economic Behavior*, *91*, 237–257.

Barberà, S., Berga, D., & Moreno, B. (2016). Group strategy-proofness in private good economies. *American Economic Review*, *106*, 1073–1099.

Calsamiglia, X. (1977). Decentralized resource allocation and increasing returns. *Journal of Economic Theory*, *14*, 263–283.

Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*. Cowles Foundation: New Haven, Yale University Press.

Groves, T., & Loeb, M. (1975). Incentives and public goods. *Journal of Public Economics*, *4*, 211–226.

Hurwicz, L. (1972). On informationally decentralized systems. In C. B. McGuire & R. Radner (Eds.), *Decision and organization*. North Holland: Amsterdam (chap. 4).

Mas-Colell, A., & Sonnenschein, H. (1972). General possibility theorems for group decisions. *The Review of Economic Studies*, *39*, 185–192.

Moulin, H. (1980). On strategy-proofness and single-peakedness. *Public Choice*, *35*, 437–455.

Postlewaite, A., & Roberts, J. (1976). The incentives for price-taking behavior in large exchange economies. *Econometrica*, *44*, 115–127.

Satterthwaite, M. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, *10*, 187–217.

Sen, A. (1970). *Collective Choice and Social Welfare*. Cambridge: Harvard University Press.

Silvestre, J. (1978). Increasing returns in general non-competitive analysis. *Econometrica*, *46*, 397–402.

John Broome

Richard Bradley and Marc Fleurbaey

1 Your Intellectual Journey

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

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

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

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

R. Bradley
London School of Economics and Political Science, London, UK
e-mail: r.bradley@lse.ac.uk

M. Fleurbaey (✉)
CNRS, Paris, France
e-mail: marc.fleurbaey@gmail.com

Paris School of Economics, Paris, France

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_7

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

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

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

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

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

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

A more recent, smaller discovery (which I think genuinely was mine) is that the best account of the logic of requirements is built on something called a 'neighbourhood semantics'. This is a technical matter, but it also helps a lot of things to fall into place. Another example was my realization that fairness requires, not the maximization of anything, but proportionality in the satisfaction of claims. This provides a good explanation of the fairness of lotteries, and it also gave me my account of the value of equality, which I presented in my book *Weighing Goods*.

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

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

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

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

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

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

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

When I took my MA in philosophy at Birkbeck, my thesis was about the philosophical foundations of welfare economics. During that year, I also wrote a paper about the value of human life in economics. The value of life is a topic where philosophy and economics are very tightly connected. It suited me very well, since I worked in an Economics Department but was interested in philosophy. I have pursued this question ever since. Since the economic theory of the value of life is wrapped around with risk and uncertainty, I learnt about decision theory. Since the value of life has important practical applications to medicine and public health, I found myself involved in those subjects. And then by accident, I became involved in another application of this same topic, which is the ethics of climate change.

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

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

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

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

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

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

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

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

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

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

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

2 Utilitarianism

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

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

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

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

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

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

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

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

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

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

I then realized that the theory of uncertainty provides a useful analytical tool in value theory. This was Harsanyi's discovery. I remember working on it when I was a visitor in All Souls College, Oxford, in 1982. I was fascinated that Harsanyi's Theorem could derive such a powerful conclusion from such seemingly anodyne assumptions. I found the mathematics almost magical. The additive structure of decision theory and the additive structure of the utilitarian theory of value emerge from premises that do not mention additivity. Moreover, I discovered this is true also in the Bolker–Jeffrey version of decision theory, which has quite different mathematical foundations. Additivity evidently has deep mathematical roots that I still don't really understand.

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

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

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

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

3 Bernouilli's Hypothesis and the Representation of Betterness

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

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

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

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

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

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

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

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

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

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

4 Interpersonal Addition

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

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

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

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

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

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

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

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

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

5 Personal Goodness and Interpersonal Comparisons

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

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

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

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

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

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

A: (m, m) if heads; (n, n) if tails
B: (m, n) if heads, (n, m) if tails
C: (m, n) for sure
D: (n, m) for sure
E: (m, m) for sure
F: (n, n) for sure

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

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

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

6 The Intuition of Neutrality

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

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

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

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

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

Gabrielle Demange

Karine Van der Straeten

1 Introduction and Background

This article is an edited version of a series of interviews conducted in July 2018.

Karine Van der Straeten (KVDS)*: Thank you very much Gabrielle for agreeing to have these interviews. I am extremely honored to help in this process. Having known you first as a student, and then as a colleague at the Paris School of Economics and as a co-author, I am very impressed by the breadth and the depth of your scientific contributions and by your vision of economics in general. This makes the task of interviewing you all the more challenging!*

If you agree, I would suggest structuring the discussion as follows, according to the now established tradition of these interviews. At first, we could talk about your years of training, and the reasons or intellectual influences that led you to work in social choice. Then we could focus on two or three of your major contributions. Finally, I would like to take advantage of the broad scope of your themes of expertise and interests to discuss what you think are the most promising recent developments in social choice—or at least the ones that have interested you the most—and some of the new questions to explore.

Gabrielle Demange (GD): Thank you very much Karine for spending time on these interviews. It is a great pleasure and honor to participate in these series. It is also a great opportunity for me to go back to my previous works and think about my motivations!

KVDS: *So, could you please start by telling us about what drove you into the field of social choice?*

Manuscript prepared for a book of interviews on Social Choice coedited by Marc Fleurbaey and Maurice Salles.

K. Van der Straeten (✉)
Toulouse School of Economics & CNRS, Toulouse, France
e-mail: karine.van-der-straeten@tse-fr.eu

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_8

GD: I was a master student in mathematics in Paris when I discovered game theory, strategic or cooperative, and social choice. I was very fortunate to follow a course on game theory given by Hervé Moulin, which was my very first introduction to these topics. I became so much interested that I started a PhD with him. I don't remember exactly the title of the thesis but it included 'imperfect competition' as game theory was becoming a main tool in industrial organization. Similarly, I did not see clear boundaries between social choice theory and game theory. Maybe because the main impossibility theorems bear both on aggregation of preferences and manipulability.

When I started in the early 80s, these impossibility theorems had been found and refined along various directions so part of the research expended to escape them and find positive results in more specific settings. It was the time where social choice theory somewhat created itself as a field. Maurice Salles organized a large conference in Caen with most people working in implementation, welfare, preference aggregation, fairness and so on. Maurice told me that the creation of the Society was decided there. While still working on my thesis, I attended the conference (but did not at the time realize how lucky I was). Many other young French scientists attended as well, such as, if I remember correctly, Marc Fleurbaey, Gilbert Laffond, Jean-François Laslier, Michel Le Breton, Alain Trannoy, who all later on pursued their own path in studying social choice.

The journal *Mathematical Social Choice* was also founded at that time and Bernard Monjardet proposed me to submit a work that became my first published paper (Demange 1983). This paper was a positive result, on the existence of a Condorcet winner in a domain of single-peaked preferences I called 'single-peaked preferences on a tree' (I remember Alan Kirman laughing, finding this name ridiculous!).

KVDS: *Where were you based at that time?*

GD: I was teaching mathematics in the French University (Paris VI) and doing research in the Laboratoire d'économétrie de l'Ecole Polytechnique (though no econometricians were there, but econometrics meant quantitative and formalized economics). This might seem strange outside France to be involved in two institutions but actually my environment was very good, thanks in particular to the director Claude Henry who played the role of a 'benevolent dictator', keeping us away from any administrative duties.

2 Articles

KVDS: *I would like now to come back to some of your main contributions to the field, and discuss them in some detail. Given the wide variety of topics you have worked on, it is very difficult to select a couple of representative or most significant contributions! During preliminary exchanges we had to prepare these interviews, you mentioned one paper on fair allocations, and one paper on stable coalitions, which we could discuss.*

Going in chronological order, maybe we can start with the article on fair allocations, co-*authored with Ahmet Alkan and David Gale and published in Econometrica* *in 1991 (Alkan et al. 1991). For those who are not familiar with this article, could you please synthesize what seems to you to be the main contribution of this work?*

GD: We study the problem of distributing a set of 'objects' together with an amount of money, positive or negative, to a group of individuals. Each individual receives at most one object. For example, the objects are rooms to be assigned to students. The question here is to assign the rooms and the money in a manner that is fair, which means both (Pareto) efficient and envy-free, that is, everyone likes his/her own allocation (room together with the rent) at least as well as that of anyone else. When rooms are of different qualities (size, sunshine, room with a view…) and students' preferences differ, fairness requires to account for these differences both in the assignment and the rents' levels. The objects can be undesirable, for example, the members of an academic department must perform administrative tasks and the department has a fixed administrative budget to compensate them. When tasks and preferences differ, fairness can be achieved only by differentiating the compensations (and the budget is large enough!). We derive simple necessary and sufficient conditions for the existence of fair allocations. The notable feature here is that the existence proof is constructive, which allows us to derive qualitative properties of fair allocations. In particular, we show that if there are at least as many people as objects and a fair allocation exists, then if the amount of money is increased, there is a new fair allocation which makes everyone strictly better off.

KVDS: *How did you start working on these issues? What was your motivation?*

GD: I was fascinated by the 'divide and choose' method and the cake-cutting problem. It illustrates the many different aspects of fairness problems: Are there 'fair' shares? How to achieve them? You have two individuals, say you Karine and me, and a cake to split between us. The cake has strawberries, cream, almonds and so on. Fairness encompasses two properties: no-envy and Pareto efficiency. If we each get an identical share, with exactly the same composition, no one prefers the share of the other one: the split is envy-free. But there is little chance that it is Pareto efficient, say because you don't like strawberries and I don't like cream. Reaching simultaneously a Pareto efficient and envy-free split raises non-trivial questions: Is it possible? How will the divider cut the cake if she knows the other's preferences? And if she does not know them and is risk-averse?

The divide and choose method works for two individuals. My first work on this type of subject was inspired by a kind of extension of the method introduced by Crawford (1979). The procedure works with more than two people, starting with an auction in which agent bids for being a proposer. It implements allocations that are 'egalitarian-equivalent'. Egalitarian-equivalence is a notion of fairness that is less universal than envy-free as it depends on a reference bundle. Though, it is easier to handle and solves existence issues.

A simple and very elegant procedure to reach a fair allocation works in an exchange economy: Distribute the endowments equally between the agents and reach an equilibrium. But this procedure does not always work, for example because you cannot divide and distribute equally the goods. This is the case in the fancy cake-cutting problem if you must split the cake into slices. This is also the case in more serious problems if individuals differ in their abilities—their ability is like a non-transferable endowment—or if the goods are indivisible, as in the paper. This is why fair allocations may not exist when goods are indivisible; in our paper we relied on a constructive method to find them.

KVDS: *Could you tell us more about your two co-authors on this paper?*

GD: David (Gale) was a well-known mathematician interested in game theory and economics. He spent a sabbatical year in Paris when I started my thesis. I was working on assignment games (Shapley and Shubik 1971) and after a while we started to collaborate together. It was a great opportunity for me and the fairness paper is our fourth joint paper. From a technical point of view, all our joint papers deal with the same type of problems, in which there are indivisible goods in the same category, say houses, tasks, and each individual is interested in having at most one of these goods. It includes situations without transfers (marriage or matching models) or with transfers (assignments of workers to tasks). Though the setting might seem restrictive, there are many applications and extensions currently developed under what is called market design.

Ahmet (Alkan) was a former student of David at Berkeley and came to Paris to visit David. This is how we met. He had also worked on assignment and matching games and this is how we end up working together the three of us.

KVDS: *In the paper, you insist on the "comparative statics" part (more money, more objects). What exactly is your conception of this comparative static exercise? You mention in Section 1C that in this literature, it is common to have results such as "more money or more things can make some individuals worse off". What was the vision/conception of these "paradoxical" results? And how did you interpret your "more positive" results, which you describe as "surprising", i.e. the results that under certain conditions, we can have good properties of monotonicity?*

First, to place these results in context, there were works on the 'transfer paradox': At a competitive equilibrium, an agent may end up worse off if some agent gives him part of her endowment. The interpretation of the transfer paradox in international economics has of course important consequences: a country may be worse off by accepting a gift from another country. This paradoxical result is due to price changes and their effects on the endowments' value. In the context we consider, such effects are not present as there is no ownership. We simply ask: if there is more money, can we make everybody better off at a fair allocation? Our positive result has to be contrasted with the impossibility result of Moulin and Thomson (1988) in the context of divisible goods. This is why we call the result 'surprising'.

KVDS: *What, in your opinion, have been the most important subsequent works that have taken up these results/this framework, in social choice or in other fields?*

GD: As I said previously, in many contexts, fair allocations may be impossible to reach simply because they do not exist. A question then is to find allocations that are approximately envy-free (or satisfies any other equity concept), or that 'minimize' envy. Another issue is to design procedures or algorithms that work well in dynamic contexts. Actually, I think that the closest works to mine are conducted by researchers in computer science. Interestingly, some tools that have been developed on Internet, such as the Spliddit Website (http://www.spliddit.org/apps/goods) in which you enter some data and it computes fair division of goods, credits, or tasks. One of the algorithm implements the allocation introduced in our paper.

KVDS: *If you agree, let us now move on to the other paper, on group stability in hierarchies and networks. This very interesting paper was published in the Journal of Political Economy in 2004 (Demange 2004).*

It seems to me to be another example of an attempt at getting "positive results" (stability and efficiency), in a literature when the general problem usually gets more negative results. Would you agree with this statement? How would you summarize the main contribution of this paper?

GD: Yes indeed, it is again a result that one can qualify as positive. It has moreover some links with the possibility of preference aggregation, as I will explain in a little while.

The main contribution is to show that under a specific power structure linking individuals and subgroups/coalitions—a structure that I will interpret as associated with a 'hierarchy'—there always exists a stable outcome in the sense that no eligible (or effective) subgroup in this hierarchy has any interest in 'blocking' it. The notion of stability here is that of the core for a cooperative game. One difficulty with this notion is that there are some non-pathological situations where a stable outcome does not exist (as characterized by Shapley or Bondareva for transferable utility games). Very schematically, to ensure stability, intermediate coalitions should not have too large incentives to block.

I postulate that in a hierarchy, only some coalitions defined by the hierarchical structure can block. In particular, people at the same level cannot collaborate directly without a common superior. The simplest hierarchy is that of the 'principal-agent' model, in which the isolated agents have some power, but obviously minimal in that any other coalition with power must contain the principal. The power structure that I consider in a hierarchy can be seen as a generalization of this 'principal-agent' model with several levels. What is important is that two individuals can be in the same coalition only if their closest common supervisor also belongs to this coalition as well as all the intermediate between them (in formal terms, this means that the coalition is connected in the graph describing the hierarchy).

Provided there are no spill-over between coalitions, the stability result holds whatever the problem faced by the hierarchy (allocation of costs, provision of computing facilities, etc.), and whether it is super-additive or not. When the problem is super-additive, the whole society has an interest in coordinating itself on a single decision; whereas when the problem is not super-additive, the society may be more efficient by splitting into several independent subgroups, each taking decisions that apply to its members. One of the paper's interests is also to propose a very intuitive algorithm to calculate a stable allocation. Starting from the bottom of the hierarchy, one gives each individual his/her incremental contribution to all his/her successors.

KVDS: *You mentioned earlier the similarities with the problem of preference aggregation. Could you please make the link more explicit?*

In problems of preference aggregation, two broad lines can be followed to obtain 'positive' results. Let me describe them with the example of the majority voting game. A majority winner, often called a 'Condorcet winner', exists only if no coalition with at least half of the voters 'blocks' it, i.e. unanimously prefers another candidate. The majority winner(s) thus coincide with the core of the majority voting game. As we know since Condorcet, often there is no majority winner. Two different approaches have addressed this majority instability. One approach is to restrict eligible preferences, as when one assumes single-peaked preferences or intermediate preferences. Another approach is to restrict the blocking power of some coalitions. A first paper along this line was Nakamura's paper and the fascinating 'Nakamura number' (Nakamura 1979). Nakamura considers super-majority rules where an effective coalition has more than a given fraction of the voters, say 60%. When the number of alternatives is smaller than the Nakamura number, the core is non-empty *whatever* the preferences. When the number of alternatives is larger, cycles may occur and the core may be empty. [To illustrate, for standard majority, the Nakamura number is 2, with the possibility of a Condorcet cycle as soon as we have three alternatives. (I am simplifying a little bit as the Nakamura number depends on the number of voters Nakamura (1979).)]

The result on hierarchies is of this type: the core is non-empty *whatever* the preferences.

KVDS: *How did this paper fit in your research agenda at that time?*

GD: I had always been interested in the stability of decisions in a collective context (and still am!), in particular the stability in relation to the blocking power of coalitions. I had already some works published in the *Journal of Economic Theory* (with Dominique Henriet, (Demange and Henriet 1991)) and the *Journal of Mathematical Economics (Demange* 1994*)*, which exhibited a context where a society can agree on a stable outcome—in general a partition of the society in independent sub-groups. I had in mind the modeling of two forces acting in opposite directions on the formation of coalitions and their size: the larger a coalition, the greater its power, but also the more numerous the causes of disagreement due to the dispersion of preferences. As an illustration, you can think of the (endogenous) splitting of the society into communities, each choosing a public good and a tax level to finance it (as first considered by Guesnerie and Oddou). I showed that under certain conditions on preferences—I called intermediate preferences on a tree—there was always a stable outcome. Under these preferences, an unstable outcome is surely blocked by a connected coalition in the tree, so one only needs to consider blocking by connected coalitions, and this ensures stability. The *Journal of Political Economy* article does not make such assumptions about preferences and directly considers which coalitions can or cannot be formed.

KVDS: *If I understood the paper correctly, in a hierarchy, some blocking coalitions are exogenously ruled out. Yet, individuals have agree about this hierarchy at some point. So it seems like it pushes the problem one step further. Is this a question you have been interested in?*

GD: Yes indeed, I take the hierarchical structures as given and I study their properties, regardless of the number of levels, the number of direct subordinates, etc. I do not say anything about the acceptability of the structure. My main argument is that it is a structure that facilitates decision-making for a multitude of problems, not for a particular problem, and this may explain its prevalence in many areas.

KVDS: *In the paper, you also compare stability in hierarchies and stability in networks, with the former being much more stable. In particular, you explain that the restricted conditions for stability in networks (Proposition 1 in the paper) might seem at odd with the large diversity of networks we actually observe in reality. It seems to me that in the article, you propose to explain this fact by the fact that most existing networks share information, rather than take actions. Could you please elaborate a bit more on the fundamental difference between information sharing and action taking? If sharing information or not can impact the kind of decision a group can make, there obviously seems to be some similarities.*

GD: There are so many different contexts where information plays a role that it is difficult to answer this question in any general way. In a financial market, for example, the sharing of privileged information does not make much sense and does not lead to an analysis in terms of coalitions. In other contexts where agents benefit from some coordination, it is true that the sharing of private information that allows for a better coordination may have similarities with my framework. But I don't know precise works exploiting such similarities.

KVDS: *Could you please tell us how the paper was perceived as fitting in the literature at the time of its publication? Who were the people the most interested in these results at the time? And now?*

GD: I think that the paper has been quite well received especially from researchers working in social choice and cooperative games, and economists who started working on networks. Some researchers have later on defined and provided axiomatizations for allocations built on hierarchical outcomes (such as the 'average-tree' allocations). There are also connections with the 'sharing the river problem', which analyzes how much water cities along a river consume and how much possibly they pay for it. Starting from the source, the river and its tributaries are represented by a graph similar to a hierarchical structure. Furthermore it is natural to assume that only connected coalitions can form. The problem may be more complex due to externalities across coalitions (I did not know it at that time, but this 'sharing the river problem' had already been considered before my paper, although not from a coalitional point of view.)

Though the principal-agent model is a particular hierarchy (as I discussed above), the paper did not have any echoes with the principal-agent literature in industrial organization. This literature is mainly concerned with problems of information and moral hazard. Some works have extended the principal-agent model by incorporating an additional intermediate level, with a focus on the delegation-centralization trade-off induced by better local information but different objectives. I do not address information problems. Now the paper is much more cited by works in computer science.

KVDS: *This is very interesting! Could you please elaborate a bit more on these interactions with computer science?*

GD: Computer scientists are very active in the domain, and actually have now close connections with social scientists and game theorists, through joint conferences and participation in editorial boards. The types of questions some computer scientists study closely relate to social choice.

For example, the following questions are very close to their core competence: How 'hard' is it to find a stable solution (in different contexts, cooperative or not)? How complex is it to find a winner in an election? How often is it possible to manipulate a voting rule? These questions make sense when the data (agents, actions) is large, say when there are many voters. Some works are directly related to the papers cited above: some study the complexity of finding a fair allocation; some propose algorithms to determine whether a set of coalitions is associated to a hierarchical-tree structure and algorithms to determine whether a set of preferences satisfy the 'single-peakedness on a tree' condition or the 'intermediateness' condition; some others compute how many profiles satisfy these conditions….

Computer scientists also use their access to data and investigate new problems raised by new technologies, such as the so-called 'ride-sharing problem', where a set of commuters arrange one-time rides at short notice. A group of researchers take a cooperative approach by considering the social network of the commuters, assuming they can form coalitions between connected agents. The issue then is to define stability, to determine the rides and the payments. Interestingly, the theory can be confronted to data.

3 Recent Development and Future Avenues for Research

KVDS: *When discussing the two papers on fairness and hierarchies, you mentioned the very interesting developments in computer science that followed. Do you think, more generally, that computer science has some important contributions to make to social choice?*

GD: Yes, I think that one of the most active communities today in social choice comes from computer science. Actually, computer scientists are active in the broader field of economic design, including social choice and mechanism design (See for example Brandt et al. (2016)). In a forthcoming volume entitled '*The Future of Economic Design*', researchers share their views on the future of the field (Laslier et al. forthcoming). I don't have the exact numbers, but a large number of the contributions are from computer scientists.

I see at least two main reasons why this should be the case, and why more fruitful interactions should be expected.

First, recent progress in computing facilities allows some new mechanisms and collective choice procedures to be designed and implemented.

Second, at a time when the increased reliance on complex 'algorithms' in many public and social areas raises some defiance in (some parts of) the public opinion, standard reasoning in social choice may provide ways to address these concerns. In particular, the "axiomatization approach" could help compare various algorithms in a clear and understandable way.

KVDS: *You say that recent progress in computing facilities allow new mechanisms and aggregation procedures to be implemented in practice. Could you give us some specific examples?*

GD: In a paper in the volume just cited, I discuss in some detail how, thanks to new computing facilities, new voting procedures can be designed. More specifically, I describe two recent promising set of experiments allowing for a fairer expression of voters' preferences.

The first set of experiments you know quite well Karine, since you were involved in some of them. In the past fifteen years, several experiments have been conducted in France and elsewhere to test new voting procedures. Taking advantage of Internet, it has become very easy to test on very large sample of voters various voting procedures, such as Approval Voting, Borda rule, Plurality, Single transferable vote, Majority judgment, and to compare their results. These tests are encouraging, as they show an interest from the general public and the media. I hope that more are to come.

The second set of experiments, conducted in Switzerland, led to an actual electoral reform. This experimentation started from some severe dissatisfaction with the electoral system used in Switzerland until 2004 for cantonal elections, which was deemed 'unfair'. During a trial period starting in 2004, a new method—called the New Apportionment Procedure (NAP)—was used in the Zurich canton to allocate seats to parties and districts. This new method is based on bi-divisor methods, which I introduced with Michel Balinski in the late 80 s. At the time I wrote these theoretical articles, I doubted that the methods could be used for real political elections, mainly because they are very difficult to compute by hand. The mathematician Friedrich Pukelsheim made a tremendous job to get this procedure implemented. This new method is now definitely adopted in the Zurich canton as well as in some other cantons in Switzerland.

The unfairness of the previous system in the Swiss cantons appears in many other elections when representatives are elected in areas of very different sizes, such as the countries in EU. I am not advocating for a bi-divisor method in the EU, but social choice theorists could be more involved in the design of electoral systems.

KVDS: *Coming back to the second reason you mentioned when highlighting cross-fertilization between social choice theory and computer science, could you please now explain why you think that the "axiomatization approach", very standard in social choice theory, can help shed some light on the desirable properties of algorithms?*

GD: On a theoretical level, there are close connections between the tools developed to rank Webpages (such as *PageRank* of Google) and aggregation methods. Computer scientists have recognized that and some works use the axiomatization method to characterize ranking or recommendation systems.

On a more practical level, algorithms often solve a social choice problem. They use data that are voluntarily provided by the citizens or extracted by the machines, say the search engines. When the designer is a governmental agency, there is a legitimate demand for explanation of the mechanism/algorithm and the axiomatization approach might be helpful in that.

In a forthcoming paper (to be published in an edited volume in the honor of Leonid Hurwicz, Demange forthcoming), I illustrate this point with a recent example in France. The French high education system is mostly public so an admission system has to deal with a high number of candidates. A procedure called APB ('Admission

Post Bac') was put in place in 2009 for assigning students at their entrance to the French universities. It was based on the centralized deferred-acceptance algorithm introduced by Gale and Shapley (1962), which, as you know, has many good properties. Though, APB turned out to be a failure, resulting in its replacement in 2018. I don't want to enter into details but one of the reasons is that APB modified the deferred-acceptance algorithm in an important way. To cope with the required 'no selection' principle, according to which any student with the 'Baccalauréat' is entitled to a seat in any field, no priority was set for the universities. When the number of applicants to some slots largely exceeded the number of seats, students were allocated at random to satisfy the no selection principle. The result was that some students lacking the background for succeeding in a field and almost certain to fail got a seat while some others, much better qualified, did not. The absurdity of the system led to its rejection in 2017.

The public blamed the 'non-human' aspect of the procedure, because it was implemented by an algorithm and used random draws. The result of APB's failure is a clear defiance towards 'algorithms' from the French population. In my view, the failure was due to the absence of consistency and transparency in the policy, not in the way it is computed. Taking the viewpoint of social choice theory would have been beneficial: explain the desirable properties the government wants to achieve and make explicit the constraints. It would have made clear that the joint effect of the no selection principle and the space constraints implied random draws. But this was not politically admissible.

KVDS: *Thank you these very interesting thoughts.*

To conclude these interviews, are there any other promising research directions you would like to mention?

GD: Well, it depends what defines the field. The description of the topics for the journal Social Choice and Welfare is now very broad, including many aspects of economic design and even accepting empirical works. There are many promising researches in these directions, such as the design of voting procedures I mentioned previously. One might think that the very distinctive features that characterized Social Choice Theory at its beginning are lost: the aim of addressing 'deep' almost philosophical issues related to democracy, fairness, through very elegant and parsimonious models, rigorous and robust methods with a minimum of assumptions (no Bayesian priors, no specific games), the definition of general principles and properties .... But, the future and new technologies such as artificial intelligence might raise such type of issues worth studying by social choice researchers.

KVDS: *You say that the distinctive features that characterized Social Choice Theory at its beginning are lost. How would you explain this transformation in the field?*

For example, in your opinion, is it because the main theoretical results have already been discovered? Or do you think that such general abstract questions could/should still be explored, but that the general trend we see today in Economics—with theory becoming less and less 'fashionable' and taught at the undergrad and master levels—is the main explanation?

How do you personally feel about this transformation?

GD: No, no, it was not general remark about the current place of empirics in economics. Though, true, I think that empirical works are overly represented at the moment. But it might have been true for theoretical studies thirty years ago! As you said there are fashions in research. What bothers me more is that basics in social choice and welfare (or general equilibrium) are almost absent in many master programs.

What I had in mind is that a large part of current research in social choice aims at obtaining possibility results, and we know that they can be obtained by restricting the framework or by weakening the requirements. The scope is necessarily more limited and more applied. There is no negative judgment in that (my own works were of this type!). These works integrate tools and approaches from different fields such as social choice, game theory, computer science. It might lead to the creation of a new field (economic design?) with its own 'community'.

References

Alkan, A., Demange, G., & Gale, D. (1991). Fair allocation of indivisible goods and money and criteria of justice. *Econometrica, 59,* 1023–1040.

Brandt, F., Conitzer, V., Endriss, U., Lang, J., & Procaccia, A. D. (Eds.). (2016). *Handbook of computational social choice.* Cambridge: Cambridge University Press.

Crawford, V. P. (1979). A procedure for generating Pareto-efficient egalitarian-equivalent allocations. *Econometrica*, 49–60.

Demange, G. (1983). Single-peaked orders on a tree. *Mathematical Social Sciences, 3*(3), 389–396.

Demange, G. (1984). Implementing efficient egalitarian equivalent allocations. *Econometrica,* 1167–1177.

Demange, G. (1994). Intermediate preferences and stable coalition structures. *Journal of Mathematical Economics*, 45–58.

Demange, G. (2004). On group stability in hierarchies and networks. *Journal of Political Economy, 112,* 754–778.

Demange, G. (forthcoming). Mechanisms in a digitalized world. In W. Trockel (Ed.), *Social design: Essays in memory of Leonid Hurwicz.* Cham, Switzerland: Springer.

Demange, G., & Henriet, D. (1991). Sustainable oligopolies. *Journal of Economic Theory, 54,* 417–428.

Gale, D., & Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly, 69*(1), 9–15.

Laslier, J.-F., Moulin, H., Sanver, R., & Zwicker, W. S. (Eds.). (forthcoming). *The future of economic design. Studies in economic design.* Cham, Switzerland: Springer.

Moulin, H., & Thomson, W. (1988). Can everyone benefit from growth? Two difficulties. *Journal of Mathematical Economics, 17*(4), 339–345.

Nakamura, K. (1979). The vetoers in a simple game with ordinal preferences. *International Journal of Game Theory, 8,* 55–61.

Shapley, L. S., & Shubik, M. (1971). The assignment game I: The core. *International Journal of Game Theory, 1*(1), 111–130.

David Donaldson

Nick Baigent and Walter Bossert

NB & WB: *You began your academic career in mathematics as an undergraduate student at the University of Toronto. What are the reasons that motivated you to move to economics?*

DD: I went to graduate school in economics to escape from mathematics. One day, when I was an undergraduate, I had tea with the head of the mathematics department and I said I didn't like mathematics because it felt like puzzle-solving to me. He responded very seriously, "Then you don't belong in a mathematics department." I was completely offended at first, but he was right: if you don't have the right kind of mind, you simply won't enjoy doing it. So I got into economics. I was not a good math student and, although I performed a little better in graduate school, I was definitely not a star.

NB & WB: *You escaped to California and did your Ph.D. at Stanford. Was it a surprise for you to find out that mathematics played a major role in much of your work as an economist?*

DD: When I went into economics at Stanford in 1960, mathematics was just beginning to take over in economic theory. So a mathematics background was a help. And I didn't mind applied mathematics: I knew what it was for. I was there when Kenneth Arrow gave several excellent topics courses, based on his current research. I foolishly passed up a course in general equilibrium theory (I was a shy first-year student), but

N. Baigent (✉)
London School of Economics, London, UK
e-mail: nickbaigent@waitrose.com

W. Bossert
CIREQ, Montreal, Canada
e-mail: walter.bossert@videotron.ca

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_9

I did take his course on uncertainty. Arrow presented his papers while he was writing them.[1] He'd assign a student to take notes in each class and they were distributed to all of us. Arrow was just superb in that course. And, in his Public Finance course, he gave us a proof of his impossibility theorem on a foot-and-a-half square on the blackboard.[2] He later on became my thesis supervisor and was extremely helpful. It was also a great pleasure to talk to him because he is an exceptionally encouraging, nice person.

NB & WB: *Are there any early academic influences in addition to Arrow?*

DD: Hirofumi Uzawa was one of my favorites at Stanford. He was a Marxist and he later became a Marxist *guru* after his return to Japan. I took a course from Paul Sweezy when he visited in 1960–1961. And I was influenced by almost everyone in the department.

NB & WB: *Did you take any courses from Uzawa?*

DD: I didn't, but he hired me along with several other students for a summer project; we each got $1000 for our work. On Friday afternoons, he'd come and get us from our offices and invite us over to his place. His wife ended up having to prepare a meal without much warning, then having her own dinner in the kitchen, which made me uncomfortable, but I suppose that's what happened in many Japanese families at the time. He told us all kinds of stories, about what it was like to be a communist in Japan and so on. He'd get us graduate students drunk and the problem was we all had to get back home, which in my case meant riding a motorcycle. But no one ever got injured. There was a rumor he went back to Japan because his children were discriminated against when they lived in Chicago but I don't know the background of this. He was a very interesting person and I really liked him. There was a time when I thought I was going to visit Japan and, if that had worked out, I would definitely have looked him up.

NB & WB: *He passed away just recently; he must have been well into his eighties and he certainly was quite a character. After his return to Japan, he spearheaded an initiative to get the University of Tokyo to provide beer vending machines on campus. Sadly, he did not succeed. But Hirofumi Uzawa was not the one who got you into social choice, was he?*

DD: No, that came much later when I read Amartya Sen's book.[3] It was inspiring and it made me think that I could do something in philosophy using mathematics. But I didn't write anything in social choice until I met Charlie Blackorby.

NB & WB: *But there is more to your attraction to social choice theory than that, isn't there? You are very interested in theories of justice and how we can think about improving the social good—and social choice theory seems to be a natural field for this.*

---

[1] Arrow (1970).

[2] Arrow (1963/1963).

[3] Sen (1970).

DD: Yes, I like normative arguments a lot and I always wondered what good ones there were. Social choice theory can accommodate normative arguments as long as the underlying values are explicit and defensible. And a natural focus is the social good.

NB & WB: *The normative part of economics clearly seems to be present in much—if not all—of your work, including what you did in public economics, cost-benefit analysis, inequality and poverty, consumer theory and population ethics, for example.*

DD: Yes, that's true. I enjoyed working on normative aspects of welfare economics using duality theory, which I learned from Charlie Blackorby, to shed some light on the main issues of welfare economics and social choice theory. I used to find ordinary price theory tedious until I understood duality.

NB & WB: *Are there any other key events or encounters, including those before your doctoral studies, that with hindsight seem to have influenced you becoming a social choice theorist and welfare economist?*

DD: There wasn't much before graduate school. I do, however, recall that there were a few people saying rather silly things about utilitarianism when I was an undergraduate student—and I haven't forgotten that. And I was a lapsed Christian (United Church of Canada) and belonged to the atheists and agnostics group in the Student Christian Movement. That background, reinforced by my mother's first-rate moral intuitions, certainly pushed me toward moral questions. After Charlie Blackorby arrived at UBC, we took a chance on social choice theory and got into it together.

NB & WB: *Are there any other areas of economics that you might have specialized in, had you not chosen social choice theory and welfare economics? There is at least one very nice IO paper that Charles Blackorby presented at some point.*

DD: That probably was a spin-off[4] of the first population paper I wrote with Charlie[5]; it used similar techniques. I also played around with growth models at some point. But I never liked them all that much. Uzawa hired me to do some two-sector growth models in Marxist economics that involved capital accumulation and all that. I don't think I would have gotten into that area by myself.

NB & WB: *Had you not become an economist, what might you have become?*

DD: My father wanted me to become an engineer; he was an engineer himself. He didn't have a university degree and did it the hard way—through the professional engineering system. But I rebelled; I didn't want any part of engineering. I started off in mathematics, physics and chemistry and, after the first year, I specialized in mathematics. My father wasn't happy at first but in the end accepted it. I was quite good in mathematics in high school and the first two years of university but if I had really thought about what I wanted to do, I would have gone into philosophy. I don't know how that would have worked out.

[4]Blackorby et al. (1982).

[5]Blackorby and Donaldson (1984).

NB & WB: *After you came to the University of British Columbia (UBC) in 1964, you also did some work with Chris Archibald. What topics were these early articles on?*

DD: That was work in welfare economics. We looked at issues such as (non-)paternalism and the fundamental theorems of welfare economics. There was also some work on the relationship between paternalism and price formation, and on economic equality.[6] Chris certainly encouraged me when I was young. I was not exactly a fast learner in those days and I got a really slow start in the business. Chris helped me get over those early difficulties.

NB & WB: *Charles Blackorby must have been very influential, too. Once you started working with him, the two of you kept going for decades. He turned out to be your main coauthor over the years.*

DD: Yes, we really hit it off in the seventies after he arrived at UBC. Our first paper was a response to a paper by Brian Barry and Douglas Rae.[7] We got into quasi-orderings at that time and it was great fun; I really enjoyed it. We work well together and have a good time doing it.

NB & WB: *Quasi-orderings are also important for your later work on inequality measurement—after all, the Lorenz dominance criterion is usually expressed in terms of a quasi-ordering.*[8] *You also used duality in your work on inequality. We are thinking of two of your joint papers with Charles Blackorby on inequality that used tools from duality theory such as distance or transformation functions in the context of relative and absolute inequality measures.*[9] *This was very influential work; using duality in this area is something that hadn't been done before in this literature.*

DD: We thought of a different way of doing something similar to some earlier work by others in this area. That is, we wanted to get a function with certain properties (such as being a relative or an absolute index) without having a social evaluation function that gives them to you automatically. Distance functions and other peculiar dual representations turned out to do the trick.

NB & WB: *These were just the beginnings of your collaboration with Charles. There was much more to follow—among other topics, your work in population ethics.*

DD: The way we got into population ethics was interesting. I spent a sabbatical on Salt Spring Island and Charlie visited every other weekend so we could work together. One time he showed up with two articles by Kevin Roberts that essentially scooped everything we had worked on over the previous year.[10] So we thought we

---

[6] Archibald and Donaldson (1976a, b, 1979).

[7] Barry and Rae (1975), Blackorby and Donaldson (1977).

[8] Lorenz (1905).

[9] Blackorby and Donaldson (1978, 1980).

[10] Roberts (1980a, b).

should do something no one ever works on. I said I had always wanted to know something about population issues and why don't we do that? So we did and took it from there. The first paper that came out of this was our article in the *Journal of Public Economics.*[11]

NB & WB: *And you kept the project going over many decades.*

DD: Yes we did. Later on, Walter Bossert joined us and persuaded us to investigate the population problem in an intertemporal setting. We opted for a period analysis because that was what we knew. The three of us wrote quite a few articles, ending with a book that covered pretty much everything we had done (some with, some without Walter) over the years.[12] The book doesn't seem to sell very much, though, and I kept thinking it was a mistake to write it—no one would want to read it and no one was interested. And that's almost true.

NB & WB: *But it turned out that there is quite some interest, also among philosophers. The book gets cited more and more and it seems that many people are sufficiently influenced by it so that they have started working on the topic themselves. For instance, you were invited to give a talk on population issues at Oxford, weren't you?*

DD: Yes, I gave a paper I wrote with Krishna Pendakur[13]; it's related to some earlier work I did with him.[14] The paper deals with the application of population principles. Happily, the number of people interested in population ethics in the context of climate change appears to be increasing.

NB & WB: *In your work with Charles Blackorby on population ethics, you introduced an alternative to classical utilitarianism and average utilitarianism that is based on assigning different utility numbers to neutrality and to a critical level.*

DD: Our proposal is what we call critical level utilitarianism (CLU) with a fixed critical level above neutrality. A life at neutrality is a life that, from the viewpoint of the person living it, is as good as a life without any experiences. Once we normalize the utility level corresponding to neutrality to some number—zero, say—a life is worth living if the lifetime utility associated with this life is positive. Using this zero-normalization for neutrality, we choose a fixed positive critical level α which is the utility value that represents the value of a life at this critical level.[15] Thus, the value function for CLU is very simple: for any utility vector $u = (u_1, \cdots, u_n)$ involving n people, we calculate the sum of the differences between the individual

[11] Blackorby and Donaldson (1984).

[12] Blackorby et al. (2005).

[13] Donaldson and Pendakur (2014).

[14] Donaldson and Pendakur (2004, 2006).

[15] If utilities are individually cardinally measurable and interpersonally comparable at two utility levels, neutrality and a level above neutrality, for example, and if, for each individual, a utility function is chosen which is zero at neutrality and equal to an arbitrary positive number at the other level, numerically measurable, fully interpersonally comparable utilities result.

utilities u_i and α to get the value function $W(u) = \sum_{i=1}^{n}[u_i - \alpha]$ (or any increasing transformation of this expression). To rank any two utility vectors u and v (not necessarily with the same population or the same population size), we just have to compare the corresponding values of W.

CLU has some nice features. To begin with, the interpretation of the critical level is very transparent: if we add a person at this critical level to a utility-unaffected population, social goodness is unchanged. Because the critical level is positive, this criterion is different from classical (or total) utilitarianism, where the critical level is equal to zero—the level that represents neutrality. A positive critical level puts a floor on the trade-off between average utility and population numbers. This allows us to resolve one of the major criticisms Derek Parfit directed toward classical utilitarianism. He accuses the classical principle of leading to the repugnant conclusion. A population principle implies the repugnant conclusion if every alternative in which each person experiences a utility level above neutrality is ranked as worse than some alternative in which each member of a larger population has a utility level that is above neutrality but arbitrarily close to it.[16] This conclusion is not implied by CLU with a positive critical level, and the difficulty is avoided in a clear and intuitively appealing way.

In addition to avoiding the repugnant conclusion, CLU and its generalized additively separable extension allow an evaluator to ignore the existence and well-being of individuals who are long dead or exist in the far future, a property we call existence independence.[17] Parfit also discusses principles that have a 'range' of critical levels.

NB & WB: *You did something like that yourself—the critical-band CLU principles that generate quasi-orderings.*[18] *Instead of a single critical level, these principles consider an interval (or 'band') of critical levels and make statements regarding social goodness only if the requisite pairwise comparisons are the same for all critical levels in the interval.*

DD: Yes, we did this at some point but I have changed my mind about it over the years. I now think the quasi-ordering approach is too clumsy. The reason is that, to use it to select alternatives from a feasible set, standard rationality recommends looking for undominated alternatives, those feasible alternatives without feasible ones that are better. But, in the critical-band case, although the best alternatives for each critical level in the band are undominated, there are undominated alternatives that are not best for any critical level in the interval. If the interval represents uncertainty about where the critical level should be, there is a reason to focus on the critical level principles that correspond to critical levels in the interval. An alternative to the undominated-alternative approach is, therefore, to find the best alternatives for each critical level in the band and recommend that choices be made from that set.

[16] Parfit (1976, 1982, 1984, Chap. 19), Blackorby et al. (2005, pp. 4–5).

[17] Blackorby et al. (2005, p. 159).

[18] Blackorby et al. (1996).

NB & WB: *You don't think that always requiring the social ranking to be an ordering is too demanding? John Broome, for example, doesn't; he claims that goodness is, by definition, a quantity and therefore transitive.*[19]

DD: Broome presents an argument like that in his book *Weighing Lives*,[20] I do think social betterness relations are transitive. But I'm agnostic about allowing quasi-transitivity (transitivity of social betterness) rather than full transitivity. And, in the context of individual well-being, the claim that one alternative is better than another for an individual can only mean the person's well-being is greater in the first. If true, individual no-worse-than relations must be fully transitive. But I also think there are a lot of ways of complicating things that are not very useful. I am persuaded by Broome's discussion of the arguments (against transitivity of betterness) by Rachels and Temkin.[21]

NB & WB: *There have been contributions that criticized CLU, especially in philosophy. For example, the sadistic conclusion has been used as an argument against CLU with a positive critical level.*

DD: Yes, the sadistic conclusion is Gustaf Arrhenius's idea. A principle implies the sadistic conclusion if and only if, when adding people to a utility-unaffected population, it can be better to add people with negative utilities rather than a larger number with positive utilities.[22] Arrhenius considers this an unattractive property, but if you add a requirement like that to a set of reasonable axioms, you get an impossibility result—and there are many impossibility results in this area. Principles can avoid the repugnant and sadistic conclusions but, if they do, they necessarily do not satisfy existence independence. So one has to make a choice. I think Parfit was right about the repugnant conclusion, and I think existence independence must be satisfied because it is necessary for piecemeal social evaluation.

Suppose we suddenly discover there has been a race of humans living on some far distant planet for the last million years but never had anything to do with them, and never will. And suppose we discover their existence. In that case, the social goodness relation for us should not change. But that is what criteria that do not satisfy existence independence do: they may make the past important in determining what social goodness is now. Similarly, the levels of well-being of future people who aren't affected by what we do now may matter if we don't have existence independence (additive separability).

Additive separability is easy to come by, for the reasons just explained. It is also linked to how we treat population subgroups. Under some mild assumptions, if we assume the utilities of one subgroup have a separability property then, because all the principles we consider are anonymous, *every* subgroup is separable from its com-

[19] Broome (2004).

[20] Broome (2004).

[21] Broome (2004, pp. 50–63), Rachels (1998), Temkin (2012).

[22] Arrhenius (2000), Blackorby et al. (2005, p. 163).

plement and you get additivity.[23] Charlie and I used this result in a characterization of CLU and its transformed generalizations.[24]

NB & WB: *You and Charles Blackorby also worked on the implementation of interpersonal comparisons of well-being.*

DD: Yes, Charlie and I did a series of papers on alternatives to utility measurement such as welfare ratios and equivalence scales, and the use of compensating or equivalent variations in cost-benefit analysis.[25] Robin Boadway wrote a very nice paper on the problems with the use of compensating or equivalent variations.[26] He showed that if relative prices change due to a change in the distribution of endowments so there is no efficiency change, the sum of compensating variations is *always* positive. Consequently, it cannot be used as a welfare criterion,

One of the equivalence-scale papers, I wrote with Charlie Blackorby has an interesting result that no one believed for a long time: if you structure equivalence scales in a particular way and the reference preferences that you use to estimate demand systems are not PIGLOG,[27] you can estimate equivalence scales uniquely—and you have interpersonal comparisons.[28]

The equivalence scales Charlie and I used are independent of household expenditure but can depend on prices. Krishna Pendakur thought of a way to structure expenditure-dependent equivalence scales. They permit estimation under approximately the same conditions expenditure-independent scales need.[29]

NB & WB: *There used to be quite a bit of hostility within the profession when it comes to the use of interpersonal comparisons. You must have experienced this at some point.*

DD: I certainly did. But I always thought the whole taboo in economics around interpersonal comparisons was silly. People discovered that you didn't have to make them in order to do demand theory, for example. But if you want to work on normative issues, there is no serious substitute for their use.

NB & WB: *In the late 1970s, some prominent journals moved against social choice theory and some of them were very explicit about it. What was, and is, your reaction to this?*

DD: I can't say I liked it at the time. But this behavior of the journals involved is also why the Society for Social Choice and Welfare came into being, so some good things came out of it. Of course, the society would not exist without the hard work of Maurice Salles.

---

[23]Gorman (1968), Blackorby et al. (1978).

[24]Blackorby and Donaldson (1984).

[25]Blackorby and Donaldson (1985, 1987, 1990, 1991, 1993a, b).

[26]Boadway (1974).

[27]PIGLOG stands for the logarithmic subclass of price-independent generalized linear (PIGL) preferences; Muellbauer (1975).

[28]Blackorby and Donaldson (1993b).

[29]Donaldson and Pendakur (2004, 2006).

NB & WB: *There was some controversy among social choice theorists regarding the creation of the journal* Social Choice and Welfare. *Some people argued that its existence might make it even easier for other journals to reject, telling us to send our papers to SCW instead.*

DD: I was never worried about this. When Charlie and I had a hard time getting a paper accepted, we used to say we'd send it to our black hole—the *International Economic Review.* Whatever we sent them they kept, so our papers just disappeared into print.

NB & WB: *And the IER was very fortunate to get many excellent papers of yours. Your paper with Charles Blackorby and John Weymark on the diagrammatic approach to social choice with interpersonal comparisons is published in the IER as well.*[30] *Is there a specific journal—or journals—that left an impression on you over the years?*

DD: Another journal I really like is the *Journal of Economic Theory.* I know the founding editor Karl Shell very well—we were roommates at Stanford. I actually helped him along a little when he met his partner Susan Schulze. Karl and I were driving home in his car and there was a young woman by the side of the road with a car that wouldn't start. I said, "Karl, we have to stop to help her." He was reluctant at first but finally he agreed—and they have been together ever since. I am still in touch with them after all these years. He is not involved in the journal these days, though. John Weymark, one of my students, went to do his Ph.D. with Karl at the University of Pennsylvania. I thought Karl would be a great supervisor for him and it worked out very well. UBC was very fortunate to get John on the faculty later on, so I was able to collaborate with him on several papers.

NB & WB: *Did you ever feel thoroughly fed up or disillusioned with social choice theory?*

DD: There are some things that pass for social choice theory that don't fit with my prejudices; I think the field should be mostly concerned with normative issues but much of the matching literature, for instance, doesn't seem to fit into that. But this does not make me disillusioned with the field; I like the social choice scene very much.

NB & WB: *What is your take on welfarist-consequentialism as a criterion for social evaluation?*

DD: Almost all my normative work is about social evaluation. The investigation is about discovering the social good; no guidance about actions is included. I certainly am a welfare-consequentialist though, but not a maximizing consequentialist. In my view, every other way of thinking about social evaluation is just misguided. Some of Amartya Sen's work is a good example. He has this elaborate theory of functionings and capabilities.[31] The philosophers like to think of it as an account of well-being. But Sen won't have any of it—he doesn't want it to be welfarist. Taking it as an

[30] Blackorby et al. (1984).

[31] Sen (1985).

account of well-being seems to be perfectly reasonable to me but the denial does not. But Sen is not really hostile toward welfarism, unlike many others.

NB & WB: *You are one of the strongest proponents of utilitarianism. Would you have liked more engagement between utilitarians and its critics? Why do you think there has not been more engagement?*

DD: There has been in philosophy—it is one of those ongoing battles. In social choice theory and economics in general, utilitarianism took a big hit because of what happened in the 1930s when people made interpersonal comparisons a taboo—and it took us a long time to get out of that. At least we've made some progress since then.

NB & WB: *What kind of utilitarianism do you favor—act utilitarianism or rule utilitarianism?*

DD: I am neither an act utilitarian nor a rule utilitarian. But there are two main issues. One is axiological—the task of finding out what is good. Another one is once a notion of the good is established, how does this inform actions?

I am rather like Mill who thought that utility is the ultimate judge of everything normative.[32] But Mill's own work often didn't look very utilitarian. I don't believe we should be maximizers. For example, utilitarianism justifies certain arrangements we make—like marriage. People who marry and live together for a while usually look out for each other and that is a good thing. However, this means that they typically favor their spouses over everyone else, which many act utilitarians think inappropriate—we should care equally about everyone. But I think it is silly to quibble about that. The institution is justified and we take into account that people will favor their spouses when we make the calculations.

I read several articles on rule utilitarianism and I realized, from reading some psychology, that most people think of morality in terms of rules. It's the ten commandments all over again in different forms—instead of 'don't kill,' we get 'don't cheat on your taxes.' But even rule utilitarians behave like act utilitarians on occasion. Suppose your aunt Muriel shows up at your home with the ugliest hat you have ever seen. If she asks you how you like it, you tell her it's a beautiful hat. We do things like that—including telling a lie in this case—because they are good overall and rules such as 'never lie' can be suspended on some occasions. So I must be a hybrid of a rule and act utilitarian.

I have become interested in psychology and have been reading quite a bit recently. There seems to be some evidence that moral agents follow rules. In that sense, some psychological studies support rule utilitarianism. And it has the advantage of having a simple code they can follow. When people say they have principles, that's what they mean—they have a set of rules and they try to follow them.

Act utilitarianism can also require people to make great sacrifices for others. There are debates such as whether we should give away most of our money so more people in poverty can be better off? Peter Singer is a supporter of this view. Singer is a very

[32] Mill (1979a,1979b).

important person; he basically founded the field of applied ethics.[33] There are very few people who create a subdiscipline that's really an accomplishment. The only two people other than Peter Singer I know who did that are Kenneth Arrow and Charles Seeger, Pete Seeger's father, who created ethnomusicology.

Peter Singer has had a lot of influence in several areas. He also took on the issue of possibly terminating pregnancies in cases where the person, if born, would be very severely handicapped. He said some sensible things about it, not to everyone's liking. Some of the responses made the claim that, when a disabled person says "I am glad to be alive," they take this to imply that terminations such as the ones Singer considers should not occur. But it doesn't make sense to say that someone is better off alive than in a situation where he or she was never born: there is no person on one side of the comparison. Singer is very well aware of the distinction between people who exist and people who are not born. But the kind of unfounded case his opponents tried to make represents a trap many people fall into.

NB & WB: *Your own proposal—CLU—deals with issues of that nature. If the critical level is chosen above neutrality, it does not mean that people who live above neutrality but below the critical level should be terminated. It clearly recognizes that there is a difference between existing people and non-existing people.*

DD: Yes. Once a person is alive, she or he is alive in every feasible history of the world; people cannot be 'unborn.' This means that any moral discussion about a person can only be concerned with questions such as lengthening or shortening of her or his life; it is not possible to move the person to the group of the non-existent. There was a lot of confusion regarding this issue, triggered by intuitions that were based on fixed-population considerations that didn't carry over to variable-population issues. But I think the point has sunk in.

NB & WB: *Yet another topic where Peter Singer's interests overlap with yours is animal rights, which he was—and probably still is—quite actively involved in.*[34] *You and Charles Blackorby did some work on the ethical treatment of animals as well.*

DD: Charlie and I were, at different times, the token faculty-of-arts people on the UBC animal care committee. The committee's job is to ensure that the treatment of animals used in research is in line with ethical principles. It was an interesting experience and we even wrote a paper about it.[35] What is interesting is that it had an audience. We gave it to the animal care director of UBC and then I was asked to give it to a group at Simon Fraser University, including their director of animal care. They just wanted anything that would give them something to hang their hat on for guidance. They knew hurting research animals is a bad thing in itself. They also knew that using animals in the world for us humans creates different numbers of animals than there otherwise would be. So there is a variable-population dimension. No one else was concerned, though. Even the *Canadian Animal Care Standards* only

[33] Singer (1979).

[34] Singer (1975).

[35] Blackorby and Donaldson (1992).

talk about how the animals in the experiments are treated and not about how many of them are used. Nowadays, they try to minimize the number of animals used. We are talking about big numbers here. When I was on the committee at UBC, about 50,000 vertebrates (mostly rodents) were killed every year.

There are also a lot of interesting anecdotes we heard about. In one laboratory where they kept families of mice, one night a technician left a little bottle at the bottom of the cage. The next day, the bottle was full of mouse pee, so they found out that mice wanted to be clean—they wanted to have their own loo.

The other thing that came up was a really important problem. I don't know whether they still do this but, at the time, they produced monoclonal antibodies by infecting mice with cancer so they would grow antibodies in their stomachs which would be drained off. A mouse could last for two or three rounds of this, and they were not happy mice. So the question was whether we should have mice that last for three rounds or more mice that last for two rounds each. If the mice habituate and the suffering becomes less in the next round, or if the critical level for mice is positive, CLU recommends fewer mice and more suffering per mouse. That's the message of CLU: keep the population down even when it hurts. I know they followed our recommendation for a while but Jim Love, who was the director of animal care in these days, is now retired so things may have changed.

Interestingly, all the researchers who were on the committee sounded like utilitarians. I remember one of them saying that he was well aware that these experiments are bad for the animals. At the same time, the consequences of the experiments are very good and outweigh the suffering of the animals and that justified doing them—a straight utilitarian argument, except that they hadn't thought of the variable-population issue involved.

In some laboratories, though, they seemed to be quite into animal suffering. One of the things they do with rabbits when they test their eyes, for example, is that they give them drugs that paralyze them. This means there is no response from the animals that tells you anything, which is something that concerns me. An important case at UBC was the treatment of the baboons in a neurology laboratory, who were restrained for long periods. When the director shut down the baboon experiments, he gave me a report. He knew I was against these practices so maybe I had something to do with the decision.

One lesson to be drawn from all this is that most of the conventional methods proposed by economists won't work. For instance, taxing the animals will just lead to fewer of them being used but it won't improve their quality of life; likely, it'd be the opposite. So some regulations are definitely needed. The best animal care legislation is in place in Sweden. This is because Astrid Lindgren, the author of the Pippi Longstocking books, was in favor of better animal care and the prime minister at the time said what Astrid wants, she gets. So the Swedes were the real pioneers when it comes to animal care. Britain followed and passed a bill that is still quite good. But Canada has had a planned reform on the books for at least fifteen years now—but no action has been taken whatsoever.

NB & WB: *You and your family were vegetarians for quite a long time—for ethical reasons, we suppose.*

DD: Yes, we were vegetarians for twenty years. We read Peter Singer's book on animal liberation[36] and we just found it amazing. We were also living on Salt Spring Island during a sabbatical around that time and we had chickens. At one point, our kids asked what the difference was between Henny Penny in the hen house and the chickens in the supermarket. I said, well, the ones in the supermarket are dead. So the kids were very keen on becoming vegetarian. They were seven and three years old at the time.

NB & WB: *What do you take to be the relevant concept of utility?*

DD: My own view is that utility numbers are indexes of the quality of people's lives and people can have good lives in various ways. For example, happiness helps and things like accomplishments, skills and education also play an important role. All these things go into making up a good life, in addition to having shoes that don't leak and a place to sleep. I also think people can be mistaken about how good their lives are. For example, it is a real problem with old people who have dementia. In these cases, we have to be paternalistic to some extent. In general, people are also better in determining their well-being if they have some experience. For instance, someone may know that if they go to an AC/DC concert, that will make them happy but a classical concert won't. But if you ask people whether they want to live until they are 95, most people don't have a good idea about its contribution to the value of their lives.

NB & WB: *One of the criticisms of utilitarianism is that agents have control over and choose their preferences. Another is that preferences are not independent of the states over which preferences are defined. What are your responses to these?*

DD: I think the view that people choose their preferences is partly right. They do it in order to make their lives better, like having a preference to play the violin or the tin whistle, and they also can be right about this. But they can also choose preferences that are altruistic and their preferences may be slightly different from their well-being. I may have a preference for other people to do well or to have fewer people die from a terrible disease. But having fewer people die of Ebola doesn't make me better off. It is a good thing and I like it but I am not convinced about it being part of my well-being. So choosing preferences is not much of a criticism of utilitarianism in my view. Our preferences may or may not accord with our well-being.

But there is a kind of utilitarianism called preference utilitarianism, which means preferences are taken as fundamental. I think this is a mistake. We all know not everyone has well-formed preferences, like children who are seven years old, people with dementia, people with serious illnesses, people in comas.

As to state-dependent preferences, I do not think this is a viable way to go.

NB & WB: *In the nineteenth century, utilitarianism led to radical reforms. Is there a radical reform today that you think utilitarianism supports?*

---

[36] Singer (1975).

DD: Yes. Utilitarianism really says ‘Have the revolution right now’ but we can’t do that so it says instead ‘Do the best you can.’ Things like the increases in inequality we have been experiencing over the last forty years have been appalling, a claim supported by utilitarianism. I think utilitarianism is a kind of radical doctrine. But there are people who think of it as conservative. For example, John Harsanyi once gave a talk at a conference and he was pushing utilitarianism as a conservative view of social justice. I must admit this didn’t resonate with me.

NB & WB: *Are you a utilitarian in the strict traditional sense—that is, using the sum of well-being as the criterion to assess social goodness? Or would you also consider modified versions that allow for inequality aversion in individual utilities?*

DD: I certainly looked into the modified possibilities and, thanks to my coauthors, I have a good grasp of it. But I don’t really buy it. If you want inequality aversion in utilities, you find that additive separability is very attractive—it has a lot of desirable properties. This means you end up with transformed utilities and, mathematically, they are equivalent to utility functions. So the problem shifts away from assigning utilities to people and alternatives to assigning transformed utilities. I can’t quite see the point of doing that. It is hard enough to get utilities anyway and when we have to dream up some transforms that give us inequality aversion, this is very hard to do.

NB & WB: *There is also an issue regarding the information assumptions that are required. For the standard utilitarian criterion, you need cardinal unit comparability; for arbitrary transforms of utilities, the information assumptions are likely to be more convoluted and less transparent—and they have to be transform-dependent.*

DD: Yes, these transforms have to be very complex. But I am not all that worried about this particular difficulty. It is possible to assign numerically significant utilities to individuals. If we have cardinally measurable individual utilities and two utility levels with interpersonal comparability, you end up with numerical significance. It is easy to assign numbers to those two levels.[37]

NB & WB: *One natural normalization is the utility level that represents neutrality.*

DD: Yes, neutrality is interesting because it contains the idea of interpersonal comparisons: everybody has to have the same degree of well-being at neutrality.

NB & WB: *You received prizes both for your research and for your teaching, something that not many people have achieved. So in addition to your research accomplishments, you must have been an excellent teacher, too.*

DD: I found teaching very stressful at first but ended up really liking it once I got over the initial difficulties. The worst course I ever taught was in a huge lecture hall with about five hundred students and I thought nobody could have possibly liked that class. But there was an article in the Vancouver Sun recently in which the writer referred to the course. I had assigned John Kenneth Galbraith’s *The Affluent Society;* he thought it was wonderful and would never have read it if it hadn’t been for my

[37] Blackorby et al. (2005, Theorem 2.11).

class. But for me, it was not a very good experience; I found it awful. The class was full of engineers who were hissing and throwing paper airplanes, etc., all the time. They kept disrupting the class to the extent that I had to walk out at one point.

I also got a death threat written in blood on my door and I had to get the campus police involved. The note used very violent language; apparently, the likelihood of actual violence is higher the more violent the note is. I learned something from this experience but it was not good. I had ordered a new recorder that arrived on that day and I never liked it—the association with this dreadful event was just too strong.

The department finally decided to stop the experiment of running Econ 100 in an enormous class like that; it was hated by all teachers. After that the course was taught in smaller sections, which was a definite improvement.

NB & WB: *You must also have had pleasant and rewarding experiences in your teaching days.*

DD: I enjoyed teaching in the Arts I program; the experience was really good for me. I had gone through a mathematics program as an undergraduate and then went straight into economics so I didn't get much of a broad education at the university level. I thought if I went into the Arts I program (an interdisciplinary program for first-year students), I'd learn a lot myself—which was certainly true. I taught in a great-books program in Arts I. We started with Plato's *Republic,* read selections from the Bible and, eventually, we got to Hobbes, Rousseau, Marx and Mill. We did literature as well, of course, so we also covered books such as *Paradise Lost.*

NB & WB: *You taught this for several years—right?*

DD: I did it for four years: two years in the '60s and two more a couple of decades later.

NB & WB: *Did the course change much between the '60s and the '80s?*

DD: Not much as far as the material is concerned. But the students changed. In the '60s, you'd say, suppose that you were to form a commune, how would you write a constitution for it? They'd say what a great idea it was and they'd go ahead and do it. In the '80s, they'd just look at you like you didn't know which way was up. But there were other things they really went for in the '80s. For example, they were really good when it came to civil liberties. The two generations were different but they were always interesting.

I also got the best essay I have ever seen in the Arts I program. One year we read Nietzsche's *Beyond Good and Evil* and the novel *Jane Eyre.* One student wrote an essay that was a story of Rochester and Jane visiting Nietzsche. Rochester left to have his hands looked at by his doctor because of the fire. Then, Jane and Nietzsche had a (civilized) argument. Of course, Jane won. But every single word either of them spoke was taken from the respective book. This was long before you could search all books with a computer; it was really impressive.

NB & WB: *Did you keep in touch with some of your Arts I students?*

DD: Yes, John Weymark was in the program in the '60s. I tried to steer him into economics; I thought he'd be a very good economist and he certainly turned out that way. And I'm still in touch with a few Arts I students from the eighties. I've kept in touch or collaborated with quite a few of my economics students too. Walter Bossert and Krishna Pendakur are examples.

NB & WB: *Any recollections on teaching more advanced material?*

DD: The first time I taught the expected-utility theorem, I looked up Luce and Raiffa's book.[38] Their proof is just superb; the whole book is like that. The other books I really appreciated are Ken Binmore's mathematics books. His mathematical-analysis book really is terrific.[39] He also wrote two volumes on the foundations of analysis which are just as good.[40] Binmore's books do not attempt to oversimplify the material and yet succeed in making everything accessible, which is not easy to do. In contrast, it seems that math courses have been watered down over the last couple of decades. When I arrived at UBC, they had an introductory calculus course that was reasonably hard and, over the years, it declined into just applying formulas. But most students get their serious training in analysis and, fortunately, this continues to be a good course.

The part of mathematics that I actually enjoyed as an undergraduate was number theory. I also took a course in projective geometry. It is very weird and that makes it really fun; some non-standard geometries are just great in my view. Number theory was also useful for some of my work in economics. I have a paper with John Weymark where we used the fundamental theorem of arithmetic (which says that every integer can be uniquely decomposed into primes).[41] We got the idea for the paper when John sent me his earlier solo piece[42] and I wrote back to tell him that his generalized Gini measures do not necessarily satisfy replication invariance. So we decided to get together and figure it out, which is what we did. It was really fun to do.

NB & WB: *Unlike most of us, you have strong interests in philosophy, especially in applied ethics. You did a lot of reading in that area and presented at philosophy conferences and workshops. How would you describe your experiences interacting with philosophers?*

DD: I really enjoy it; I like to hang out with philosophers. They are like anyone else in the academic business—they do some interesting work, they make mistakes. They also tend to be a little wordy, though, which I sometimes find difficult to put up with. But all told, I like it. It also tends to be virtually mathematics-free, which always pleases me. Walter and I went to a few conferences of the International Society for Utilitarian Studies; three of them, I believe.

NB & WB: *Yes, they took place in New Orleans in 1997, in Winston-Salem in 2000, and in Hanover in 2005.*

---

[38]Luce and Raiffa (1957).

[39]Binmore (1977).

[40]Binmore (1981a, b).

[41]Donaldson and Weymark (1980).

[42]Weymark (1981).

DD: That's right. I think the best of these was the one in New Orleans. On that occasion, it was a real interdisciplinary group. I remember Peter Hammond was there, too. I always learned a lot from his comments at conferences.

NB & WB: *Are you still connected to the Center for Applied Ethics at UBC?*

DD: I have no connection now. They had some good visitors, though, like John Broome and Bob Sugden. Last, I heard was that they changed from having well-known visitors to inviting people who mainly work in their own areas. I went to many of their seminars even after I retired. Some were better than others, just as is the case in economics.

NB & WB: *You spent your entire career at UBC, which is quite unusual in our profession.*

DD: Yes, I retired in 1999 after 34 years. Retirement was good because I had serious health problems and most of what I did in the '90s was work. It was just the right time. I had a serious cancer in 2002, a hip replacement in 2005 and heart surgery in 2014.

NB & WB: *You must have witnessed many changes at UBC over the decades.*

DD: I certainly did. But over all these years, UBC treated its social choice theorists very well. They didn't intend to hire us as social choice theorists but the department supported us all the time. All of the social choice theorists have left the department but there still is a very strong commitment to public economics. Unfortunately, this does not happen in many major US universities where they don't seem to teach public economics at all. This is a mystery to me, considering that half of the economy is based on the public sector.

NB & WB: *Your own research focuses on economic theory and philosophy. Any interests in applied work?*

DD: Collaborating with applied people is very interesting. I did some work with Krishna Pendakur and learned a lot from him. But I like the elegant parts of economic theory; there are some really beautiful results. One problem some economists have is a belief that our models are 'true.' For instance, the Chicago school seems to believe that Milton Friedman's economics is true. There is evidence all over the place that it isn't. Joseph Stiglitz might whisper in their ears, "This stuff is false."

NB & WB: *What is your position on rights in social choice theory and how they should be respected?*

DD: I think it is worth having rights but they are instrumental in the sense that rights serve the purpose of increasing utility. The idea of rights is to make people's lives better. They are rules we enforce in the hope they have good consequences. My position on this is close to Jeremy Bentham's.[43] He claimed that, "*Natural rights* is simple nonsense: natural and imprescriptible rights, rhetorical nonsense,—nonsense

[43] Bentham (1843, p. 501).

upon stilts." I am not quite in that camp but I think rights are OK as long as it is understood that they are instrumental.

NB & WB: *Do you have a specific theory of justice that you support?*

DD: My view is that a just society is a good society. To me, utilitarianism gives you the answers in a way. But the theory of justice I think is the worst is supported by a group of philosophers who call themselves egalitarians. The problem is that they are mostly not—they spend most of their time arguing against equality and in favor of a limited amount of inequality. Often they say things like "People own themselves and therefore they get to keep whatever they earn, given just institutions." This just seems completely crazy to me, but the egalitarian view is very popular in philosophy. What I don't like in many of these approaches is that they give a lot of weight to individual responsibility. I don't think there is a coherent view of individual responsibility. The genetic component is big and the learning component is also large. To say, free choice implies responsibility neglects the way real people choose. Free choices may be made for many reasons or no reason at all. And the theory also has an out, which is the right-wing out. They say people are poor because they choose to be poor that they are lazy or they don't like to work, etc. When I said that to John Roemer, he responded that he would never say anything like it. I answered that I knew he wouldn't but it is consistent with the egalitarian project in general. The approach allows this notion of individual responsibility and there is nothing to prevent people from arguing that poverty is a matter of individual responsibility. People actually do argue this, all the time—and I don't want any of it.

Another view taken by some philosophers is that people should have equal opportunities for welfare. But that doesn't work either because the principle doesn't care whether people actually do well or they make lots of mistakes when faced with opportunities. We should care about what happens to people and opportunities are not enough for that. It's like what we do with kids. We want to give them opportunities, so we pay for post-secondary education. But the opportunity is not what we care about—we want them to get educated so that they have better lives.

NB & WB: *Utilitarianism (or, more generally, welfarism) certainly doesn't suffer from these shortcomings. Another argument in favor of welfarism is provided by Goodin,*[44] *and we believe that you find it very persuasive.*

DD: Welfarism depends critically on Pareto indifference: the idea that, in same person comparisons, equal utilities in two alternatives implies equal social goodness. Goodin gave us a great axiom: if you declare an alternative better than another, then there has to be at least one person who is better off in the former. If the social ordering of utilities is complete, the axiom implies Pareto indifference. And, in the presence of completeness and independence, welfarism is implied. Goodin's is a very natural and uncontroversial claim. His article was a great find for us.

Welfarism went out of style for a while. But people who worked on alternatives to welfarism didn't seem to get very far. They got results but they were not very satisfactory.

[44] Goodin (1991).

NB & WB: *They do seem to have some success, though; there has been a lot of follow-up work.*

DD: Yes, that's right—and we as a profession need to do this. We need to explore the possibilities. I think it is similar to what happened in social choice theory when people started to look at social choice functions. An the end that died out because we kept getting the same results as earlier theory did. But it was important work; we needed to find out. I like social choice theory that focuses on rationality. I once talked to the philosopher James Griffin at a conference; he said economists have done a lot more than philosophers on rationality.[45] It provided a lot of tools for the rest of us.

My coauthors and I benefited a lot from János Aczél's work on functional equations.[46] He is an extremely generous academic. I remember whenever there was a question on functional equations, you could just write to Aczél and he'd tell you right away if it had been solved; if not, he'd get back to you with an answer.

NB & WB: *Functional equations certainly played an important role in much of your work on population ethics, especially in an intertemporal setting. Another issue that arose in this context was the question of discounting the utilities of future generations. You always argued against discounting, although there was—and still is—some considerable resistance to this view within economics.*

DD: Yes, many people seem to think that, without discounting, the sacrifices imposed on the current population would be too big. But I don't see why you should change your idea of the good on the basis of this argument. Discounting also involves a violation of anonymity: it does not treat people impartially because it pays attention to their birth dates. If there are overlapping generations, discounting favors the old at the expense of the young and I don't think this is ethically attractive. Of course, there is room in many views of lifetime utilities for individual discounting of the value of future experiences.

But I am more concerned about some assumptions other than the perceived need to discount. For instance, there are economists who tend to conflate preferences with well-being. There are some issues (such as inequality and poverty) where there are good reasons to do this; it would be difficult to proceed otherwise. But the decisions economic agents make regarding their lives are not always in their long-term interests. Philosophers are much more aware of this problem. In our book, on population,[47] we follow the philosophers and talk about well-being rather than preferences.

Also, people come in households and that's often not taken into account properly. In an article with Charles Blackorby on poverty measurement,[48] we made a comment on incomes being assigned to households so we have to account for individuals somehow. We mentioned equivalence scales but got one of our results backward. We were saved by a referee, fortunately. Having good referees is very important.

[45] Samuelson (1938a, b), Arrow (1959), Sen (1971).

[46] Aczel (1966).

[47] Blackorby et al. (2005).

[48] Blackorby and Donaldson (1993a).

NB & WB: *To do a good refereeing job can take a lot of time. There seem to be more and more papers around nowadays as compared to a few decades ago. Do you think there are, in some sense, 'too many' papers written in economics?*

DD: I think so. The tenure pressures are much higher than they used to be and people are tempted to make tenure decisions on the basis of the number of published papers; they often do not pay enough attention to as yet unpublished work. Things were much easier when I started out. The universities were still expanding and were desperate to get graduate students. Sputnik was the best thing that ever happened to me—it got me into graduate school because the Americans started to panic that they would lose the space race.

But even nowadays I enjoy the academic scene. For all the bad things that have happened to universities, we have managed to keep the activity going. It also is a good thing for people like us—the existence of universities gives us a way to be employed. I often wonder what would happen to theorists if the university option didn't exist.

I once worked for IBM during the summer while I was a graduate student at Stanford. It was an education in itself but it convinced me that I didn't want to work for a business. It was interesting, though. They played all these intellectual board games—like Kriegspiel and Go—at lunchtime. I thought that was very cool.

NB & WB: *And you participated?*

DD: On occasion. I just couldn't play Kriegspiel. This is a form of blind chess where there are three chess boards. Each player has a separate board, and so does the umpire. A player can only infer the position of the opponent's forces from limited information provided by the umpire who disallows illegal moves. So you have to figure out where the other guy is moving. You try to make a move and if the umpire tells you that you can't do that, this tells you something about the possible positions of the opposing pieces. It is intriguing—and it takes a lot of lunch breaks to finish a game. The same is true for Go, of course. They had other games, too, but these were the two that really attracted me. I had a girlfriend at the time whose father played Go. But I never really learned it. I guess I don't have the game-playing gene; I don't like playing games all that much.

NB & WB: *In the mid-1980s, you were involved in writing a book on Expo '86 (which took place in Vancouver).*[49] *You and your fellow contributors certainly had something to say on the subject—and not everybody wanted to hear it. What triggered that?*

DD: We had a bad provincial government in British Columbia at the time. It promoted the standard right-wing (neo-libertarian) positions—kill the unions and all that. So eleven of us formed a policy group which the university didn't like. But they tolerated us.

NB & WB: *And you had tenure.*

[49] Allen and Rosenbluth (1986).

DD: All but two of us. The BC minister of education at the time, Pat McGeer, said that he would love to fire us all if he could but we had tenure so he couldn't. But it wasn't exactly true: Angela Redish and Margaret Slade were in the group and they didn't have tenure at the time.

NB & WB: *This was not your only involvement in politics. You continue to play a very active role in the causes that you care about.*

DD: I do, I always have. I got my political education in California, mainly from the civil rights movement.

NB & WB: *And the Vietnam war was going on at the time you were there, too.*

DD: Yes, Vietnam was such a crime that I couldn't help getting involved.

NB & WB: *Kenneth Arrow was on your side, too.*

DD: He was very good on political issues. He was shy about politics but he had the right kind of commitments, and I appreciated that. I remember one time he asked me if he should sign his name to be a sponsor for an anti-Vietnam war event on campus, although he knew he would be out of town at that time, and did I think that this was an ethical thing to do. I answered, "Well, that depends on whether you want to end the war or not." He responded, "Oh yes," and signed immediately.

NB & WB: *You kept active in the peace movement as well. For instance, you participate in the peace walk every year, we believe.*

DD: Yes, whenever they organize them. I missed the last one because of my heart surgery. I have also been to several climate change demonstrations.

NB & WB: *Music seems to be an enduring activity for you; our understanding is that this has always been and continues to be a real passion. Is there anything in the way you compose or play music that is common to the way you do economics?*

DD: I suppose I use roughly the same method in writing a paper or a piece of music, but I am not really aware of it.

Music was first. I started to play the piano at the age of five and moved to organ and harpsichord, along with trombone, other brass instruments, and bassoon in high school. Needless to say, I was not equally proficient in them all. When I was an undergraduate, I got into guitar and folk music and I enjoyed that. It continued when I was in graduate school. In 1981, I got the rheumatism that I still have, and I had to give up piano and guitar. So I got back to wind instruments: recorders and whistles this time. My father was a singer; his specialties were Baroque music (which he sang in churches and at concerts) and musical theater. He also loved Harry Lauder's songs—he was a Scottish immigrant. So I got to know all Harry Lauder's songs when I was growing up.

I was an undergraduate in a mathematics department where most of the faculty members were musicians. So that made me feel good (I always thought I didn't belong, as I told you earlier). And my interests in economics and social choice spill over into the music. I'm interested in peace and social justice and wrote a piece called *Swords into Plowshares.*

So I guess there is a connection of sorts, but it's a weak one. But I certainly like playing, writing and recording music. I also have two Web sites that have music on them. One is http://www.davidsmusic.net the other http://www.dark-willow.net. The second of these is my folk duo with Ellen van der Hoeven. Everything on both Web sites is downloadable and I keep them updated, more or less. I also made two CDs. I recorded them, edited them, mastered them and got them reproduced. I had to learn a lot to do all that.

NB & WB: *You wrote the pieces on the CDs yourself?*

DD: Not all of them. On the more recent CD, one piece is traditional; I wrote the rest. On the first, there is a mix. It has music written by other people and traditional pieces as well as a few of mine. There is also more recorder playing because my tremor killed my recorder playing before I finished the second CD. In spite of the tremor, I can still play keyboards and I may get back to composing, which I haven't done since I got back from the hospital last time.

NB & WB: *And we, along with all your other friends, hope very much that you will continue to enjoy your music for many more years to come.*

References

Aczél, J. (1966). *Lectures on functional equations and their applications*. New York: Academic Press.

Allen, R., & Rosenbluth, G. (1986). *Restraining the economy: Social credit economic policies for B.C. in the eighties*. Vancouver: New Star Books.

Archibald, G., & Donaldson, D. (1976a). Non-paternalism and the basic theorems of welfare economics. *Canadian Journal of Economics*, *9*, 492–507.

Archibald, G., & Donaldson, D. (1976b). Paternalism and prices. In M. Allingham & M. Burstein (Eds.), *Resource allocation and economic policy* (pp. 26–34). London: MacMillan.

Archibald, G., & Donaldson, D. (1979). Notes on economic equality. *Journal of Public Economics*, *12*, 205–214.

Arrhenius, G. (2000). An impossibility theorem for welfarist axiologies. *Economics and Philosophy*, *16*, 247–266.

Arrow, K. (1951). Social choice and individual values. New York: Wiley, second ed. 1963.

Arrow, K. (1959). Rational choice functions and orderings. *Economica*, *26*, 121–127.

Arrow, K. (1970). *Essays in the theory of risk-bearing*. Amsterdam: North-Holland.

Barry, B., & Rae, D. (1975). Political evaluation. In N. Polsby & F. Greenstein (Eds.), *The handbook of political science* (vol. 1). Addison-Wesley, Reading.

Bentham, J. (1843). Anarchical fallacies. In J. Bowring (Ed.), *The works of Jeremy Bentham, Vol. II* (pp. 489-534). Edinburgh: William Tait.

Binmore, K. (1977). *Mathematical analysis: A straightforward approach*. Cambridge: Cambridge University Press, second ed. 1983.

Binmore, K. (1981a). *The Foundations of analysis: A straightforward introduction, Book 1 logic sets and numbers*. Cambridge: Cambridge University Press.

Binmore, K. (1981b). *The foundations of analysis: A straightforward introduction, Book 2 topological ideas*. Cambridge: Cambridge University Press.

Blackorby, C., Bossert, W., & Donaldson, D. (1996). Quasi-orderings and population ethics. *Social Choice and Welfare*, *13*, 129–150.

Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social choice theory, welfare economics, and ethics*. Cambridge: Cambridge University Press.

Blackorby, C., & Donaldson, D. (1977). Utility vs equity: Some plausible quasi-orderings. *Journal of Public Economics*, *7*, 365–381.

Blackorby, C., & Donaldson, D. (1978). Measures of relative equality and their meaning in terms of social welfare. *Journal of Economic Theory*, *18*, 59–80.

Blackorby, C., & Donaldson, D. (1980). A theoretical treatment of indices of absolute inequality. *International Economic Review*, *21*, 107–136.

Blackorby, C., & Donaldson, D. (1984). Social criteria for evaluating population change. *Journal of Public Economics*, *25*, 13–33.

Blackorby, C., & Donaldson, D. (1985). Consumers' surpluses and consistent cost-benefit tests. *Social Choice and Welfare*, *1*, 251–262.

Blackorby, C., & Donaldson, D. (1987). Welfare ratios and distributionally sensitive cost-benefit analysis. *Journal of Public Economics*, *34*, 265–290.

Blackorby, C., & Donaldson, D. (1990). A review article: The case against the use of the sum of compensating variations in cost-benefit analysis. *Canadian Journal of Economics*, *23*, 471–494.

Blackorby, C., & Donaldson, D. (1991). Adult-equivalence scales, interpersonal comparisons of well-being and applied welfare economics. In J. Elster & J. Roemer (Eds.), *Interpersonal Comparisons of Well-Being* (pp. 164–199). Cambridge: Cambridge University Press.

Blackorby, C., & Donaldson, D. (1992). Pigs and guinea pigs: A note on the ethics of animal exploitation. *Economic Journal*, *102*, 1345–1369.

Blackorby, C., & Donaldson, D. (1993a). Household equivalence scales and welfare comparisons: A comment. *Journal of Public Economics*, *50*, 143–146.

Blackorby, C., & Donaldson, D. (1993b). Adult-equivalence scales and the economic implementation of interpersonal comparisons of well-being. *Social Choice and Welfare*, *10*, 335–361.

Blackorby, C., Donaldson, D., & Weymark, J. (1982). A normative approach to industrial-performance evaluation and concentration indices. *European Economic Review*, *19*, 89–121.

Blackorby, C., Donaldson, D., & Weymark, J. (1984). Social choice with interpersonal utility comparisons: A diagrammatic introduction. *International Economic Review*, *25*, 327–356.

Blackorby, C., Primont, D., & Russell, R. (1978). *Duality, separability, and functional structure: Theory and economic applications*. Amsterdam: North-Holland.

Boadway, R. (1974). The welfare foundations of cost-benefit analysis. *Economic Journal*, *84*, 926–939.

Broome, J. (2004). *Weighing lives*. Oxford: Oxford University Press.

Donaldson, D., & Pendakur, K. (2004). Equivalent-expenditure functions and expenditure-dependent equivalence scales. *Journal of Public Economics*, *88*, 175–208.

Donaldson, D., & Pendakur, K. (2006). The identification of fixed costs from consumer behavior. *Journal of Business and Economic Statistics*, *24*, 255–265.

Donaldson, D., & Pendakur, K. (2014). *Applications of population principles: A note*. Unpublished manuscript.

Donaldson, D., & Weymark, J. (1980). A single-parameter generalization of the Gini indices of inequality. *Journal of Economic Theory*, *22*, 67–86.

Goodin, R. (1991). Utility and the good. In P. Singer (Ed.), *A companion to ethics* (pp. 241–248). Oxford: Basil Blackwell.

Gorman, W. (1968). The structure of utility functions. *Review of Economic Studies*, *32*, 369–390.

Lorenz, M. (1905). Methods of measuring the concentration of wealth. *Publications of the American Statistical Association*, *9*, 209–219.

Luce, R., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.

Mill, J. (1979a). On liberty. In: J. Mill (Ed.), *Utilitarianism; On Liberty; Essay on Bentham* (pp. 126–250). Collins, Glasgow, originally published in 1859.

Mill, J. (1979b). Utilitarianism. In J. Mill (Ed.), *Utilitarianism; On Liberty; Essay on Bentham* (pp. 251–321). Collins, Glasgow, originally published in 1861.

Muellbauer, J. (1975). Aggregation, income distribution and consumer demand. *Review of Economic Studies*, *42*, 525–543.

Parfit, D. (1976). On doing the best for our children. In M. Bayles (Ed.), *Ethics and population* (pp. 100–102). Cambridge: Schenkman.

Parfit, D. (1982). Future generations, further problems. *Philosophy and Public Affairs*, *11*, 113–172.

Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.

Rachels, S. (1998). Counterexamples to the transitivity of better than. *Australasian Journal of Philosophy*, *76*, 71–83.

Roberts, K. (1980a). Possibility theorems with interpersonally comparable welfare levels. *Review of Economic Studies*, *47*, 409–420.

Roberts, K. (1980b). Interpersonal comparability and social choice theory. *Review of Economic Studies*, *47*, 421–439.

Samuelson, P. (1938a). A note on the pure theory of consumer's behavior. *Economica*, *5*, 61–71.

Samuelson, P. (1938b). A note on the pure theory of consumer's behavior: An addendum. *Economica*, *5*, 353–354.

Sen, A. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.

Sen, A. (1971). Choice functions and revealed preference. *Review of Economic Studies*, *38*, 307–317.

Sen, A. (1985). *Commodities and capabilities*. Amsterdam: Elsevier.

Singer, P. (1975). *Animal liberation: A new ethics for our treatment of animals*. New York: Avon.

Singer, P. (1979). *Practical ethics*. Cambridge: Cambridge University Press.

Temkin, L. (2012). *Rethinking the good: Moral ideals and the nature of practical reasoning*. Oxford: Oxford University Press.

Weymark, J. (1981). Generalized Gini inequality indices. *Mathematical Social Sciences*, *1*, 409–430.

Peter Fishburn

Steven Brams, William Gehrlein, Fred Roberts, and Maurice Salles

1. Who were the most influential people who acted as mentors during your career:
   (a) In your general work in mathematics?
   (b) In social choice and voting theory?

1a. During my career, I was supported and encouraged by people who allowed me to pursue my own interests. Principals were Russell Ackoff (operations research and decision theory), Duncan Luce (utility theory and mathematical psychology), Nicholas Smith (decision and utility), Jimmie Savage (decision under uncertainty and subjective probability), and Ronald Graham and Andrew Odlyzko (discrete math).

1b. I had no guiding presence in social choice and voting, but two special people—William Gehrlein and Steven Brams—were invaluable collaborators.

2. What attracted you to the study of social choice/voting in general and to approval voting in particular?

I was the undergraduate chair of elections at Penn State in 1957–1958 where I learned about voting procedures and the contretemps of a few candidates. That experience influenced my focus on theory rather than practice in social choice. The

S. Brams
New York University, New York, USA

W. Gehrlein (✉)
University of Delaware, Newark, USA
e-mail: wvg@udel.edu

F. Roberts
Rutgers University, New Brunswick, USA

M. Salles
University of Caen-Normandy, Caen, France

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_10

books of Kenneth Arrow, Duncan Black and Amartya Sen affected my later research. Duncan Luce provided the means to write *The Theory of Social Choice* at the Institute for Advanced Study in Princeton during 1970–1971. Steve Brams guided me into approval voting in the mid-1970s.

3. What results that were obtained in social choice/voting by you and/or your many co-authors personally strike you as being the most important contributions? What particular papers or books from your long list of publications do you consider to be the highlights of your career?

Important personal contributions include the aforementioned book, results on Arrow's theorem with infinite voters (1970), single-peaked preferences (1972) and paradoxes of voting (1974). Important collaborations with Bill Gehrlein involve probabilities of voting paradoxes (1976) and coincidence probabilities for simple majority and positional voting (1978), and with Steve Brams on the foundations of approval voting (1978). Many later works have precedents in the preceding contributions.

Other career highlights are the books *Utility Theory for Decision Making* (1970) and *Nonlinear Preference and Utility Theory* (1988), an extensive retrospective on the utility theory of von Neumann and Morgenstern (1989), and results in the theory of ordered sets in *Interval Orders and Interval Graphs* (1985).

4. You collaborated with a great number of researchers, including several over a period of many years. What, in your experience, makes for a fruitful and long-lasting collaboration?

Fruitful long-term collaboration requires congeniality, mutual respect, open communication, willingness to share, an ability to ask penetrating questions, and a desire for joint exploration.

5. In the study of subjects like voting and fair division, how much importance do you attach to laying the theoretical foundations for their study compared to seeing your findings, where practical, tested empirically or actually tried out in real-world settings?

My abiding interest has been theoretical foundations. Real-world tests and applications are clearly important but were of marginal interest.

6. What are the contributions by other researchers in social choice/voting that you view as being the most important or influential?

Contributions by others that seem important in hindsight include Condorcet's essay on cyclical majorities (1785), Kenneth May's axiomatization of simple majority (1952), Ken Arrow's impossibility theorem (1960), John Banzhaf's results on weighted voting (1965), Y. Murakami's formalization of representative majority (1966), Richard Niemi and Herbert Weisberg's voting paradox probabilities (1968), Richard Zeckhauser's analysis of majority rule with lotteries (1969) and Amartya Sen's impossibility of a Paretian liberal (1970).

This clearly is an old geezer's list. I have not kept up with the literature since I retired, so recent contributions are absent.

7. We know that you did a lot of research in other areas of mathematics, mathematical psychology, measurement, decision theory, economics, etc. Could you tell us a bit about your work in those areas, and what attracted you to them?

As noted above, Russ Ackoff, Duncan Luce and Jimmie Savage were instrumental to my interests in decision theory, utility theory, choice under risk or uncertainty, subjective probability, and allied areas of economics and mathematical psychology. What attracted me to these subjects were (a) the possibility of formulating axioms that justify models of choice, (b) methods of measuring the terms in those models and (c) ways of applying the models under partial or complete information.

When I joined Bell Labs in 1978, Ron Graham and others directed my attention to discrete math topics, including ordered sets, graph theory and combinatorics. This opened up new opportunities to do the kind of mathematics I enjoyed and involved many new collaborators. Along with Graham and Odlyzko, I am deeply grateful to Larry Shepp, Jim Reeds, Bill Gehrlein, Tom Trotter and Fred Roberts. Others who involved me in their research areas were Robert Calderbank (coding theory) and Paul Erdos (convex sets and planar geometry).

8. Your work does not include a large volume of work that is based in game theory, except for some foundational aspects of utility theory, and for some work in approval voting. Can you tell us why you chose not to place more emphasis in game theory?

I do not recall a conscious decision to avoid doing more in game theory. Perhaps other fields offered more opportunities to do the mathematics that I enjoyed most.

9. You have completed much of your research in a number of different areas of mathematics but you did not actually pursue any academic degrees in mathematics itself. What led you to your choices?

My undergraduate degree in industrial engineering and my Ph.D. in operations research used lots of math. The latter was heavy in both pure and applied mathematics. The undergraduate course led naturally to a graduate program in operations research.

10. Your career trajectory was quite different from that of the typical life-long academic researcher in social choice/voting. What drew you to this field?

During my years in advanced research at the Research Analysis Corporation (RAC) in the 1960s, I developed an interest in social choice as an offshoot of decision theory. I then devoted my year at the Institute for Advanced Study to research and writing in social choice theory. Of those who encouraged this new direction later, Steve Brams deserves special mention.

11. You spent a lot of your career as a researcher in an industrial laboratory. Can you tell us how much you needed to justify the relevance of your work there? Also, how did the role of research in such a lab change over the years?

As noted at the outset, I chose most of my research topics. This was true at RAC (1963–1970), Penn State (1971–1978) and Bell Labs (1978–2001). Very little justification was needed beyond my employers' faith in my qualifications and productivity.

I was hired by Bell Labs in an economics department, but a few years later, the department was abolished. I moved into mathematics and stayed there until I retired in 2001. In the meantime, the breakup of the Bell System had serious repercussions for the Labs, and when its math group was decimated, I remained in a reduced department. The pervasive orientation of the department was discrete mathematics. When I strayed too far into social choice, word came from on high that, to put it tactfully, I should do less in that area.

Many years earlier, the advanced research department at RAC was abolished. I have the odd distinction of being the last regular researcher to leave as the department disintegrated. Then, after my year at the Institute, it was on to Penn State as a Research Professor. During our time there, my wife Jan completed a Ph.D. in Religious Studies. She was then hired by the faculty of the Theological School at Drew University, and I ended up at Bell Labs in the bargain.

12. As someone who worked at the interface among many disciplines, what kinds of background and skills made it possible for you to do this so successfully?

My broad training in pure and applied mathematics made it possible to do research in many subjects and to engage in interdisciplinary work with very bright people. Curiosity and a willingness to share and be challenged by others surely helped.

13. You had an extraordinarily productive career over a long period of time. What was it that motivated you to remain so productive over the entire span of your career?

I loved doing research and discovering new results. I liked to write and see things into print. Of vast importance were great places to work, gifted co-authors and marvelous support personnel. Throughout my career, I had a private secretary or a dedicated mathematical typist, so there was always excellent help to turn my handwritten documents into good copy. A desire for recognition cannot be gainsaid, but was a secondary concern.

14. When you retired, it seems that you ceased to do research of the kind that you were so noted for and, instead, decided to pursue other activities. What led you to this decision? What other activities have engaged you since you have retired?

Times changed. Fewer resources were being devoted to basic research, the technical support I had enjoyed diminished, and I was unwilling to retrain for a computerized world. It was a blessing to earn a living doing what I loved, but as we entered a new century it was time to move on.

My wife retired in 1995 as Dean of Drew's Theological School to pursue her own interests. I retired six years later. My favorite things since then are yard work, cornet practice, crossword puzzle solving, family activities, a daily Scrabble session with Jan, reading novels and the New York Times, and TV.

15. Are there any other thoughts that you would like to share regarding anything that we have not touched on?

The history and breadth of my research areas were always attractive and resulted in a few dozen surveys. A case in point is the monograph *Interprofile Conditions and Impossibility* (1987), which exposits Arrow's impossibility theorem and numerous contributions to social choice based on his approach and succeeding developments.

Allan Gibbard

Matthew D. Adler and John A. Weymark

This interview was conducted on October 15, 2015 at the home of Allan Gibbard (**AG**) in Ann Arbor by Matthew Adler (**MA**) and John Weymark (**JW**). Adler participated by Skype from his office at the Duke Law School. The text of the interview has been edited to improve its readability, to clarify some of what was originally said, and to provide bibliographic details for the works cited.

JW: Allan, let us begin with some questions about your background. You grew up in West Virginia where your father, Harold Gibbard, was a prominent Professor of Sociology. Did growing up in an academic family play an important role in your own decision to become an academic?

AG: Well, the psychologists tell us that we don't know the causes of our actions just by introspection, but I'm sure it had an important influence. I grew up with academic values, and for a long time I thought that being a perpetual student was the ideal life.

JW: You went to Swarthmore College as an undergraduate where you majored in Mathematics and minored in Physics and Philosophy. Why did you choose this combination?

AG: Well, I had a wonderful introductory Philosophy teacher, Jerome Shaffer, and had no idea I was interested in Philosophy before that. I thought that philosophers specialized in giving fallacious proofs of the existence of God. I'd gotten fascinated

M. D. Adler
Duke Law School, 210 Science Drive, Durham, NC 27708, USA
e-mail: adler@law.duke.edu

J. A. Weymark (✉)
Departments of Economics and Philosophy, Vanderbilt University, VU Station B #35189, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA
e-mail: john.weymark@vanderbilt.edu

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_11

with Physics, I think in 5th grade, but then laboratory Physics didn't work very well for me, so Mathematics was more down my line, although I don't think I would have been a good enough mathematician to thrive as a professional mathematician.

JW: In your 2006 interview with Alex Voorhoeve (Voorhoeve 2009), you mentioned that both Shaffer and Richard Brandt were influential. Later, we'll talk about how Brandt influenced your choice of Ph.D. thesis. Are there other ways that Philosophy faculty at Swarthmore influenced the kind of issues you considered?

AG: Well, Brandt was the most important figure for me, and he had a very direct way of trying to figure out what's really at stake with a problem that people were discussing, and that very much appealed to me. I didn't study with him until the Fall of my senior year when I took his Moral Philosophy seminar, although my freshman year roommate was in his introductory class and we discussed things a great deal. So, the appeal of Brandt's approach to Philosophy was just very strong for me.

MA: This is just a quick follow up. Brandt was later at Michigan and you, of course, have been at Michigan for a while. Were those connected?

AG: Yes, I think so. Brandt left Swarthmore the year after I graduated, and during the era when he was there, Swarthmore had one of the top departments and it was really built by Brandt. And then in the 1960s universities came to dominate colleges, so there weren't going to be college departments of the eminence of the Swarthmore one by the time the 1960s came. But, it was remarkable when I came to Michigan that two other members of the department had been teaching as young Assistant Professors at Swarthmore when I was there and one other member of the department had been a student at Swarthmore when I was there.

MA: Who were those people?

AG: Larry Sklar and Jaegwon Kim had both been Assistant Professors at Swarthmore. I got to know Larry there, but never took a class from him. But I took a Symbolic Logic course from Kim. John Bennett was the one who I think was a freshman at Swarthmore when I first got to know him.

JW: You have partly answered the next question because in your later work in social choice theory you employ formal logic and set theory. You have already explained how you learned about logic as a student. Did you also learn about set theory from your days at Swarthmore?

AG: Yes, I think the Swarthmore program was much better then than it was later when my son went there, because there was a real interest in the abstract aspects of Mathematics. Those of us who had finished second year calculus as freshmen were put in a course that studied things like the Peano axioms and set theory and various other sorts of highly abstract aspects of Mathematics and that was the part of Mathematics that fascinated me.

JW: Following your graduation from Swarthmore, you spent two years with the US Peace Corps as a high school teacher in Ghana. What prompted you to do this and has this experience had an influence on your research?

AG: Well, I was, like everyone, worried about the balance of my obligations to others and doing the things I wanted to do. I didn't think I could be an Albert Schweitzer who did the things he wanted to do until he was thirty and then became a doctor in Gabon, but I decided I could spend two years trying to be of service before I went on to do the things I most wanted to do. I discovered in the Peace Corps that virtually nobody else in the Peace Corps had such a motivation, or would at least avow it. They were of course quite devoted to their teaching. For me as well there were great benefits to being in the Peace Corps. I'd never been outside North America, and getting to know another culture broadened my view of human possibilities in an important way. So, I got a lot of personal benefit from being in it.

JW: Do you think it influenced how you approach moral issues?

AG: Yes, I think so. I wrote a paper for John Rawls in graduate school where I was trying to argue that there are lots of people who don't have the kinds of goals that Rawls attributes to his parties in the Original Position (Rawls 1971). I was very struck that, whereas I had grown up in an academic family where the highest value was having a free mind and enquiring into things, in Ghana the highest values were behaving well and being obedient. Of course, you can get that within American culture too. I had grown up in a small slice of American culture. Indeed it occurred to me later that I'd grown up in a colonial situation, and that the university was full of people who weren't from West Virginia or similar places but were surrounded by people who were. I imagine that's the way it is where you are too.

MA: Allan, was the concern there that people would have different conceptions of the personal good or was it rather more of a concern like Tim Scanlon's that people behind the veil would be motivated by altruistic considerations as well as considerations of personal good (Scanlon 1998)?

AG: I always thought that people would be motivated by altruistic considerations as well as by their own good, but Rawls takes freedom to form and revise one's plans as a top value that the social contract is to promote. The people I met in Ghana may have had this view of their own lives—I'm not sure—but it didn't seem to be part of their child-raising, as it is among the American liberals I know. Good behavior and obedience seemed to be what was stressed.

MA: Right. But does one get around that by looking into fully informed—ideal—preferences? Of course, ideal preferences are something you talk about a lot in *Wise Choices, Apt Feelings* (Gibbard 1990b) and *Thinking How to Live* (Gibbard 2003).

AG: Well, the people in Ghana certainly had preferences, but as I say, it didn't seem that they had been formed in an ideal way by free-thinking about what to want. That didn't seem to loom very large for people who were concerned that their children were going to grow up well behaved but not concerned that their children grow up with enquiring minds. This isn't to say that people didn't have enquiring minds, but it didn't seem to be the same sort of value as it was among the academics I'd grown up with.

JW: Let us move on to the next stage of your career. In 1969, you started graduate school at Harvard in Philosophy. In your first term, you took a course with John Rawls in which you read parts of the manuscript of his *A Theory of Justice* (Rawls 1971). How did Rawls' book influence your own views about these issues then and subsequently?

AG: Well, his argument that altruism is not the strongest motivation, that some sort of reciprocity is a stronger motivation, made a big impression on me. And it was a very exciting class. Rawls was thinking of the Original Position as a site for bargaining, and one day some of us stayed after class and argued with him for about an hour that people in the Original Position wouldn't have clashing interests that they knew about, so that the bargaining wasn't really genuine bargaining. I think that's the position that Rawls took eventually. All that was very exciting. The importance of the idea of fair reciprocity was one thing that struck me from being exposed to Rawls that I hadn't been struck by being exposed to Brandt.

JW: Willard Quine and Hilary Putnam were also on the Harvard faculty at that time. Did your interactions with them or any of the other Harvard Philosophy faculty play a major role in shaping your views about philosophical issues?

AG: Well Quine had a big influence. I didn't end up having all of his views, but he was sort of the dominant figure, and at Swarthmore I was the sort of logical positivist he criticized. Swarthmore was very caught up in the ethos of the, say, late 1930s to 1950s in English analytic philosophy. I remember reading Quine's "Two dogmas of empiricism" (Quine 1951) sitting under a tree and feeling quite dismayed and then trying to figure out how to integrate those sorts of views into my view of the subject. My most recent book, *Meaning and Normativity* (Gibbard 2012), is really trying to see how much of the older approaches one could save if one thought that questions of meaning were normative questions and not purely empirical questions.

And Putnam, well, Putnam was a sort of avid communist. Engels was I think his favorite philosopher, and I had had enough exposure to what communism was like to be resistant to that. So I listened to Putnam a lot, but I certainly didn't become a follower. I guess it was later that he had his transformation that resulted in his transformative philosophical article, "The meaning of 'meaning"' (Putnam 1975), and at that time he gave up his hard-line communism. I'm not sure whether it was Maoism or what; most of the graduate students were enthusiastic Maoists. Putnam also taught a course on the advanced logic of the Continuum Hypothesis and things like that. I learned a lot from that.

JW: I understand that your 1971 Ph.D. thesis, *Utilitarianism and Coordination* (Gibbard 1990a), is a subject that you started thinking about when you were still at Swarthmore. Can you tell us a little bit about the development of the origins of your thesis and what the basic ideas in the thesis were?

AG: I took Brandt's Moral Philosophy seminar in my senior year, and that was, I suppose, the biggest undergraduate influence on me. And one week the subject was rule utilitarianism and variations on it. Brandt each week would have a list of articles that were central to the subject for the week, and each student would every two weeks write a short paper. This was my week to produce a paper, and as he was talking about the readings for that week, he said, "Oh, and I have a little thing that I've placed on reserve" (Brandt 1963). But somehow I forgot about Brandt's "little thing" and read the other papers and formed a view and wrote my paper. And then Brandt said, "Oh, did you get a look at my 'little thing'?" And I turned red and realized I'd forgotten it, and so I felt I had to write another paper. But by then I'd formed independent views on the issue his paper concerned. He argued in his paper that a fairly straightforward form of rule utilitarianism would be what he called "extensionally equivalent" to act utilitarianism. I had concluded in my work before reading his article that they weren't, that there were situations where the question "What if everyone did the same?" made a real difference. So, I wrote the second paper and Brandt said, "Well, if this is true, it ought to be publishable." Of course that's an exciting thing to hear as an undergraduate. So, I worked on a revised version, and he read it and said, "If this is true, it ought to be publishable. Of course, it isn't publishable yet." And then I had to do a few iterations.

When I was in Ghana I got a letter from him saying that a young genius named John Troyer in his seminar had produced a paper on my paper, and Brandt wrote a two-page letter along with Troyer's paper. I thought Brandt and Troyer were both getting matters completely messed up, and I decided that the only way to write this paper was to pretend I was talking with my very intelligent little sister when she was twelve. I was fifteen when she was twelve, and in that period, I was explaining lots of things to her, and so I had lots of practice at explanation with immediate feedback on whether I was getting things across. And so imagining explaining things to her produced a paper that everybody liked, and it was accepted by the *Australasian Journal of Philosophy* (Gibbard 1965) before I got to graduate school.

I had a four-year fellowship at Harvard, and most people took longer than four years to finish the Ph.D. But if I took longer than four years, I thought, I was going to have to teach "Hum 5," the course Humanities 5, which is a very broad survey of so-called western philosophical thinking, and it seemed to me that teaching that course was going to be a full-time job. I wouldn't be making any progress, so I was very interested in finishing up in my fourth year. I spent a summer trying to figure out a thesis on ethical relativism and came up without much to say. And then I got a further idea of how to develop the arguments about the non-equivalence of rule utilitarianism and act utilitarianism.

And there was also a book by a man named D. H. Hodgson (Hodgson 1967), who was an Australian judge, and he was arguing that ideally rational act utilitarians would not be able to make and keep agreements, because the motivation to keep the agreement depends on the expectations that others will be keeping it. I got interested in that. I don't think I'd ever heard of a Nash equilibrium at that point. David Lewis, though, was writing his book *Convention* (Lewis 1969), drawing on game theory to produce a theory of meaning. Lewis had worked for the Rand Corporation, I think, and he knew a lot about these things. So I developed another part of the thesis that addressed Hodgson, exploring what kinds of agreements people would keep if they could establish the expectation that the agreement would be kept. So that made a dissertation. Those weren't the subject matters I expected to be pursuing later, but the thesis did get me through without having to teach Hum 5.

MA: I'm struck by the fact that so many people, great minds, you and others, worked on this problem of the divergence or convergence of act and rule utilitarianism and the related question of coordination among utilitarians. As you mentioned Lewis did so, and of course David Lyons (Lyons 1965), Hodgson—and I believe your colleague Don Regan, who now teaches at Michigan Law School, also wrote a book on a similar topic (Regan 1980). And then of course, Derek Parfit later in the first part of *Reasons and Persons* (Parfit 1984) talks about this and I'm struck by the fact that there's not been subsequent scholarship; it sort of dies out. I wonder, what are your thoughts about that? Is it that so much attention then shifted to Rawls; rather than looking at the different variants of utilitarianism, the focus after Rawls is on the debate between justice and utilitarianism or consequentialism? Anyway, it is striking to me that there was so much work on this (the divergence or convergence of act and rule utilitarianism), including yours, between, 1950 and the mid-1980s and since then there seems to be less scholarship.

AG: Yes, that's a good question and I don't entirely know what to say. But it seems to be that what you suggest is right: There is a sort of vast culture shift from thinking that utilitarianism basically had it right and one had to work out some problems, to thinking that the right view was somehow contained in Kant and that utilitarianism had it all wrong. I guess I was never convinced that the truth was all contained in Kant. Rawls had a tremendous influence on this shift, I think.

JW: Speaking of Rawls, we would like to hear something about your experience having Rawls as your thesis supervisor. While you both address issues related to how a community should coordinate activities so as to realize the benefits of cooperation on reasonable terms, you were considering broadly utilitarian principles, whereas Rawls rejects utilitarianism in favor of a form of liberal egalitarianism. Does this difference in views affect how you interacted? Do you think that having to defend your ideas to Rawls helped you to refine and sharpen your arguments?

AG: Well, we interacted on working out his views and implications of it. He wasn't very interested in the subject of my thesis and so the advice amounted to: "Yes, I read it. It's very good. Some parts of it are a little obscure." "Oh, what's obscure? I'll try working on that." And then he gave an example, so I rewrote that part. He claimed I didn't need much supervision and certainly I didn't get much supervision.

JW: Also, during that period, you have recounted that you first came across Ken Arrow's *Social Choice and Individual Values* (Arrow 1951) in the Harvard Philosophy department's Robbins Library.

AG: Yes. It's a good thing that it has a bright blue cover.

JW: Do you recall what your reaction to reading the book was?

AG: Oh, I was just amazed and very puzzled. And then when I heard that Arrow was actually coming to Harvard and that he was collaborating with Rawls on a seminar with a young economist I'd never heard of named Amartya Sen, I was very excited, and I told Rawls I wanted to take it. He said, "Oh that's just for Economics graduate students." But I insisted—and the seminar was of course an utterly amazing experience. Howard Raiffa was there, Franklin Fisher, his co-author Jerome Rothenberg, and a young man people said was national bridge champion—Richard Zeckhauser. I sort of sat there gaping and thinking, "I've never witnessed such intelligence all gathered in one room before."

JW: In that course, Sen circulated drafts of some of the chapters of *Collective Choice and Social Welfare* (Sen 1970a). Other than Arrow's book, had you read anything else in social choice theory before this? How did Sen's book influence you?

AG: Sen's book was sort of encyclopedic. I don't think it transformed my view of the subject, but it certainly had things worked out marvelously. But I was still focused on what are we to make of Arrow's Theorem. That's what I—well, you've read the seminar paper that I ended up writing.

JW: O.K. Well, let me ask about that. The term paper you wrote for that course won the Goldsmith prize that year for the best paper in an Economics course or seminar and it's recently been published in *Economics and Philosophy* (Gibbard 2014b). In it, you established your well-known oligarchy theorem: If Arrow's collective rationality condition is weakened to quasi-transitivity while maintaining the rest of his axioms, then the social choice procedure is oligarchic, what you call a "liberum veto oligarchy." What motivated you to consider weakening Arrow's transitivity assumption and why did you use the term "liberum veto oligarchy?"

AG: Well, Sen had devoted a session to advocating weakening transitivity to what he called quasi-transitivity. So, I don't remember how I discovered that that wasn't going to be much help, but in the seminar paper I was addressing Sen. I'd read a lot of history growing up, including Polish history, so the liberum veto was part of my background knowledge and seemed like the right term to adopt for it.

JW: When you subsequently rewrote this part of your term paper around 1970 while an Assistant Professor of Philosophy at the University of Chicago, you recast it more formally as a response to a paper by Schick (1969) that argued that Arrow's Impossibility Theorem is not particularly troubling because the transitivity of social preferences is, in his words, "untenable." Why did you rewrite your theorem in this way and why did you not publish the paper then? It was subsequently published quite recently in the *Review of Economic Design* (Gibbard 2014a).

AG: Well, the problem I had with the seminar paper was that I was worried there was nothing to respond to that was in print. Sen's lecture on the subject had been, as far as I knew, a lecture to the seminar. So I didn't quite know how to place the significance of the theorem. I guess I shouldn't have been so diffident, but when Fred Schick's paper came out, then there was a clear target. So my theorem spoke to something that someone had actually advocated in print.

MA: Let me just follow up quickly on transitivity. As you probably know, there's been more recently a debate in moral philosophy about transitivity. Larry Temkin has argued for giving up transitivity in his big new book *Rethinking the Good* (Temkin 2012). He argues for giving up not just transitivity of indifference, but transitivity of strict preference. On the other hand, people like John Broome (Broome 2004) have said that transitivity is just analytic in "betterness." Do you have a view about that? Whether transitivity is just sort of built into consequentialist thinking or not?

AG: I think my view would be much more Broome's. I haven't worked afresh on it, but it's always seemed to me that the transitivity of "better than" is obvious.

JW: My understanding is that you actually never submitted that paper. Can you tell us why?

AG: Oh, well, Hugo Sonnenschein sent a paper, his joint paper with Andreu Mas-Colell (Mas-Colell and Sonnenschein 1972). And then there was another paper by Guha (1972). So the result had been proved and there didn't seem to be a lot of point in duplicating those proofs.

JW: In your Harvard term paper, you also consider the implications of dropping Arrow's independence axiom. Using a version of your well-known Edwin, Angelina, and the judge example that subsequently appeared in your 1974 *Journal of Economic Theory* article (Gibbard 1974), you showed that the Pareto condition is incompatible with a particular kind of liberal right—the right of two people to get married should they want to. As in Sen's Impossibility of a Paretian Liberal Theorem (Sen 1970b), a right links individual and social preferences by requiring the social preferences over a pair of alternatives to coincide with an individual's preference when it lies within his private sphere. Although Sen had presented his Paretian liberalism theorem to the Econometric Society before going to Harvard, am I correct in understanding that he did not discuss this result or circulate that part of his book in the seminar and that you were unaware of it when you wrote your term paper? Also, what led you to consider modeling rights in a social choice framework?

AG: Well, if I can remember this, I wasn't aware of Sen's paper, although maybe I should have been. But as I seem to remember, after I presented my paper, he gave me a paper. I was somewhat confused about its status and didn't get that straight until much, much later. So I think when I presented the theorem to the seminar, I didn't know that Sen had done that.

JW: What motivated you to try and model rights using the social choice framework that Arrow had developed?

AG: I don't remember in specific terms. It did seem that one could use the social choice framework to talk about rights, and that then important and interesting things happened, including things that bore on what the significance of the Arrow Theorem was and wasn't. I don't remember having an epiphany that rights could be treated this way, but it certainly seemed that they could be.

JW: In that term paper, you suggested that the way you modeled rights is not completely satisfactory and raised the possibility, but didn't explore, that it may be desirable to waive a right. These are issues that you explored in your 1974 *Journal* *of Economic Theory* article (Gibbard 1974). In that article, you modeled a set of alternatives as the Cartesian product of the personal features that are available for each individual. Using this framework, you showed that a natural way of assigning rights is inconsistent, without any appeal to a Pareto principle. You also allowed individuals to waive rights and showed that this formulation of rights is both self-consistent and compatible with the Pareto principle. What advantages does this way that you modeled alternatives and rights-assignments have over the way that they were modeled by Sen?

AG: Well, Sen was trying to show that the Pareto principle was defective, and his argument was that the Pareto principle was inconsistent with assignments of rights. That seemed very puzzling to me, since even though, as Sen formulated things, the awarding of rights was not inconsistent without bringing in the Pareto principle, it seemed that the motivation for thinking that people had rights was a principle that was going to turn out to be inconsistent by itself. So I didn't understand how what Sen was doing was going to discredit the Pareto principle. I think Sen has always stuck to the view that it does, and I'm as puzzled as I ever was.

JW: In *Anarchy, State, and Utopia* (Nozick 1974), Robert Nozick took issue with the way that both you and Sen model rights. In your approach, rights are linked to preferences—when an individual has a preference over a pair of alternatives in his private sphere, then that is the social ranking unless the right is waived. Nozick instead argued that by exercising their rights, individuals pick some features of an alternative and that social choice considerations only apply when determining any remaining features. How would you respond to Nozick? Do you still think that your way of formalizing rights in a social choice framework is a good way?

AG: Social choice theory concerns what histories are morally O.K. in the ways that what happens depends on people's preferences. And that seems like a question one can pose when one thinks about rights. And if you pose it that way, then the social choice theory apparatus applies.

JW: How do you respond to Nozick's way of thinking about rights, which is more game theoretic? In his approach, each person independently chooses within his or her own sphere and then there may be some features left over, and that's the only part of the decision that's left to social choice considerations.

AG: Well, I don't see why one shouldn't apply the social choice framework anywhere where your assumptions are good assumptions, and so I don't think of the two ways of doing things as incompatible.

JW: Among economists, you are best known for what is now called the Gibbard–Satterthwaite Theorem (Gibbard 1973; Satterthwaite 1975). This theorem shows that any social choice function that maps profiles of individual preference orderings into a single choice from the set of available alternatives must be dictatorial if (1) the domain is unrestricted, (2) there are at least three alternatives, and (3) the social choice function is strategy-proof—that is, nobody can ever gain from misreporting his preferences. What prompted you to work on this problem? When you started working on it, did you expect that the conclusions would be so nihilistic? And what challenges did you have to overcome to establish your theorem?

AG: As I remember it, Arrow, when we were discussing—"we" is the wrong term, because I don't think I said more than two things the whole seminar until I presented my paper—but when the seminar was discussing Independence of Irrelevant Alternatives, which I thought was what was problematic in the assumptions of the Arrow Theorem, Arrow said something about it being equivalent to strategy-proofness. And so the next year, I had started teaching at the University of Chicago and they had assigned me a social choice theory seminar which had, I think, two registered students and two auditors, including Bernie Grofman. Bernie was quite an auditor, and I was probably learning more from him than he was from me. Preparing for one session, I thought, "Well, I should work in the part about being immune to strategic manipulation being equivalent to IIA." I assumed it would take about five minutes to think it through. "So let's see, I'd better prepare this class. How does that work?" And then I got stuck and I was stuck for several weeks. And then I can remember the afternoon when I was visiting my fiancée Mary in Urbana and we decided to spend a couple of hours sitting at her card table getting some work done. The basic idea of how to prove what I was looking for occurred to me, and my mind started racing.

MA: Let me just add a quick follow up. You mentioned in passing there that you found Arrow's Independence of Irrelevant Alternatives to be problematic. Can you expand a bit on that? Why do you think that the axiom is problematic? This has been a big issue in social choice theory.

AG: Well, first, of course there's confusion in Arrow about what's governed by the Independence of Irrelevant Alternatives as he formulates it, and what is a matter of social choice being guided by a preference function that is fixed independently of what's feasible and what isn't. Once we get all that straight, the reason to reject IIA as Arrow formulates it is basically the one I gave in the seminar paper, that one's ordering of non-feasible alternatives, along with feasible alternatives, gives some indication of the strength of one's preferences. My example, I think, was Patrick Henry, "Give me liberty or give me death," which seems to indicate a strong preference for liberty. Arrow was a thorough ordinalist. In another version of the theorem, you have cardinal utility scales that aren't interpersonally comparable, and only ratios of preference strength matter. But either way, people's preferences involving infeasible alternatives provide some indication of preference strength. And so I thought the kind of information you can glean from preferences involving infeasible alternatives was relevant to the moral weight that a preference should bear.

JW: *Econometrica* was not where you first submitted your article. Please tell us something about what happened when you first sent your paper off for publication.

AG: Well, there was a new journal called *Theory and Decision* and I submitted it to that. And I was still puzzled that such a simple theorem didn't have a simple proof, and I included a covering note to the effect, "Maybe this is trivial, but I certainly can't see that it's trivial." I got a response from one Editor that said, if I remember, "It is trivial. Arrow's Theorem says that any social welfare function is imposed or dictatorial. And so it is imposed or dictatorial or manipulable, and it is imposed, dictatorial, or a hot dog." So it was a completely fallacious, a complete misstatement of Arrow's Theorem. And then I got a two-page letter from the other Co-Editor who said the same thing in two long pages. So I decided I'd better try to send it to a competent place, and I sent it to *Econometrica*. I think I got indications later that Arrow and Sen had been the referees, and so there it was certainly competently refereed, to say the least.

JW: In order to establish your theorem, you introduced the concept of a game form, which specifies an outcome as a function of the strategies chosen by the individuals. In terms of a game form, strategy-proofness is then the requirement that truth-telling is a dominant strategy in the direct mechanism in which a strategy is a preference. Did you realize when you wrote your article that game forms would be a useful tool for analyzing other problems involving asymmetric information?

AG: Well, it seemed to me they were a useful tool. The first version of the theorem I proved for myself was the one that just talks about truth-telling and expressing one's preferences. It was later on, maybe in the summer of 1970 which I spent hanging around Cambridge University, that the game-form version occurred to me. I loved its generality. In lots of systems of voting, you aren't voting by doing something that constitutes making a statement of what your preference ordering over the alternatives is. At some point, it occurred to me that there's a more general way of putting the result. The game-form formulation has the advantage that it doesn't actually talk about truth-telling. It just says that for each preference ordering you might have, there's a dominant strategy, whether or not it takes the form of reporting your preferences truthfully. So that seemed obviously a much more general way to formulate things that would apply to any social mechanism whatsoever, and I was pleased with that—although of course it makes the statements of proofs more complicated.

JW: Mark Satterthwaite independently established a version of this theorem. His constructive proof first appeared in his Ph.D. thesis. In the published version (Satterthwaite 1975), in addition to providing this constructive proof, he also shows that the Gibbard–Satterthwaite Theorem follows from Arrow's Theorem and vice versa. In his article, Satterthwaite said that he got the insight for this correspondence result from reading your article. When did you first learn of Satterthwaite's work and did you have any inkling that your theorem and that of Arrow were so tightly linked before you read his article?

AG: My own proof proceeded via the Arrow Theorem, and that made for a tight link. A beautiful thing about Mark's proofs is that he got the strategy-proofness theorem independently of the Arrow Theorem, and then used all this to draw connections that I had not myself drawn. I remember getting Mark's letter when I was at Chicago—I think it was a letter, or perhaps he phoned. As I say, I didn't have a direct proof of the manipulability theorem—my proof went by way of Arrow's Theorem—and he did have a direct proof. I think he said that he had submitted it and

Sen had pointed out that I had proved the result, but then Mark did quite ingenious things with the implications of having a direct proof of the theorem.

JW: I guess a better way to phrase the second part of my question was that he seemed to sharpen the connection between Arrow's Theorem and the Gibbard–Satterthwaite Theorem, in effect showing that you can go back and forth between the two.

AG: Yes, right.

JW: In your article, you emphasize that the social choice is required to be deterministic. You later allow for the social choice to be a probability distribution over the alternatives (Gibbard 1977). You show that strategy-proofness implies that the social choice function must be a probability mixture of functions that are either dictatorial or only choose from two fixed alternatives. Do you regard such social choice procedures as being satisfactory? More generally, do you think that lotteries have a useful role to play in social decision-making?

AG: Well, I think obviously the narrow sorts of schemes that turn out to be non-manipulable are not satisfactory, and one should tolerate some manipulability rather than have such blatantly unsatisfactory ways of choosing things.

JW: Lotteries are sometimes used when there are indivisibilities for which there is no way to split things up in some fair way between people. Do you think that that's a reasonable way to make decisions in such circumstances?

AG: Yes, I don't think that all possible uses of lotteries in social choice are unacceptable. Sometimes, a lottery is a reasonable way of resolving ties and the like. But I don't think that the indeterministic version of the theorem says, "Ah, there's a satisfactory way, there's a strategy-proof satisfactory way of making social choices after all." The kinds of indeterministic schemes that the theorem show can be strategy-proof are blatantly defective, in my view. And so you're still faced with trade-offs among desirable features, and in light of these trade-offs, full strategy-proofness will need to go.

JW: From what you said earlier about Independence, do you also have concerns that strategy-proofness is too strong of a requirement?

AG: We learn it can be had only at too high a price, so it's clearly too strong a requirement in that sense. It seemed like a nice feature in isolation, but the price for it is prohibitive. It seems to me that that's what the theorems show. After the impossibility theorems, people worked on what to do in the face of the finding that strategy-proofness can come only at an unacceptable cost. I didn't master the results that emerged, but that seems to me to be the right approach.

JW: Rawls' Difference Principle requires that social institutions be designed so as to maximize the prospects of the least advantaged as measured by an index of primary goods (Rawls 1971). This index is supposed to be a measure of an individual's command over basic social goods like income, rights, and opportunities. This raises the problem of how such an index can be constructed. In your 1979 *Theory and Decision* article (Gibbard 1979), you tackled this problem using the formalism of social choice theory and welfare economics. You reformulated Rawls' Difference Principle in terms of a possibly partial ordering of the opportunities facing an individual. Among other results, you showed that if individuals have preferences over the relevant goods that satisfy the standard assumptions used in microeconomics, then your version of the Difference Principle is incompatible with the Pareto principle. What lessons for Rawls' theory do you think follow from this result?

AG: I haven't kept that paper well in mind over the decades, but I keep hoping that it offers the definitive interpretation of Rawls' Difference Principle, pursuing his motivations and avoiding definite mistakes. Rawls' system has a number of layers, but the crucial ones, to my mind, are these: First, the Original Position with the information and motives of parties who are choosing principles of justice to be accepted and realized in their society, and second, those principles themselves—which include the Difference Principle. I myself think that this system is at least consistent with a kind of indirect utilitarianism. Rawls, as far as I know, never precisely denied this; what he denied is that his system amounts to direct utilitarianism. I think that John Harsanyi (Harsanyi 1955) was right that the standard of what would be chosen in the Original Position is utilitarian, and I think he held too that the principles that meet this standard need not amount to direct utilitarianism. Direct utilitarianism of course must entail the Pareto principle, but it doesn't follow that the principles chosen as a public conception of justice will conform to the Pareto principle. My paper is about that.

JW: About primary goods? About the index of primary goods?

AG: Yes, that's what I was trying to understand in that paper. Rawls' central idea is appealing: that it's the primary goods that society is responsible for making available, and what you do with them is your responsibility. But then that does raise a big question of how it's determined what the primary goods are, how to index bundles of primary goods. You have to answer those questions to make the theory say something definite.

JW: I think what your work showed was that that's problematic. I guess that the indexing number problem has never ever been really satisfactorily dealt with.

AG: I'm not sure whether it has. I try in that paper to propose how we might get something that does the job of such an index. My more general view about Rawls is that the arguments he gives for the Difference Principle really say that the situation in the Original Position is such that the Difference Principle and what utilitarianism recommends are not substantially different. Rawls' view thus isn't really an incompatible alternative to rule utilitarianism, but amounts to saying how rule utilitarianism works out as applied to questions of economic justice.

MA: Let me just quickly follow up there. There's a famous dispute between Harsanyi's conception of the veil, which says that the veil involves equal probabilities of ending up as any person, and the Rawlsian conception, which says that it's a decision under ignorance without probability. Rawls uses his conception of the veil and a conception of how to choose under ignorance in order to get to the Difference Principle. So, are you suggesting that Harsanyi's conception of the veil was better than Rawls'? I believe that Parfit in *On What Matters* (Parfit 2011) takes Harsanyi's side in that dispute. Parfit's view, I believe, is that if we're trying to model impartiality, the stipulation of equal probability as opposed to ignorance is a better model.

AG: It's puzzling whether ruling out probabilities the way Rawls does makes a real difference. What parties in the Rawls' Original Position are ignorant of, in Rawls' scheme, is what their society is like—beyond that the circumstances of justice obtain, including moderate scarcity. If parties knew exactly what their society was like, including the proportion of people with each set of relevant characteristics, then clearly they could use probabilistic reasoning. They would each take it that one had roughly an even chance of being female or male, and that one has a 1% chance of being in the top 1% and a 20% chance of being in the bottom 20%. So if they could specify, for each way their society might turn out to be, such things as which distribution of income was to be realized, then numbers would very much count, and they would specify whatever setup gives a person the best prospects. Parties could specify a function from what the society is like, with its economic possibilities, to what the economic institutions should be like.

But that isn't how Rawls sets things up. What the parties choose is not a standard that the economic setup is to satisfy, but a public conception of justice, a conception that the society is not only to match but to adhere to and be motivated by. The question the parties face is what the public conception of justice is to be, and this choice is made under non-probabilistic uncertainty about what their society is like and hence what the consequences of the choice of a public conception of justice will be. Rawls thought that his Difference Principle would be chosen as a part of the public conception of justice.

It's important to note also, though, that his rationale for this includes what amounts to an argument that under circumstances that the parties know to obtain, instituting the Difference Principle will maximize expected utility. The parties, he says, care little about the gains they could make above what the Difference Principle accords them, but alternative principles allow outcomes they abhor. In light of all that, an indirect utilitarian would choose the Difference Principle just as surely as would parties to the Original Position as Rawls specifies it.

JW: Is what's relevant that Rawls is stripping someone behind the veil from information which you think is relevant for making a decision?

AG: Yes, he strips them of information that would be relevant if they had it. He strips them of information as to what the economic possibilities are like that they could realize by a choice of principles to govern economic arrangements. But even more central to what he does is this: What the parties choose in the Original Position is not just the standards that economic arrangements are to meet, but what conception of justice shall be public among them. What conception of justice will be publicly accepted makes a crucial difference as to what economic arrangements will be realized and whether they will be stable.

Rawls misses an argument I just gave, but sets things up in such a way that the upshot of this argument is inconclusive. The argument he misses is this: That if in the Original Position, what's to be chosen are the standards that economic arrangements are to meet, with the assurance that whatever they choose will be implemented, then Rawls' system will amount to an indirect form of utilitarianism. They won't know what their society is like and what its economic possibilities are, but they can choose a complicated standard that, in effect, chooses for each way their society might be whatever economic arrangements would maximize people's economic prospects.

But as Rawls sets things up, they have to choose a conception of justice that is to be publicly accepted no matter what the economic possibilities turn out to be like. One candidate for playing this role goes as follows: "Give those in the worst-off starting positions the best prospects possible." That's the candidate that Rawls argues they would choose. Other candidates would be more directly utilitarian: "Realize whatever economic arrangements would give the representative person the best economic prospects." Rawls argues, in effect, that absent information about what one's society's economic possibilities are, parties to the Original Position will prefer the Difference Principle. But as I have been suggesting, how all this might depend on features of the Original Position as Rawls stipulates them is a tricky matter. And beyond that, there's the more basic question of why the Original Position should have those features. Giving it these features is supposed to yield the outcomes that Rawls wants, to be sure, but more importantly, it's supposed to illuminate the kind of rationale that ascribing these features to the Original Position might have.

MA: So is it fair to say that you strip away information from a contractor about who she is? But then the question is: For all the different possible people, does the contractor assign those equal probabilities or does she choose without any probabilities at all? Rawls seems to take the latter approach and from there he gets to the Difference Principle by virtue of the view that choosing under complete ignorance you should maximize the worst possible position. But again, if one were to strip away information about your identity, but then assign equal probabilities to the different possible societies and your being one or another person in those societies, one could then just apply expected utility theory to that choice.

AG: I'm saying that as Rawls sets things up, the prime uncertainty that the parties can't have subjective probabilities about concerns what their society is like, with its economic possibilities. If the parties knew all that, they could ascribe probabilities easily. If they knew everything there was to know about the economic circumstances of the bottom 20%, then trivially, they would have a 20% subjective probability of being in those circumstances. But they don't know such things as what the distribution of income and opportunities is for those in the bottom 20% of starting positions. So I agree with Rawls that it doesn't make any sense to ask what's the probability that this society is twenty-first-century America or ancient Greece or something else; in the Original Position, I take it I am in any of the vastly different possible societies where the circumstances of justice apply. We can agree with Rawls that it doesn't make any sense to say that people in the Original Position assign probabilities to being in certain kinds of societies.

I was also insisting that given any particular way society might be apart from who's who, the thing to choose is to maximize expected utility. So, there's a sure-thing principle that says: "Even without knowledge of what your society is like, maximize expected utility." But that's not an option that parties to the Original Position can choose. They have to choose a conception of justice to be accepted in their society, and they have to make this choice in ignorance of what their society is like in its economic possibilities.

MA: O.K, I'm going to take over asking questions at this point. Welfarism is the view that social alternatives should be evaluated solely in terms of the individual well-beings associated with them. The Difference Principle, with its focus on opportunities, is non-welfarist. In a provocative 2001 article, Kaplow and Shavell (2001) argued that any non-welfarist principle violates the Pareto principle. Faced with such a conflict, what do you think should be abandoned? Or does it depend on the non-welfarist principle that's being considered?

AG: My sympathies are with the Pareto principle and welfarism. I talked a long time ago about the "intrinsic reward" of a life (Gibbard 1986), and setting up whatever system gives the best prospects for intrinsic rewards of people's lives seems to me to be a good thing to favor. That will require satisfying the Pareto principle, couched in terms not of preferences but of the intrinsic reward of lives. But the conception of justice whose public acceptance would most foster intrinsic reward needn't satisfy the Pareto principle in any form.

JW: Faced with such conflicts, we're interested in how you would resolve the impasse of the incompatibility.

AG: It's a difference between direct utilitarianism and various forms of indirect utilitarianism. Indirect utilitarianism evaluates not acts or economies directly, but such things as possible ethoses for moral standards or standards of economic justice. Such evaluations will satisfy a kind of Pareto principle: If one ethos yields prospects that are better for someone and worse for no one, then that ethos is preferable. But the dicta that comprise the ethos don't have to include the Pareto principle and don't have to be consistent with the Pareto principle.

That said, one feature of utilitarianism that worries me is that you can try to base things on reciprocity, and then utilitarianism says that even if others aren't reciprocating, one must take their good equally into account except as taking their good into account produces incentives that aren't good-promoting. So I do have that worry about utilitarianism that it doesn't sufficiently cater to demands of reciprocity. I'm worried about Rawls' contention, which seems right, that motives of fair reciprocity are stronger than motives of unconditional altruism. Still, if you can get everybody to cooperate in a scheme, if I have my choice of what scheme everybody will cooperate in, I think we should go for one that produces the highest expected intrinsic reward of people's lives.

MA: Let me ask a different question and then we'll come back to this issue of the strains of commitment in utilitarianism. Your analysis of Rawls' Difference

Principle employs a formal economic model in order to address normative questions. In a 1978 article with Hal Varian (Gibbard and Varian 1978), you considered how economic models can be used to help explain features of the real world. You argued that the usefulness of an economic model for explanatory purposes when applied to a real-world situation is due to the assumptions being either sufficiently close to the truth or a caricature of the key features of a situation with the conclusions being robust to variations in the caricature. What relevance does this analysis of the value of economic models for positive economics have for normative purposes?

AG: Well, let me say first that that paper came from my being asked to give an American Philosophical Association paper on Philosophy of Economics, and I thought I didn't really have the subject well enough under my command to do it all by myself, and Hal agreed to collaborate on it. So, my motivation in writing it didn't have a lot to do with normative questions. But I kept finding that philosophers thought that economic models were ridiculous because they had assumptions that obviously weren't true of the world. And so I was just trying to educate philosophers about how economic theorists think of their models that economic theorists aren't the idiots that philosophers were assuming that they were. Of course, we use models a lot for normative purposes, as with the model of the Original Position or Harsanyi's model, but that wasn't what our paper was about. It was really about positive applications of economic models.

JW: In many normative analyses, simple models drawn from economics are used to work out ideas or test ideas, and they involve making unrealistic assumptions. Is this a shortcoming? Or are there some kinds of rationales analogous to what you did for positive economics that justify the use of simple models of production and distribution to, say, test a theory of distributive justice?

AG: I would think the latter. I don't have elaborately worked out views on that. But in order to scrutinize the things you might want to say about messy reality, you need to have some tractable way of thinking about matters, and thinking that messy reality has some important relation to a tractable model is about the only strategy of inquiry that is available. I would think one should always keep one's eye on what the tractable oversimplified model might have to do with reality, but—well, it's pretty much the way Arrow talks about positive models. You say, "Well if reality were this way, here's what we could conclude." And then the next question is, "Are there features of reality that make a difference and tell us we shouldn't conclude that?" So, say a normative argument such as an invisible hand argument for laissez-faire abstracts from externalities, and it also abstracts from information, the way the Arrow–Debreu model abstracts from information. If things were the way that the Arrow–Debreu model (Arrow and Debreu 1954) supposes, then laissez-faire would be Pareto efficient. That leaves a number of important things to note.

First, of course, the Pareto principle tells you nothing about distribution, and it has always seemed to me that the old arguments that utilitarianism will have an egalitarian tendency are good arguments, despite being despised from the 1930s on. And so, laissez-faire is not the thing to go for. It's true that feasible ways of trying to increase equality will violate a version of the Pareto principle, but that's a cost to be born. And it's a version of the Pareto principle that basically talks about the ways you could organize society with perfect knowledge and unlimited calculating power. A Pareto principle that abstracts from that may have something to do with reasonable policy choice, but not a lot.

And then, of course, the story that standard economics tells says that we have ignored externalities, and we'd better bring those in. We've ignored information, and later on, people studied intensively systematic ways of bringing information in. And besides all that, there are the ways in which we can't be perfectly rational expected utility maximizers, so that nudges and things like that may come in further on in the discussion. So, I would think that the idealized models for normative purposes, as well as positive purposes, tell us what would hold if certain things were true, and the relevance of what would hold if certain things were true to what actually holds is going to be a complicated question.

MA: A different way in which this question might be relevant depends upon whether utilitarianism is simply a criterion of goodness or a decision procedure. Some people say that utilitarianism simply specifies the conditions under which an act or a possible world would be better, but is not necessarily operational as a decision procedure. On that view, models might not have a lot of direct relevance to morality. A different construal of utilitarianism says, "maximizing total well-being should be a direct decision procedure." In which case, models are very important because humans, at least, can't hold representations of whole possible worlds in their heads. Humans will need to use models of those worlds, and then we have a question whether those models are adequate. So let me ask, on your view is utilitarianism simply a criterion of goodness, or does it have more direct decisional role?

AG: Well, I think I would take the same sort of view as R. M. Hare took, that utilitarianism would be fine for what he calls archangels, but not as a full ethos for society (Hare 1980). For the latter, the question is more, "What ethos would produce the best results?" And the ethos that would produce the best results probably is not asking one to calculate each decision in a utilitarian way. As I said before, I think that the question Rawls asks amounts to this too: His question is what would most appeal to parties in the Original Position as the ethos to have as our public conception of justice.

MA: O.K. So let me ask you about causal decision theory. Leonard Savage (Savage 1954) modeled decision-making under uncertainty as a ranking of acts, where an act maps states of the world onto outcomes. For Savage, states are act independent, but this has been generalized to allow for act dependence. Few economists seem to be aware that there's an extensive philosophy literature dealing with an alternative approach to decision-making under uncertainty known as causal decision theory in which judgments of causality play an important role. Your 1978 article with William Harper (Gibbard and Harper 1978) is one of the seminal papers in this literature. For those unfamiliar with causal decision theory, please provide a brief explanation of what this theory is and why it is important. In what kinds of circumstances do you think that it is safe for economists to ignore issues of causality in decision theory?

AG: Well, first, a couple of preliminary things. One is that the Gibbard–Harper article was drawing very heavily on a discussion involving Robert Stalnaker and David Lewis (Stalnaker 1981), and we buried this acknowledgment too far in the footnotes. It was Stalnaker whose idea we were mostly developing. We thought of ourselves as writing a paper about a proposal by Richard Jeffrey (Jeffrey 1965), but in order to fill in the background, we had to explain Stalnaker's unpublished proposal. So in the last reprinting (Gibbard and Harper 1981), we started out with something like, "this paper develops a proposal by Robert Stalnaker." Second, I don't think what we're doing is an alternative to Savage. Savage requires act-independence, but then there are two kinds of independence that one might demand. One is probabilistic independence and the other is known as causal independence. I've talked to various decision theorists about this, and the people I've talked to seem to split about 50/50 about which Savage intended. I guess I tend to think that Savage must have intended what we're advocating, what came to be called causal decision theory. I would have thought that Savage wants the agent to be sure that which state obtains doesn't depend causally on what one does.

Stalnaker was responding to a theorem of David Lewis (Lewis 1976); this was at a symposium that I wasn't at, but that Harper, Lewis, and Stalnaker were, and that Harper told me and others about. Stalnaker had thought that there was a conditional if-then operator that produces a conditional proposition as a function of two propositions. So on his view, "If A, then B" is a proposition, true for some ways the world might have been and false for others. He proposed that this propositional operator had the nice feature that the probability of "If A, then B" is the corresponding conditional probability, the probability of B given A. Lewis, at that session that Harper told me about, showed that such a thing could hold in general only for excessively trivial cases. And Stalnaker said, "Well, suppose, suppose R. A. Fisher had been right when he conjectured—well, I guess Fisher just said it was a possibility for all the statistics tell us—that the correlation between smoking and cancer is not because smoking tends to cause cancer but because there's a common cause of both of them." [See Pearl (2000) for a discussion of Fisher's views.] Suppose there's a genetic factor that leads both to smoking and to cancer, so the news that I'm about to smoke becomes an indication to me that I have this genetic factor and that I'm going to have cancer. So smoking, the story is, doesn't cause cancer, but it's an indication that I'll get cancer.

One interpretation of Savage is that he requires that the states be epistemically independent of the acts—that, for instance, the epistemic probability of one's having this genetic factor be independent of whether one smokes. Such epistemic act-independence produces the kind of decision theory that says you shouldn't smoke because smoking is an indication of something you strongly and reasonably don't want. What I like to call "instrumental expected utility theory" says, in contrast, that since in this fantasy of Fisher's smoking doesn't in any way tend to cause cancer, it makes sense to go ahead and smoke if you enjoy it.

MA: So would that in fact be your view? That is, your view would be the instrumental or causal, as opposed to the evidential, approach, namely that in this scenario where smoking doesn't cause cancer and simply indicates cancer, if you like smoking, you should smoke?

AG: Yes—and this applies to Calvinism also. Some people say that Calvinists think that because leading a godly life is an indication that one's among the elect, wanting to get to heaven is ground for leading a godly life, even at some sacrifice of earthly goods. My sister, who's a historian of that era, says, "No, the idea is that the elect really care intrinsically about godliness. So, Calvinists are said by some people to be epistemic utility maximizers but, according to my sister, they were actually instrumental expected utility maximizers.

MA: I see. But surely it's the case that economists, at least implicitly, are causal or instrumental and not evidential. In the case you are describing, I imagine economists would say that you should smoke. And in the classic case of Newcomb's paradox (Nozick 1969), I imagine economists would say, "Take both boxes, not one," even though your taking both boxes is evidence that the predictor has not put the big prize in one box. And in the case of twins playing a prisoner's dilemma, economists are surely going to say that you should rat even though that's evidence that your twin is going to rat. So, if this is right, then economists really owe you a great debt. And it also seems to me that economists, although they use the Savage framework or the Anscombe–Aumann framework (Anscombe and Aumann 1963), which is a variation, are very sloppy about ensuring that the states are causally independent from the acts.

AG: Yes, apart from not knowing about the Anscombe–Aumann framework, I'm 100% with you on all that. I do know about the Savage framework, and it, I maintain, should be read instrumentally, not epistemically.

MA: All right. You are best known to philosophers for your work on meta-ethics. Let us now turn to that subject and how it relates to your approach to utilitarianism. You characterize normative questions as planning questions, that is, questions about how one ought to act, what one ought to believe, and what one ought to feel. Please explain why you view normative "oughts" in terms of plans. What are the advantages of this perspective over competing views?

AG: Well, first, it has a disadvantage, namely that I have to use the word "plan" in a way that differs from the ordinary use in lots of ways, so . . .

JW: But it's a use that economists would be sympathetic with.

AG: So, to answer a normative question about what, say, Caesar ought to have done when he arrived at the Rubicon, I ask myself, "Well, suppose I'm in Caesar's situation." Under that hypothesis, I ask myself what to do. And if my hypothetical plan is not to cross, then that amounts to thinking that the rational thing for Caesar to have done was not to cross.

MA: Are you concerned about the problem of changing your personal identity, that it's metaphysically impossible for Allan Gibbard to be Caesar? To the extent that you want to plan not just for the things you're going to do, but for the contingency of you being someone else, how do you get around these issues about the necessity of personal identity?

AG: Well, it's metaphysically impossible for me to be Caesar, but it's not strictly impossible epistemically for me to get conclusive evidence that I'm Caesar. It's far-fetched, but when people are mixed up in babyhood or something like that, then they have coherent beliefs about who they are that are epistemically justified but metaphysically impossible. So, it's the epistemic possibility that matters here. I mention this sort of thing in *Meaning and Normativity* (2012, p. 133) and elaborate an entire framework for such questions in Appendix 1 of that book.

JW: To follow up on that, you wrote a paper about interpersonal comparisons in the late 1980s in the Elster–Hylland volume (Gibbard 1986) where you developed in a very nuanced way some of Harsanyi's ideas about the nature of interpersonal comparisons (Harsanyi 1955). Does what you've just said tie in with your earlier arguments about the scrutability of different individuals?

AG: Well, it ties in with the question, "Suppose I'm John Weymark. What to want?" This is an intelligible question, even though it's metaphysically impossible that I be John Weymark.

MA: Let me ask you about the motivation for thinking of "oughts" or an ought statement as the expression of a plan or commitment, as opposed to an assertion of either a natural or a non-natural reality. Can you say a little more about this, again for an audience of social choice theorists who are not familiar with expressivism and the debate between expressivism and competing positions in meta-ethics?

AG: Well, I'm basically taking off from the kinds of arguments that G. E. Moore gave in around 1900 (Moore 1903) for a view that Henry Sidgwick also had (Sidgwick 1874), that no purely naturalistic assertion will have the same meaning as a statement about what I ought to do in a situation. So, if you say that in a prisoner's dilemma with one's twin one ought to rat, then there's no empirical fact for which the settling would be tantamount to deciding what one ought to do in that situation. And so that leaves a puzzle. Moore and various other people say, "Well there's a fact, but it's not a natural fact," and that seems very obscure. So, in the 1930s, people like A. J. Ayer (Ayer 1936) said, "Well, that's because what we have to explain is not what naturalistic thesis is equivalent to saying that one ought to rat, but the state of mind of believing that one ought to rat." And Ayer thought that this state of mind is emotional. But critics say, "The question isn't really what I do feel about it, but what I ought to feel about it." And it seems to me that we do ask ourselves questions about how to feel about things. "Ought I to feel jealous?" Questions like that.

And so what I tried to do was apply Ayer's strategy, not to moral assertions but to normative assertions more generally: What is the state of mind of thinking that one, say, epistemically ought to believe the theory of natural selection? What I've tried saying is that we form something like plans for what to believe given certain kinds of evidence, and what I'm doing with such a statement is expressing having such a plan for what to believe given certain evidence. I could ask, "What was it rational for David Hume to believe about the causes of life?" Hume, I think, was basically a creationist. He thought that we could only see plants and animals as designed to do things crucial to reproduction. Before Darwin that was the rational thing to believe. Hume's most notable discussion of this is in Chap. 12 of his posthumous *Dialogues Concerning Natural Religion* (Hume 1779). How to interpret Hume's intent in that chapter, though, is immensely controversial. When I say that's the rational thing to believe in Hume's evidential circumstance, I'm forming plans for what to believe given Hume's evidence without having Darwin's proposal available.

MA: So you've just explained epistemic norms in terms of plans for belief, and in both *Wise Choices, Apt Feelings* (Gibbard 1990b) and *Thinking How to Live* (Gibbard 2003), you explain moral norms in terms of plans for feelings, that is, plans for feeling guilt or resentment. How do you respond to the objection that you can't control either your beliefs or your feelings; you can only control your decisions? And so, at most we can adopt plans for choice but not plans for belief or plans for feelings?

AG: Well, my conclusions about what it makes sense to believe do have an influence on what I actually believe. So if I am, say, running an experiment and doing a statistical analysis, statisticians always say to plan one's analysis as part of planning the experiment: Planning the analysis is planning what to believe, maybe what degrees of belief to have in various hypotheses under conditions of getting certain results in the experiment. So, you can't believe at will, but after all you can't will at will either. What I will is influenced by my beliefs about what it makes sense to will. And, likewise, what I believe is influenced by my beliefs about what it makes sense to believe.

MA: Let me come back to this issue about emotivism. In your 2006 interview with Alex Voorhoeve (Voorhoeve 2009), you mentioned that as an undergraduate at Swarthmore, you were interested in the status of morality and, quote, "disturbed by emotivism." Can you comment on the difference between expressivism and emotivism, and how in turn they relate to other positions in meta-ethics?

AG: Yes. I'm not actually sure whether the term "expressivism" stems from me or from someone else; I haven't been able to find that out. But there's something that Ayer has in common with Hare and with me, namely a strategy of explaining meanings by characterizing the states of mind that are expressed. So, Ayer explains the meanings of terms like "right" and "wrong" by saying, in effect, that to believe that something is wrong is to have negative feelings toward it (Ayer 1936). People call it the "Yay, Boo Theory," which is meant to deride the theory, but is actually a good way of thinking about its logic. R. M. Hare says that it's not a state of feeling that one's expressing with a moral assertion, but one's preferences for the case of being each person involved seriatum, or better, one's preferences for the case of being me and the case of being you and the case of being John, and the case of being each other person (Hare 1980). That is, if it's really a moral belief, then the conditional preference isn't affected by who's who in the situation. My preference for the case of being you is the same as my preference for the case of being me. Otherwise, my preferences don't constitute a moral conviction. My own view that the state of mind of making a normative judgment "is something like having a plan" takes that same form. So I take the idea of expressivism to be that you explain meanings by explaining the states of mind involved in a way that doesn't involve helping yourself to describing them as the state of mind of believing such-and-such. The philosopher Jason Stanley said to me, "Well, isn't that just what philosophers call functionalism in the philosophy of mind?" And I think he's right.

MA: Let me ask you another question about meta-ethics. Much of meta-ethics has been preoccupied with the so-called Frege–Geach problem (Geach 1964). Can you explain for an audience of social choice theorists not expert in the topic what the Frege–Geach problem is; why many ethicists have found it so difficult; and why your approach is well positioned to address it?

AG: To answer that, it's easiest to think of straightforward emotivism of the Ayer kind, so that saying "Lying is wrong" amounts to saying "Boo for lying." But then Peter Geach says, attributing this to Frege with some justification, that we've got to explain not only sentences like "Lying is wrong," which amounts to "Boo for lying," expressing feelings against lying, but also how a father can argue, "If lying is wrong, then getting your little brother to lie is wrong too. And lying is wrong, therefore, getting your little brother to lie is wrong."

So first, we have to explain a sentence like "If lying is wrong, then getting your little brother to lie is wrong too." And furthermore, we have to explain it in such a way that the inference is valid. We don't transfer between one interpretation of "Lying is wrong" when it's self-standing and a different interpretation when it's part of the conditional, "If lying is wrong, getting your little brother to lie is wrong too." Otherwise, the inference won't be valid; it will be in the form of, "If A, then B; A^*; therefore, B," which isn't a valid inference.

So what I say is that a statement determines a set of states of mind you might be in. A state of mind would include one's factual beliefs and also one's feelings or preferences. So, what the assertion "If lying is wrong, then getting your little brother to lie is wrong" does is exclude states of mind in which you have feelings against lying, but don't have feelings against getting your little brother to lie. If you think of it that way, then all sorts of standard truth-conditional logic just comes out with a different kind of explanation for its validity, and the explanation for the validity applies to prosaically factual beliefs as well as to beliefs that amount to feelings about something.

MA: Can you say a little bit more about the role of plans in rational choice? One version of expected utility theory says that the focus at any point in time should be the choices available to you at that point in time and your information about consequences going forward. It's not rational to stick to a prior plan if a different choice maximizes expected utility going forward. Do you view planning as being more fundamental to rational choice than on this traditional version of expected utility theory that focuses on acts as opposed to plans?

AG: Philosopher Michael Bratman's main lifelong work has been on why plans are important for leading one's life, and individual decisions in isolation won't do the job (Bratman 2014). So he's very interested in when one ought to reopen a question that one had already settled in one's planning. And I guess I would think that we need a kind of indirect approach here, asking what kind of propensity to reopen questions is optimal.

JW: To clarify a bit for economists, when you are talking about a "plan," you are thinking about what we would commonly call "contingent plans."

AG: Yes, right.

JW: You might find yourself either actually or hypothetically in some situation and the question is: What should you do in that circumstance? Is that correct?

AG: Right.

MA: Just a couple of more questions about meta-ethics. Your meta-ethical views emphasize the normative role of consistency. For example, in a striking passage from the beginning of your book, *Thinking How to Live* (Gibbard 2003), you suggest that inconsistency is a kind of normative mistake. Let me quote:

> Pluto, imagine, betrays his dear friend Minerva to get rich, leaving her impoverished and building a fancy house with the proceeds. Then in a fit of remorse and self-disgust, Pluto burns down the house, even cancelling the insurance so that his renunciation will be genuine. It's hard to claim that throughout this whole affair he acted without mistake. If it was fine to burn down the house, then it wasn't all right to betray Minerva in the first place. (p. 17)

Moreover, you require that preferences and beliefs satisfy the familiar consistency conditions from decision theory, what you call coherence conditions.

Can you explain the normative status of coherence? Is the normative statement that plans should be coherent itself just an expression of a plan, or does a requirement of coherence have a deeper, perhaps a natural, basis?

AG: Well, I would think vaguely that planning has a basis and that incoherent plans are ones such that if one aspect of them has a basis, then another aspect of them undermines that basis. That's vague, but that's the vague idea. Even though Peter Hammond's treatment (Hammond 1988) of what he calls consequentialism is a little too formidable for me to master entirely, I think the kind of analysis he gives is the kind that that's most fundamental.

MA: Take the statement, "You ought to have consistent plans." One view might be that this is simply another "ought" statement, and like any other "ought" statement, it doesn't describe a fact; it's simply my expression of some kind of very general plan to have consistent plans. And yet in your work, consistency seems to have a deeper status than that.

AG: Perhaps you don't need to have a plan at all, and so the dictum should be "Don't plan incoherently." This has a basis, namely that inconsistent plans are going to be in some respects self-defeating. Whatever might be to be said for such a plan, some alternative plan is guaranteed to do better. So a plan to eschew inconsistent plans is based on a logical feature.

MA: Intuitions also play an important role in your meta-ethics. Why do you think that coherent or consistent planning is not possible without relying on some intuitions?

AG: Well, how do we choose among different consistent contingency plans? It seems that we have no way of doing it without relying on judgment, and so some reliance on one's judgment seems inescapable. It has to be a critical reliance on one's judgment, but I don't see the alternative. Now it may be that it would be better to shape myself to be more spontaneous and less careful and not try to get my judgments to line up with each other. There are questions about what sort of person it has the best prospects for me to be. But even if I ask that, I have to rely on my judgment, and maybe on the judgment of the people I think about the problem with, to come up with an answer. Without some reliance on one's judgment, one doesn't have any basis for anything.

MA: All right. Let me ask one final question here about meta-ethics and then we can move on to questions about utilitarianism and individual goodness. There is of course a naturalist line in meta-ethics which looks to fully informed preferences. So Michael Smith's view, I take it, is that normative facts are facts about people's convergent fully informed preferences (Smith 2004), and I take it that Peter Railton has suggested a similar line both for facts about one person's good and for facts about moral good (Railton 2003). Fully informed preferences do play a role in your meta-ethics: I might plan to rely upon what I believe to be people's fully informed preferences. And yet you resist saying that "oughts" are simply assertions of fact, where the fact turns out to be a fact about fully informed preferences. Why can't we simply say that that's what "ought" facts are?

AG: I do think that there will be an ideal circumstance for forming preferences. But the question of what the ideal circumstances are is really a question of which preferences to trust. The answer that they must be fully informed might be too simplistic. The example I use in *Wise Choices, Apt Feelings* (Gibbard 1990b) is: What if I'm convinced that if I really were fully informed about what went on in people's innards, I would never want to eat with anyone else. Well, it seems that the thing to do there is enjoy other people's company when I'm eating, and ignore what's going on in their innards. So, it will have to be the right sort of being fully informed, and so there should be an account of what kind of information state is the ideal one for forming preferences.

And then we can ask: "What does that question mean?" And I say: "We can understand that as a question of what kinds of preferences to trust." If I know that something would be my preference under condition A and then something different would be my preference under condition B, shall I defer to the preference I would have under condition A or the preference I would have under condition B. This is a normative question; it amounts to asking what are the conditions such that I ought to adopt the preferences that I know I would have under those conditions.

JW: As I understand it, that's true not just for moral issues involving other people, but also in your prudential decisions. You have to trust some of your previous judgments, . . .

AG: Right.

JW: . . . your preferences; otherwise, you just can't get on with life.

AG: Right. In conversation, Brandt talked sometimes about choosing between the job that would be most prestigious and the job that one would be happiest in. And that's maybe mostly a prudential question, which we can put in the form of: "What would I prefer under ideal conditions." But regarding conditions as ideal is something we have to explain. And I propose we explain regarding them as ideal as deferring to the preferences that I learn I would have under those conditions.

MA: But there is a competing, descriptive analysis of the claim that I ought to do something. The descriptive analysis is, "I ought to do something given my current informational state, which might not be perfect, if everyone with full information, taking account of my current informational state, would want me to do that." And on this descriptive view, facts about what these fully informed advisors would want is a kind of descriptive fact. So, I view that as being a competitor to your expressivism.

AG: I don't think it is actually a competitor; part of it combines well with expressivism. I do think there will be a naturalistically formulable condition that's such that I ought to do a thing if and only if the thing meets that condition. But then we can dispute about what the condition is. This issue isn't entirely straightforward, as the example of having full knowledge of what's going on in people's innards suggests. So that still leaves the question: "Suppose people disagree about what the proper conditions for forming preferences are. What are they disagreeing about?" And I propose, they are really disagreeing about which sorts of preferences to trust, which conditions are such that the fact that preferences are formed in those conditions supports trusting those preferences or adopting them.

MA: Let's come back to themes of utilitarianism and the strains of commitment. You regard morality as being concerned with identifying how to live together on terms of mutual respect that nobody could reasonably reject. You argue that a contractarianism based on such terms would result in decisions being taken that utilitarianism recommends provided that the parties can be counted on to honor their agreements even in the face of strong personal countervailing motives (Gibbard 2008). You assume that there's full compliance. Critics of the veil of ignorance arguments of Harsanyi and Rawls have argued that in the non-ideal world in which we live, individuals would not necessarily abide by the agreements reached behind the veil once it is lifted. In Rawls' terminology, the "strains of commitment" are too great. How much confidence do you have that your utilitarian conclusions would survive if full compliance were not assumed? Relatedly, what relevance does ideal world theorizing have for thinking about morality in non-ideal circumstances?

AG: I think of something both in the spirit of Brandt (1979) and in the spirit of Rawls (1971) as applying here. We can imagine choosing a social ethos. Brandt talks about what moral code to support for one's society, and then has a complex description of what a moral code is. Rawls has phrases that amount to the same thing; he talks of a "public conception of justice." So in the Original Position, one chooses among alternative public conceptions of justice that we might have and institutionalize. Having something as the moral code of one's society, its public conception of justice, doesn't by itself produce full compliance. It has an influence, but not always an overriding influence. So what's to be chosen from behind the veil of ignorance will depend on what people will actually do if one set of standard or another comprises the public conception of justice. And some possible public conceptions of justice are going to elicit more compliance than others.

I think that despite the fact that Rawls seemed to deny it, his view amounts to a kind of indirect utilitarianism applied to broad features he attributed to the social world. Actually, my recollection is that on close reading, we see that Rawls avoided addressing indirect utilitarianism, and just denied that the parties behind the veil of ignorance would choose direct utilitarianism as the public conception of justice. That's quite consistent with the kind of rule utilitarianism that Brandt advocates. I thus claim that Brandt and Rawls are advocating the same sort of conception of justice, Brandt perhaps more straightforwardly and Rawls perhaps more convincingly.

MA: I have a question about a person's "good." You identify a person's "good" with whatever plays this role when we plan to live with each other on the terms described earlier. The concept of "personal goodness" is given content by specifying three axioms that this plan for living with others should satisfy. In your own words (Gibbard 2008), these axioms are as follows:

(1) "Prefer most to live with others on a basis that no one could reasonably reject on his own behalf." (p. 51)

(2) "A rejection on one's own behalf of a going social arrangement is unreasonable if, absent information about which person one would turn out to be, one would have rationally chosen that arrangement on one's own behalf." (p. 51)

(3) "One chooses rationally on one's own behalf only if one chooses what is prospectively most to one's good." (p. 52)

Why do you think that these requirements lead to a determinate quantitative measure of personal good? On what basis are these personal goods interpersonally comparable? Truistic but vague notions about my own well-being—what's good for me, what a good life would be, choosing on my own behalf—are all different ways to talk about my "goodness." But, do those notions, even for a particular person, have determinate content?

AG: I don't have a clear memory of what I was doing in setting down this account, but I'll try to rethink. What's to a person's good, on the kind of account I was advocating, is a normative question, a question of what to have play a certain kind of role in a conception of social justice. So questions of determinacy and interpersonal comparability are aspects of the broad normative question of how to live together on a basis of mutual respect. I'm trying to adopt the Rawlsian (Rawls 1971) and Scanlonian (Scanlon 1998) strategy of thinking about what no one could reasonably reject, by standards that rule out, among other things, its being the case that for every possible standard of justice there will be someone who could reasonably reject it. These specifications appeal to the notion of rejecting a standard "on one's own behalf," and don't by themselves specify what this amounts to. The idea is that one can't give determinate content to talk of "your good" or "my good" in advance of facing the ethical question of how good is to be distributed among us. In saying this, I think I am agreeing with Scanlon.

As for how determinate the notions of your good or my good are, these are then normative questions; they boil down to questions of what to choose in various hypothetical situations. We have to make actual choices, and if we are capable of thinking deeply about being in various circumstances, then we can make hypothetical choices. So the content of notions like "my good" is determined by the answers to questions that we answer by making hypothetical choices, or forming hypothetical preferences. That said, we are left with the broad question of how to live with each other on a basis of mutual respect and what standards govern this.

As for my own views on this, I'd say this: If, as I hope, we can make sense of the idea of "the intrinsic reward of ones life," I'd want that to figure in what rejecting a standard "on one's own behalf" amounts to. I think, though, that in the Tanner Lectures (Gibbard 2008), I wanted to be agnostic about much of this.

JW: In Harsanyi's Impartial Observer Theorem (Harsanyi 1955), the objective is to socially rank uncertain prospects from a moral point of view. He imagined an impartial observer being placed behind a veil of ignorance believing that he has an equal chance of being anybody once the veil is lifted. Assuming that his preferences behind the veil satisfy the expected utility axioms, Harsanyi argued that the impartial observer would rank alternatives according to their average utility. Now, you don't use this theorem in exactly the same form as Harsanyi; you talk about this in terms of "personal goodness." Can you explain what's different about your use of Harsanyi's thought experiment and his own, and why that's important?

AG: First—though I think Harsanyi would agree with me on this—one can't take a person's preferences all told as an indicator of that person's good. One's preferences are influenced not only by one's judgments of one's own good, but by one's judgments of the good of others. And so I struggle trying to see what the notion of a person's good can be. It's a notion that Scanlon (Scanlon 1998) doesn't think can be made coherent in advance of asking how a person's good is to figure in our conception of social justice.

JW: It seems then that your concept of personal goodness is designed so that you can take these considerations into account in a way that is not so transparent at least in Harsanyi.

MA: One quick follow up: As you well know, there's a big debate among preference theories of well-being about how to restrict preferences to be self-regarding.

AG: Yes.

MA: Brandt doesn't really deal with this in his book (Brandt 1979). And then Mark Overvold has a series of papers where he tries to explain self-regarding preferences in terms of preferences that are existence entailing (e.g., Overvold 1980), and that in turn has been criticized. So, are you saying that you want to circumvent that whole debate by thinking of personal good, as opposed to preferences either restricted or not? Are you saying that specifying personal good as playing these various roles is the solution to what may be an insoluble problem for preference-based views?

AG: Right. I think the question of what's to a person's good isn't strictly empirical. It is, as I have been saying, a question of what's to play the role of a person's good in our view of how to live together.

MA: Right.

AG: That was a question when we had our regular once or twice a week lunch with Richard Brandt and William Frankena for the first couple of decades after I came to Michigan. What a person's good consists in was something that we discussed quite a bit, and it never seemed to me that there was a simple, straightforward answer independent of what sort of role individuals' goods are to play in our general ethical theory.

MA: So, let me come back to utilitarianism squarely. One prominent critique of utilitarianism has been that it ignores the distribution of well-being. So-called prioritarians advocate a generalized form of utilitarianism, whereby well-being numbers are transformed by a concave transformation and then summed, which has the effect of giving greater weight to well-being changes affecting worse-off individuals. A second critique of utilitarianism has been that it ignores considerations of individual desert and responsibility. How might the utilitarian respond to these critiques?

AG: Well, on the second question, I would think the kind of answer Rawls (Rawls 1971) gives is right, that an ethos of desert and responsibility we can think of as being part of social mechanisms for promoting the goals to be promoted. So, we don't come to questions about how to arrange our social life already equipped with answers as to who deserves what. Our question is what to treat as desert in our social thinking and feelings—especially when we distinguish, as Rawls does, between desert and entitlement. Whether a team deserved to win is a different question from whether the team was entitled to the status of being the winner. And so to study desert we would have to ask, "What is the role of thought and feelings about desert, as opposed to entitlement?" And then, "How shall we arrange an ethos of desert to serve the purposes that it makes sense to want served socially?"

MA: I see. So on this view desert is a feature of social practices or understandings, and the question is which such practice maximizes utility. But there's a line in moral philosophy starting with Ronald Dworkin (Dworkin 2000), including people like Jerry Cohen (Cohen 2011) and Richard Arneson (Arneson 2011), which says that desert or responsibility has a much more bedrock role. Now, they're egalitarians, not utilitarians, but the thought would be that at the bedrock, if two people are badly off and we have to choose between giving a well-being improvement to the one who's badly off, as a matter of option luck, from bad choices, or giving it to the one who is badly off because of brute luck, we should do the latter—that as a kind of a bedrock matter of moral rightness, we should take into account responsibility or desert. So, I take it that you are resistant to that thought.

AG: Yes, I would think we have to develop an ethos of responsibility that promotes the intrinsic reward of people's lives, and part of what's going to come out of that is systems of incentives, including incentives of being regarded as deserving.

Please remind me what the other part of your question was.

JW: About prioritarianism.

MA: Right. So the second and orthogonal critique of utilitarianism comes from this idea of prioritarianism, which Parfit (Parfit 2000) originates and has been around now for at least twenty years. It says that well-being itself has diminishing marginal ethical significance. For those who are at lower well-being levels, giving them a given well-being improvement has greater ethical weight. It's a kind of welfarism, but it's not utilitarianism, which says that well-being changes have the same ethical weight regardless of the well-being level of the person receiving them. How would you respond to the prioritarian critique that utilitarianism is not sufficiently concerned about distribution? Utilitarians are concerned about the distribution of income, but they're not concerned about the distribution of well-being itself.

AG: Well, that supposes that we have a cardinal conception of well-being that is independent of what's to be preferred given difficult choices, choices under uncertainty where the intrinsic reward of one person's life may be maximized one way and the intrinsic reward of another person's life maximized a different way. Harsanyi didn't use the term "utility"' that way (Harsanyi 1955). He thought of a utility scale as indicating the choices that people are disposed to make, so that if a person stresses differences in how badly off she will be in a bad eventuality, the differences she stresses in her choices count as big differences in "utility."

I want to put this all not in terms of how a person is in fact disposed to choose but how it is rational for her to choose. Suppose, then, it is rational to be indifferent between a sure income of $20-thousand a year and an even-chance lottery of $10-thousand a year and $40-thousand a year. Then, the utility difference between $10- and $20-thousand a year will count as the same as the utility difference between $20- and $40-thousand a year. The utilitarian will be prioritarian in terms of income, but it could make no sense to be prioritarian in terms of "utility" in this sense. There's no such thing as the declining marginal utility of utility. It might still make sense to be prioritarian in terms of intrinsic reward, if some cardinal notion of intrinsic reward is sensible.

Whether we should be prioritarian in terms of intrinsic reward is another matter. On the direction I want to take Harsanyi's way of thinking, this depends on whether it is prudentially rational to be prioritarian in terms of one's own intrinsic reward. We might conclude that the declining marginal utility of income stems entirely from declining marginal intrinsic reward we derive from income.

So I take what is roughly the old-fashioned view of people like Arrow (Arrow 1977) and Savage (Savage 1954) that utility is a matter of what to prefer among uncertain prospects. It thus becomes analytic that expected utility is what is to be maximized. Utility is just an index that comes out of rational choices. Sen (2002) always seemed to be a native speaker of utility language in a way that I don't understand. I think utility just is a matter of how to choose given choices about uncertain prospects. In effect, the ethical claim is that the weightings that are prudentially rational for individuals are the one's society should adopt.

MA: We've already talked about Harsany's Impartial Observer Theorem. Let's talk about his Social Aggregation Theorem (Harsanyi 1955). In the Tanner Lectures (Gibbard 2008), you use a version of Harsanyi's Social Aggregation Theorem to justify a weighted form of utilitarianism. You argue that any outcome that is not Pareto optimal would be reasonably rejected by an ideal social contract. Individuals evaluate the goodness of outcomes on a personal goal-scale. Provided that the set of feasible combinations of goodness values is convex, any Pareto optimal outcome can be obtained by maximizing a weighted sum of the individuals' goods. This is a form of weighted utilitarianism that provides what you call a "coherent common rationale" or "common goal-scale" for us to adopt. Does your theory provide any constraint on what the weights used to aggregate the individual goods must be?

AG: First, I should mention, John Broome (Broome 2008) says that it isn't really Harsanyi's Theorem I'm using, but a theorem that has been long known and that, in a well-chosen phrase, he calls "the Tangent Theorem." He thinks, however, that the Tangent Theorem doesn't establish what I claim it does.

Now, clearly the argument I gave—my argument for the particular conclusion that I was claiming follows from what I was saying—that argument doesn't, all by itself, establish anything about what the weightings should be. Its real force, if I'm right, is to rule out a feature that the preponderance of recent philosophical views of justice share. Obviously, the question of what the weights should be is very important, maybe of central importance, but this particular result doesn't answer that question.

The weights, moreover, are by far not the only thing that this argument of mine, if it is correct, leaves unsettled. I don't start out making any assumption concerning what a person's good is, or even whether the notion of a person's good makes real sense. I do suppose that we can make sense of Scanlon's talk of a person's reasonably rejecting a proposed social contract (Scanlon 1998). I suppose that the upshot will be a social contract that no one could reasonably reject, and that people will act rationally within its constraints—where acting rationally includes satisfying the standard formal requirements of coherence in policies for action. The contract, then, will reflect the bases on which individuals might reasonably reject alternative possible contracts in favor of that one, whatever those bases might be.

The argument refutes the possibility that with the contract in force, different people would be pursuing somewhat conflicting goals. For if this were so, then a possible alternative contract would be available that served everyone's aims better, and it would have been reasonable to reject the contract that was in force in favor of this alternative that serves everyone's goals better. I'm not, then, saying that a satisfactory treatment of social justice stops at the point I reached, just that this is a point that we can get to by the kind of argument I was giving.

MA: Two further questions: If the personal goal-scale is not expectational, the feasible set may not be convex. Do your utilitarian conclusions depend on convexity? Also, you restrict attention to a single feasible set, but if planning is all about considering contingencies, then all possible feasible sets need to be considered. In your response to John Broome in the Tanner Lectures (Gibbard 2008), you offered some thoughts on this issue. Can you briefly tell us what they are and if you would revise them in any way now?

AG: The argument I gave for a common goal-scale does depend on convexity, and so that leaves the question of whether its conclusions bear on cases of non-convexity. This question raises complications that I don't claim to have worked through, but if we should all work for the same basic ends in a case where the set of possibilities is convex, then I find it hard to see why non-convexity should make a difference.

MA: You earlier said that having established weighted utilitarianism, it's a further question what the weights are. But at that point, assuming one's established weighted utilitarianism, why not just adopt an axiom of impartiality or anonymity to argue that the weights should be equal? Why should there be a serious ethical question at that point?

AG: I, myself, very much agree. But I'm also asking whether logical results that we can establish can force others to agree with you and me on this issue. David Gauthier has maintained that the social contract should efficiently realize the advantages of those who would be in a strong bargaining position in negotiating the social contract, and one's bargaining position depends on which alternatives are feasible (Gauthier 1986). I think there's a way to deal with his contentions, saying that we form the social contract from a standpoint before it's been settled who has the features that confer bargaining advantages. But that may require arguments that go beyond what I gave in my Tanner Lectures.

As for the feasible set not being fixed, I argue that the social contract must be agreed to, hypothetically, in advance of extensive information as to what's feasible and what isn't. Otherwise, it isn't much of a contract; it isn't much of a contract if people insist on renegotiating when news arrives bearing on what's at stake for them in an issue. So what parties face, as they reasonably reject various proposed contracts or not, is a set of prospects. Information keeps arising concerning how the social contract will bear on one's prospects and choices, but the contract restricts how to respond to such news. In making choices as to how to respond to this new information, one is to apply a fixed goal-scale and choose the prospects that rank highest on that goal-scale.

MA: Let me ask a related, but different question. You suggest in the Tanner Lectures that the ideal social contract will require the individuals to adopt a common goal-scale as their own goals. You note that if individuals pursue separate goals, prisoner's dilemma situations may arise where an outcome is suboptimal in light of all the goals. In short, your position seems to be not only that weighted utilitarianism is correct from the ethical standpoint, but that individuals are rationally required to adopt the ethical standpoint as their own. A different tradition in utilitarian thought, going back to Sidgwick (Sidgwick 1874), suggests that it is rationally permissible for individuals to pursue their own interests. Indeed, this seems to be Harsanyi's view as well (Harsanyi 1955). He uses the Impartial Observer and Aggregation Theorems to specify the content of ethical—that is, impartial—preferences, but Harsanyi does not argue that individuals are required to act on ethical or impartial preferences. How do you see your difference from Harsanyi in this regard?

AG: There's no such difference between me and Harsanyi, as far as I can see: It's one question what ethics requires, and another question whether rationality, in the restricted sense of coherence in one's policies for action, requires ethics. I don't have a proof that ideal coherence in action requires impartiality, and I don't think that such a proof can be given. The claim that coherence by itself requires ethics is false. That's something that I argued in my 1999 review essay on Christine Korsgaard's book, *The Sources of Normativity* (Gibbard 1999).

The ethical question, I am taking it from Rawls (1971) and Scanlon (1998), is what social arrangements and ethos would it be unreasonable for anyone to reject, given that everyone has the goal of establishing an ethos that it would be unreasonable for anyone to reject. Scanlon has this formulated much better than I'm managing to, and so take his formulation. We all have reason to prefer a social ethos where we benefit from others. We just do better if we aren't just producing an enormous prisoner's dilemma socially. So everyone has reason to prefer that social ethos given the choice. But that doesn't establish that it would be irrational to free-ride on others' adherence to the social contract, if one can.

MA: Aren't you worried not only about the strains of commitment, but the critique, which goes back to people like Shelly Kagan (Kagan 1989) and Sam Scheffler (Scheffler 1982), that says that utilitarianism as a personal standard is just incredibly demanding. That it's just not feasible in light of at least current psychology for people to actually adopt "maximizing overall well-being" without any preference for their own interests as their actual, day-to-day goal-scale.

AG: Well, the standpoint that weighs everybody equally into the social goal-scale I think of as a standard for comparing possible social ethoses. And a question about any possible social ethos is to what degree would having that as one's society's ethos actually elicit kinds of behavior that is rational for us to want from each other. So I would think these questions about demandingness and the strains of commitment and so forth come at the stage of trying to choose a social ethos. A social ethos that didn't have much influence on what anyone actually did toward others isn't one to want as a social contract. It would be a dead letter.

MA: I take it then that the common goal-scale, namely to maximize, you know, total well-being, comes in as you say at the level of a social ethos or social structure and not necessarily at the level of day-to-day choices. So, is that similar to Rawls' notion?

AG: I think so. Rawls (1971) sets up the question of what parties to the Original Position will want for their society. They then choose standards of justice to be institutionalized and publicly recognized—a social ethos, as I am putting it, or a moral code for the society, as Brandt (1979) puts it. So on any such account, the strains of commitment will bear heavily on the choice of a social ethos, the choice of a moral code for one's society.

MA: Is that in effect parallel to Rawls' notion that the principles of justice apply to the basic structure of society, but not the day-to-day choices?

AG: Yes, very much so. In preparing the published version of my Tanner Lectures, *Reconciling Our Aims* (Gibbard 2008), and cutting my reply down to size, I somehow unwisely persuaded myself to cut a passage stressing that I wasn't advocating direct utilitarianism, that the common goal-scale wasn't directly to guide our choices of what to do. That was a bad decision; reviews of the book mistakenly took me to be advocating direct utilitarianism.

In this regard, as you indicate, I'm very close to Rawls. Rawls has the Original Position as giving the standard for evaluating basic structures. The common goal-scale, I'm saying, applies to that, as what would guide people in the Original Position in choices of basic structures. The upshot is a kind of indirect utilitarianism, and I claim that what Rawls' arguments support really is a kind of utilitarianism. I should reiterate that what goes into the common goal-scale isn't decided by the argument I gave, even if the argument is successful. It's whatever is relevant to reasonable rejection, what a person might reasonably appeal to in order to reject an ethos and its implementation. The argument doesn't show this to be welfare, unless we define a person's "welfare" as whatever enters into the common goal-scale with respect to that person.

MA: I see.

JW: Allan, in concluding this conversation, Matt and I want to say how much we appreciate you taking the time to talk to us today.

AG: Well, I immensely appreciate not only you two taking the time and you, John, taking this whole trip, but especially the wonderful, thoughtful questions you put to me on the basis of such careful reading, such careful research and reflection. I'm sorry I haven't been on top of everything, since these are mostly things I did a long time ago, but it has been a delight to try to rethink matters with the two of you. Thank you both.

Acknowledgements We are grateful to Briana Brake and Susan Hinson from the Duke Law School for their assistance in preparing a transcript of this interview.

References

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statistics*, *34*, 199–205.

Arneson, R. J. (2011). Luck egalitarianism–A primer. In C. Knight & Z. Stemplowska (Eds.), *Responsibility and distributive justice* (pp. 24–50). Oxford: Oxford University Press.

Arrow, K. J. (1951). *Social choice and individual values*. New York: Wiley.

Arrow, K. J. (1977). *Essays in the theory of risk-bearing*. Chicago: Markham.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, *22*, 265–290.

Ayer, A. J. (1936). *Language, truth and logic*. London: Gollancz.

Brandt, R. B. (1963). Toward a credible form of utilitarianism. In H.-N. Castañeda & G. Nakhnikian (Eds.), *Morality and the language of conduct* (pp. 107–143). Detroit: Wayne State University Press.

Brandt, R. B. (1979). *A theory of the good and the right*. Oxford: Oxford University Press.

Bratman, M. E. (2014). *Shared agency: A planning theory of acting together*. Oxford: Oxford University Press.

Broome, J. (2004). *Weighing lives*. Oxford: Oxford University Press.

Broome, J. (2008). Comments on Allan Gibbard's Tanner Lectures. In Gibbard 2008 (pp. 102–119).

Cohen, G. A. (2011). *On the currency of egalitarian justice, and other essays in political philosophy*. Princeton, NJ: Princeton University Press [Edited by M. Otsuka.].

Dworkin, R. (2000). *Sovereign virtue: The theory and practice of equality*. Cambridge, MA: Harvard University Press.

Gauthier, D. (1986). *Morals by agreement*. Oxford: Oxford University Press.

Geach, P. T. (1964). Assertion. *Philosophical Review*, *74*, 449–465.

Gibbard, A. (1965). Rule-utilitarianism: Merely an illusory alternative? *Australasian Journal of Philosophy*, *43*, 211–220.

Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*, 587–601.

Gibbard, A. (1974). A Pareto-consistent libertarian claim. *Journal of Economic Theory*, *7*, 388–410.

Gibbard, A. (1977). Manipulation of schemes that mix voting with chance. *Econometrica*, *45*, 665–681.

Gibbard, A. (1979). Disparate goods and Rawls' difference principle: A social choice theoretic treatment. *Theory and Decision*, *11*, 267–288.

Gibbard, A. (1986). Interpersonal comparisons: Preference, good and the intrinsic reward of a life. In J. Elster & A. Hylland (Eds.), *Foundations of social choice theory* (pp. 165–193). Cambridge: Cambridge University Press.

Gibbard, A. (1990a). *Utilitarianism and coordination*. New York: Garland. [Originally submitted as a Harvard University Ph.D. thesis in 1971,]

Gibbard, A. (1990b). *Wise choices, apt feelings: A theory of normative judgment*. Oxford: Oxford University Press.

Gibbard, A. (1999). Morality as consistency in living: Korsgaard's Kantian lectures. *Ethics*, *110*, 140–164.

Gibbard, A. (2003). *Thinking how to live*. Cambridge, MA: Harvard University Press.

Gibbard, A. (2008). *Reconciling our aims: In search of bases for ethics*. New York: Oxford University Press [Edited by B. Stroud with commentaries by M. Bratman, J. Broome, & F. M. Kamm.].

Gibbard, A. (2012). *Meaning and normativity*. Oxford: Oxford University Press.

Gibbard, A. (2014a). Intransitive social indifference and the Arrow dilemma. *Review of Economic Design,* **18**, 3–10. [Originally written in 1969–1970.]

Gibbard, A. (2014b). Social choice and the Arrow conditions. *Economics and Philosophy,* **30**, 269–284. [Term paper for the Arrow–Rawls–Sen 1968–1969 Harvard seminar.]

Gibbard, A., & Harper, W. L. (1978). Counterfactuals and two kinds of expected utility. In C. A. Hooker, J. J. Leach, & E. F. McClellen (Eds.), *Foundations and applications of decision theory. Volume I: Theoretical foundations* (pp. 125–162). Dordrecht: D. Reidel.

Gibbard, A., & Harper, W. L. (1981). Counterfactuals and two kinds of expected utility. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance and time* (pp. 153–190). Dordrecht: D. Reidel.

Gibbard, A., & Varian, H. R. (1978). Economic models. *Journal of Philosophy*, *75*, 664–677.

Guha, A. S. (1972). Neutrality, monotonicity, and the right of veto. *Econometrica*, *40*, 821–826.

Hammond, P. J. (1988). Consequentialist foundations for expected utility. *Theory and Decision*, *25*, 25–78.
Hare, R. M. (1980). *Moral thinking: Its level, methods and point*. Oxford: Oxford University Press.
Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*, 309–321.
Hodgson, D. H. (1967). *Consequences of utilitarianism: A study in normative ethics and legal theory*. Oxford: Oxford University Press.
Hume, D. (1779). *Dialogues concerning natural religion*. London.
Jeffrey, R. C. (1965). *The logic of decision*. New York: McGraw-Hill.
Kagan, S. (1989). *The limits of morality*. Oxford: Oxford University Press.
Kaplow, L., & Shavell, S. (2001). Any non-welfarist method of policy assessment violates the Pareto principle. *Journal of Political Economy*, *109*, 281–286.
Lewis, D. K. (1969). *Convention: A philosophical study*. Cambridge, MA: Harvard University Press.
Lewis, D. K. (1976). Probabilities of conditionals and conditional probabilities. *Philosophical Review*, *85*, 297–315.
Lyons, D. (1965). *Forms and limits of utilitarianism*. Oxford: Oxford University Press.
Mas-Colell, A., & Sonnenschein, H. (1972). General possibility theorems for group decisions. *Review of Economic Studies*, *39*, 185–192.
Moore, G. E. (1903). *Principia Ethica*. Cambridge: Cambridge University Press.
Nozick, R. (1969). Newcomb's problem and two principles of choice. In N. Rescher (Ed.), *Essays in honor of Carl G. Hempel: A tribute on the occasion of his sixty-fifth birthday* (pp. 114–146). Dordrecht: D. Reidel.
Nozick, R. (1974). *Anarchy, state, and utopia*. New York: Basic Books.
Overvold, M. C. (1980). Self-interest and the concept of self-sacrifice. *Canadian Journal of Philosophy*, *10*, 105–118.
Parfit, D. (1984). *Reasons and persons*. Oxford: Oxford University Press.
Parfit, D. (2000). Equality or priority? In M. Clayton & A. Williams (Eds.), *The ideal of equality* (pp. 81–125). London: Palgrave. [Delivered as the Lindley Lecture at the University of Kansas in 1991.]
Parfit, D. (2011). *On what matters* (Vol. 1). Oxford: Oxford University Press.
Pearl, J. (2000). *Causality: Models, reasoning and inference*. Cambridge: Cambridge University Press.
Putnam, H. (1975). The meaning of "meaning". *Minnesota Studies in the Philosophy of Science*, *7*, 131–193.
Quine, W. V. O. (1951). Two dogmas of empiricism. *Philosophical Review*, *60*, 20–43.
Railton, P. (2003). *Facts, values, and norms: Essays toward a morality of consequence*. Cambridge: Cambridge University Press.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Regan, D. H. (1980). *Utilitarianism and co-operation*. Oxford: Oxford University Press.
Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, *10*, 187–217.
Savage, L. J. (1954). *The foundations of statistics*. New York: Wiley.
Scanlon, T. M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
Scheffler, S. (1982). *The rejection of consequentialism: A philosophical investigation of the considerations underlying rival moral conceptions*. Oxford: Oxford University Press.
Schick, F. (1969). Arrow's proof and the logic of preference. *Philosophy of Science*, *36*, 127–144.
Sen, A. K. (1970a). *Collective choice and social welfare*. San Francisco: Holden-Day.
Sen, A. K. (1970b). The impossibility of a Paretian liberal. *Journal of Political Economy*, *78*, 152–157.
Sen, A. K. (2002). *Rationality and freedom*. Cambridge, MA: Harvard University Press.
Sidgwick, H. (1874). *The methods of ethics*. London: Macmillan.

Smith, M. (2004). *Ethics and the a priori: Selected essays on moral psychology and meta-ethics*. Cambridge: Cambridge University Press.

Stalnaker, R. (1981). Letter to David Lewis of May, 21, 1972. In W. L. Harper, R. Stalnaker, & G. Pearce (Eds.), *Ifs: Conditionals, belief, decision, chance and time* (pp. 153–190). Dordrecht: D. Reidel.

Temkin, L. S. (2012). *Rethinking the good: moral ideals and the nature of practical reasoning*. New York: Oxford University Press.

Voorhoeve, A. (2009). *Conversations on ethics*. Oxford: Oxford University Press.

Peter J. Hammond

Philippe Mongin

Philippe Mongin (**PM**): *The important contributions you made to social choice theory, welfare economics, and social ethics are knit together by some common theoretical ideas, which I would suggest we take up successively in this discussion. I have selected in order, interpersonal comparisons of utility, utilitarianism, consequentialism, welfare economics under uncertainty, and incentive compatibility. Would you agree with this list and this order? What could come first, however, is some kind of intellectual biography. Perhaps you could explain to us how you first became interested in these topics and how they permeated your later academic life.*

Peter Hammond (**PJH**): Following your excellent overall plan, let us start with how I became interested in social choice and welfare, as well as in their application to public economics.

Actually, my interest in economics began in 1964 during the first year of my studies in Cambridge for the Mathematics Tripos. I was at Trinity Hall, where Geoff Harcourt was the Director of Studies for Economics. At a tea party early that October, he asked me when I was going to switch to economics, to which my reply was that I had no interest in doing so. But soon after, the Labour Party won a general election, with Harold Wilson succeeding Alec Douglas-Home as Prime Minister. The UK had a significant balance of trade deficit, and the new government was faced with a sterling crisis that Wilson blamed on the "gnomes of Zurich"—i.e., Swiss currency speculators. Devaluation of the pound was forestalled for three years, but that autumn I could not understand what I was reading in the newspapers about currency speculation and the measures that might alleviate the deficit.

2019 February 2nd, typeset from `interviewSCandWfinal.tex`.

P. Mongin (✉)
GREGHEC, CNRS & HEC Paris, 1 rue de la Libération, 78350 Jouy-en-Josas, France
e-mail: mongin@greg-hec.com

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_12

So I went into Heffer's Paperback Bookshop to look for an economics book I could afford, and bought the one which had the lowest price per page. This happened to be Paul Samuelson's *Economics*, his famous elementary textbook, which I read during the Easter vacation in 1965. At about this time, a college friend who knew of my interest in games such as chess and bridge, as well as cricket and other ball games, recommended that I read a book he had found in the library. This was another classic: von Neumann and Morgenstern's *Theory of Games and Economic Behavior*. Over the coming years I remember occasional forays into some old-fashioned works on mathematical economics such as that by R.G.D. Allen. There was also Ian Little's *Critique of Welfare Economics* which was too philosophical, I would guess, for my untutored taste at that time.

By 1966, I had several college friends studying for the Natural Sciences Tripos who were tiring of spending every afternoon in a laboratory. In those days, completing two years of Natural Sciences was enough for an honours degree provided you did some "diligent study" during your third year. So many of these friends switched to economics, and seemed to be enjoying it. Accordingly, I applied for a government funded scholarship to start studying economics myself in 1967/68, with a view to completing a Ph.D. degree in due course. My application was accepted, subject to not disgracing myself in my final year Maths exams, which I just about managed. Nick Stern, by the way, started along the same career path, though at a higher level of achievement.

After spending the summer vacation of 1967 working for a software startup company, my formal study of economics began that autumn. Cambridge offered me the privilege of being able to attend lectures by distinguished academic economists such as Michael Farrell, who taught not only basic econometrics but also the fundamental efficiency theorems of welfare economics, based on the first of Tjalling Koopman's *Three Essays on the State of Economic Science.* David Champernowne taught statistics. There were also Joan Robinson's lectures on growth and capital accumulation, in the course of which she taught us always to be clear about all our assumptions. Her teaching and writings were the inspiration for my essay on the assumptions of "contemporary neoclassical economic theology" (Hammond 1989).

I also learned much from Christopher Bliss's lectures on general equilibrium theory, and even more from his weekly supervisions. After starting with Keynes' really difficult *General Theory*, Chris had asked me next to read the much cited paper on the theory of second best by Lipsey and Lancaster (1956). He also hinted at feeling some dissatisfaction with it, and this encouraged me to find a counter-example to Lipsey and Lancaster's negative conclusions in the case of, for example, limits to free trade based on quotas rather than tariffs. Chris's dissatisfaction arose, I suspect, because he was aware of Diamond and Mirrlees' (1971) work on optimal taxation, already circulating in preprint form, which showed the desirability of production efficiency even in a second-best world where one has to use distortionary commodity taxes instead of first-best lump-sum redistribution.

The most important influence, however, was Jim Mirrlees. He gave lectures on public economics that included ideas that emerged eventually in his joint two-part paper with Peter Diamond on optimal taxation, as well as other advanced lectures on optimal growth theory. After Mirrlees moved to Oxford in 1968, that autumn I regularly attended mathematical economics seminars there to which Jim had invited me. Robert Solow, who was visiting Oxford at that time, was one of several prominent participants.[1] With Jim's powers of persuasion over the other economics fellows of Nuffield College, and Christopher Bliss's strong encouragement, I moved there one year later.

Soon after I arrived in Oxford, Jim passed on some conjectures involving an idea of his which helped resolve, in some special cases, the non-existence of an optimal growth path for an infinite horizon. I suspect he already had a clear idea of how to prove them, but I was encouraged to provide proofs anyway. They became incorporated into our joint paper on "Agreeable Plans" (Hammond and Mirrlees 1974). A few months later, Jim presented this to an IEA conference on Economic Growth, held during 1970 in Jerusalem, whose proceedings later appeared in the conference volume (Mirrlees and Stern 1974). Eventually agreeable plans were the subject of a considerable part of my Cambridge Ph.D. thesis on *Consistent Planning and Intertemporal Welfare Economics* (Hammond 1973).

PM: *Concerning our first topic, interpersonal comparisons of utility (ICUs), you stated that social choice theory would not have developed fruitfully (and would have remained a "science of the impossible", as you once wrote) if it had remained at the stage of Arrow's social welfare function (SWF), which excludes collective preferences that rely on ICUs (*Hammond 1987b*). Two papers of yours (*Hammond 1976a, 1979b*) modify the SWF concept to allow for comparisons of welfare levels and axiomatize concepts of equity in the Rawlsian sense. In later discussions, you consider other ways of introducing ICUs, for instance Sen's social welfare functionals, which permit enlarging the scope of ICUs beyond mere level comparisons. Here again, we could make a historical start and begin with the way your ideas on ICUs developed in the 1970s, when social choice theory began reconfiguring Arrow's work.*

PJH: While in Oxford, Jim Mirrlees asked me to look at the typescript version of what became Prasanta Pattanaik's book *Voting and Collective Choice* (Pattanaik 1971), which he had submitted to support his application for a senior research fellowship at Nuffield. It was a pleasure to read. Occasionally I would have a question, only to find it clearly answered in the next paragraph. Prasanta arrived at Nuffield during the academic year 1970/71, and gave a weekly lecture on social choice theory there. The focus was definitely on Arrow's *Social Choice and Individual Values* (Arrow 1951a) which I understood much better after attending Prasanta's lectures. Prasanta, perhaps influenced by meeting Allan Gibbard at Harvard, also introduced me to the idea of strategic voting (Gibbard 1973) and to Yasusuke Murakami's *Logic and Social Choice* (Murakami 1968). He also mentioned a monotonicity condition for sincere voting somewhat like Arrow's positive association. This had some resemblance to the monotonicity condition that Eric Maskin later made famous—see Dasgupta et al. (1979) and Maskin (1999).

[1] Some years after the 25th anniversary of this seminar, Michael Bacharach, Michael Dempster, and John Enos together edited a celebratory online volume.

My two years in Oxford also offered me the opportunity to meet for the first time some other prominent academics such as Amartya Sen and Joe Stiglitz, as well as the political scientist Brian Barry. I was also a friend of Iain McLean, another political scientist, who has since produced some fine work on the early history of social choice theory (MacLean and Hewitt 1994; MacLean and Urken 1995). In 1971, after my own two-year junior research fellowship at Nuffield had ended, I had the privilege of joining Christopher Bliss and Tony Atkinson at the University of Essex, where they had been recruited after Richard Lipsey and other founders of the Economics Department there had moved on. Essex is also where I first met Kevin Roberts, whose exceptional promise was already clear while he was completing his B.A. degree in mathematical economics.

Anyway, in the autumn of 1971 I read Amartya Sen's important book *Collective Choice and Social Welfare* (Sen 1970a). I had already seen his *Econometrica* paper on interpersonally comparable utilities, in which I found a minor mistake, duly acknowledged and corrected shortly thereafter (Sen 1970c, 1972b). I was prompted to start reading the books by Hare (1952, 1963, 1981), the noted Oxford philosopher, and discovering his major thesis that ethics was about universalizable prescriptive statements. And Chaps. 9 and 9* in Sen (1970a) usefully elaborate the key idea of an extended preference ordering over the Cartesian product of ordered pairs (x, i) that combine a social state x with an individual i. The idea of considering such pairs had already been taken up somewhat informally in the second edition of *Social Choice and Individual Values* (Arrow 1951a) (under the name "extended sympathy"), and more formally by Suppes (1966). In the case of two pairs (x, i) and (y, j) where individuals i and j differ, any preference between the two involves an ICU.

By my third year as a Lecturer at Essex, I had finally submitted my Ph.D. thesis (Hammond 1973). Before the examiners Terence (W.M.) Gorman and Geoffrey Heal passed it in 1974, Essex also granted me tenure as a Lecturer, which was possible in those days. During that academic year 1973/74, I was also allowed to teach a master's course in public economics. From the work of Diamond, Mirrlees, Atkinson, Dasgupta, Stiglitz and others, not to mention optimal growth theory, I had got used to writing down a utilitarian Bergson social welfare function to represent the objective of economic policy. Yet Arrow's theorem seemed to have converted politics from Bismarck's "art of the possible" to the "science of the impossible", to use the phrase in Hammond (1987b) that you have kindly recalled. I was also aware of John Rawls' *Theory of Justice*, and even his prior article on "Justice as Fairness" (Rawls 1958, 1971).

Several discussions and even criticisms of Arrow's theorem up to the early 1970s had focused on his condition of independence of irrelevant alternatives (IIA). This condition, which excludes the Borda rule and other similar SWFs based on rank order, is indeed crucially important, but it may not be the root of why avoiding dictatorship seemed so difficult. Chapters 9 and 9* of Sen (1970a) suggest that if one is to arrive at a non-dictatorial solution, in particular a Rawlsian maximin social ordering, or any basic notion of equity, the first thing to do is to add more information in the form of an extended preference ordering. Alternatively, a social welfare functional (SWFL), in the sense also explained in Sen's book, can embody this extra information. This information is compatible with a modified form of IIA, something like "independence of irrelevant interpersonal comparisons", to use the title of one of my later papers (Hammond 1991b), or to be more precise, "independence of interpersonal comparisons involving irrelevant alternatives". The Rawlsian maximin and its leximin extension satisfy this new version of IIA, but of course other axioms were needed for it to be fully characterized.

One of these axioms, whose role is to arrive at an equitable rule rather than the inequitable alternative of maximax, came to be known in the literature as "Hammond equity". It is the key postulate of the paper Hammond (1979b) which you kindly mentioned, with its improved characterizations of both Rawlsian maximin and leximin. The idea, however, is essentially already set out in Sen's short book *On Economic Inequality* (Sen 1973), if not in the earlier works of Pigou (1920) and Dalton (1920) claiming that progressive income transfers from rich to poor would increase social welfare.

I was fortunate that some preprint version of my 1976 paper had come to Ken Arrow's attention. The result was that, along with related work by Strasnick (1976) and the more general ideas of D'Aspremont and Gevers (1977), it was cited in an American Economic Association *Papers and Proceedings* issue (Arrow 1977). Arrow was right to recall the logical link to his own discussion of "extended sympathy".

PM: *As you just mentioned, d'Aspremont and Gevers were among the important contributors of these years. Their 1977 article offers a joint characterization of leximin and utilitarianism (D'Aspremont and Gevers 1977). To emphasize the duality of these rules is typical of the Belgian school of social choice; this comes out even more clearly in a paper (Deschamps and Gevers 1978) which shows that their axioms leave essentially no choice beside leximin and utilitarianism. Perhaps you could briefly assess this alternative treatment of leximin?*

PJH: Thank you indeed for reminding me of this important piece of research. Louis Gevers in particular was a good friend who was lost to us far too early. To be frank, I regarded two papers that Roberts (1980a, b) published in the same issue of the *Review of Economic Studies* as the culmination of the work during the 1970s on ICUs and their role in SWFLs. One of the lessons which emerged was that there are many ways of avoiding Arrow's impossibility theorem once one introduces the extra information required to make ICUs. It is nice to have appealing axioms that reduce the possibilities to a choice between leximin and utilitarianism. How to make this final choice, however, seems to require some deeper principles that, amongst others, I have explored in more recent work.

PM: *Now that we have discussed the rediscovery of ICUs by social choice theorists and your role in it, I would like to ask you about some of their theoretical underpinnings. Obviously, you regard them as being meaningful and feasible, unlike the new welfare economists, and at least in a common interpretation, Arrow himself, but your 1991 survey (*Hammond 1991a*) stresses that their empirical and ethical significance is often unclear. I wonder how this scepticism is compatible both with your early work on equity and your later endorsement of utilitarianism.*

PJH: First, I should probably try to clear up a common misunderstanding. It is that Lionel Robbins wrote that economists should not make interpersonal comparisons of utility. But what he actually wrote, at least in his later comment (Robbins 1938), was that such comparisons lack "scientific foundations". One interpersonal comparison that I have made in writing is that the few extra euros that a rich person might spend on a better bottle of wine are worth less than the same amount of money that a poor mother might spend on medicine that could save the life of her desperately sick child. Such a comparison is not purely factual; it has significant ethical content precisely because it crosses the divide between facts and values that British philosophers know as Hume's Law. Thus, I agree with Robbins when he writes that ICUs are "more like judgments of value than judgments of verifiable fact".

Sen's framework of SWFLs had played such a key constructive role in the developments of the 1970s. Nevertheless, by the time of my 1991 selective survey of ICUs and their use in social choice and welfare economics, I was becoming increasingly disenchanted with this framework. I think that my disquiet became even clearer when preparing my conference talk (Hammond 1996b) that emerged later as "Consequentialist Decision Theory and Utilitarian Ethics". The part of Sen's framework — which I was also using for at least a decade — that I was questioning concerned his treatment of individuals' utility functions as somehow basic, with comparisons of different individuals' utility functions added on as some sort of superstructure. Yet in the original position arguments used by Vickrey and Harsanyi, a person's utility is really just an estimate of the expected ethical value of becoming that individual upon emerging from what Rawls (1971) later called the "veil of ignorance".

Now, my favourite approach to constructing a utilitarian social objective is one that extends Adam Smith's "impartial spectator", who judges the welfare of single individuals, to an impartial benefactor who makes decisions for society as a whole. In this approach, it turns out that the most useful concept of individual utility is one that builds in social choice in original positions right from the start. Indeed, it seems best to consider "generalized" original positions with different specified hypothetical probabilities of becoming different individuals in society, as well as decision problems where even these hypothetical probabilities can be chosen.

PM: *You just described two theoretical moves, one away from Sen's social welfare functionals, and a move towards utilitarianism. The two moves are of course logically distinct. Is it a biographical accident that you made these two moves simultaneously, or is there a deeper reason?*

PJH: Well, Rawls introduced a different concept of the original position when arguing for his "difference principle", leading to the maximin or leximin SWFL. Of course Harsanyi (1974) famously criticized this as an inappropriate way of making decisions in the original position. Harsanyi maintained that, in the original position, one should maximize expected utility. The work I have done on rational choice reflects me to agree with Harsanyi here, though I do want to use a different concept of utility.

PM: *Another possible line about ICUs is that they are not the only possible interpersonal comparisons to make if one is to assess economic states of affairs. In a 2001 survey of ICUs written with Fleurbaey, you raise the question "interpersonal comparisons of what?" (Fleurbaey and Hammond 2004). Relatedly, Fleurbaey and Maniquet (2011, 2017) have thoroughly explored equity criteria based on the mere* *knowledge of the indifference curves, a direction that, as Fleurbaey and I later found out (in Fleurbaey and Mongin 2005), had already been taken by welfare economists whose work was obfuscated by Arrow's and Sen's successes. Could you possibly comment on this typical example of interpersonal comparisons that are not of utility? And perhaps also say what you think of its prolonged dismissal by social choice theory?*

PJH: The *Handbook of Utility Theory* that I co-edited with Salvador Barberà and Christian Seidl needed a chapter on ICUs. Yet I did not want merely to repeat what I had done in 1991. So I was very happy when Marc Fleurbaey agreed to join me as a co-author of the relevant chapter (Fleurbaey and Hammond 2004). The title "Interpersonally Comparable Utility" was carefully chosen with the intention of conveying the message that, rather than seeing ICUs as comparing different components of an independently constructed profile of individual utility functions, it might be better to think of this profile as being constructed right from the start so that it is "interpersonally comparable", with the eventual use of those comparisons in a social ordering very much in mind.

Before assessing the thought provoking contributions of Fleurbaey and Maniquet (2011, 2017), let us first go back to the work on SWFLs during the 1970s. D'Aspremont and Gevers (1977) proved that if a SWFL satisfies the conditions of unrestricted domain, independence of irrelevant alternatives, and Pareto indifference, then it must be welfarist—i.e., in a society of n individuals, one can represent it by a single preference ordering over the n-dimensional Euclidean space of vectors of personal welfare levels. Roberts (1980b) formulated a similar result that uses only the weak Pareto condition, though as I pointed out later (Hammond 1999b), his "weak continuity" assumption needs some slight strengthening to a condition I called "pairwise continuity". Anyway a welfarist SWFL, by definition, can use information only about welfare levels. This forces the ICUs to be independent of other considerations such as fairness. Kaplow and Shavell (2001, 2002, 2003), in their book and associated articles, offer a rather more contentious version of a similar point—namely, that imposing a condition like fairness is likely to violate the Pareto principle if one remains within a welfarist framework.

Now, social choice theorists have generally invoked an assumption such as an unrestricted domain of finite decision trees, or at least a sufficiently unrestricted domain. This plays a crucial role in those old justifications of welfarism. My own arguments for utilitarianism also put no restrictions on the impartial benefactor's judgements regarding the expected utility ascribed to the different personal consequences faced by each individual in the original position. Fleurbaey and Maniquet (2011, 2017), however, consider restricted domains of social choice problems such as those that involve dividing an aggregate bundle of several different commodities among a set of individuals. They consider social choice procedures which select allocations that maximize a Paretian welfare ordering which, as with Arrow social welfare functions, depends only on the individuals' preference orderings. Unlike Arrow, however, they not only consider a restricted domain of social choice problems, but they also relax independence of (preferences for) irrelevant alternatives. This allows their procedure to generate fair rather than dictatorial allocations. The various notions of fairness that they consider do depend on interpersonal comparisons of utility levels, of course, and these comparisons can take irrelevant alternatives into account. It is an interesting alternative to welfarist rules like utilitarianism.

Nevertheless, I cannot help wondering how well the kind of procedure they consider would perform in a dynamic or multi-period setting. After all, there can then be continuation sub-problems. In these, dependence on alternatives that have become irrelevant because they are no longer feasible may create difficulties of the kind discussed in my paper on "metastatic" choice (Hammond 1977). The difficulty is illustrated by the sensitivity of the Borda rule, or similar ranking rules, to the feasible set of options to which it is applied. It is also somewhat similar to the difficulties created in decision trees when an individual seeks to maximize a preference ordering which violates the independence axiom and which therefore may be represented by some "non-expected" utility function (Hammond 1988a, b).

Let us return, however, to the vexed question of whether maximizing the sum of individuals' expected utilities is somehow inimical to equity. This is what Diamond (1967a) argued in his famous comment on Harsanyi (1955). Obviously considering only total utility pays no attention to its distribution between different individuals. This neglect would have worried John Rawls even more than Peter Diamond, I suppose. But this valid criticism does not imply that a utilitarian is committed to ignore equity.

Before explaining this, it may be useful to consider some of Amartya Sen's other writings, apart from those that introduced the notion of an SWFL. Notably, he introduced rights into social choice theory (Sen 1970b). He also put forward the concepts of "capabilities and functionings" as what I would regard as important components of an ethically satisfactory notion of individual well-being (Sen 1985). Now, apart from any concern over the proliferation of concepts, there is one other worry, which I have written about in the case of rights (Hammond 1997). This is the thought that rights should matter only to the extent that they benefit the individuals who have them. But then, if rights really are beneficial, should they not be included in an appropriate measure of individual well-being that an impartial benefactor should be using? Pushing this idea far enough suggests that, despite our concern for fairness, we might still want to follow utilitarians at least to the extent of considering only utility, suitably measured, as well as interpersonal comparisons that are restricted to ICUs.

This brings us at last to my main point regarding Fleurbaey and Maniquet, and their new concepts of equity that you mention. Let us first recognize that as one increases the impartial benefactor's "consequence risk aversion" in the relevant original position, in optimal distribution problems that will tend to increase equality among first-best levels of utility. But our ethics may still suggest that there should be more equality than the level of consequence risk aversion on its own implies. In this case it may be more appropriate to emulate the treatment of rights suggested above, and include some measure of equity, reflecting possibly equality of both outcomes and opportunities, among each relevant consequence that affects individual welfare. Essentially the same idea has been discussed by, amongst others, Broome (1990, 1991) and Machina (1989). This line of argument reminds me, however, of Arrow's own "non-imposition" condition, as stated in his original "A Difficulty" paper, as well as subsequently in the first edition of *Social Choice and Individual Values* (Arrow 1950, 1951a). Overall, I am led to wonder if we should not refuse to give any special treatment to social *desiderata* such as equity or respect for rights unless we deem them to be inherently desirable for individuals. But in this case having more of these *desiderata* should increase any ethically relevant measure of individual "well-being" or utility.[2]

PM: *It is perhaps time to move on to the related questions of utilitarianism. When it comes to this, you offer both formal derivations and a philosophical endorsement. I see two main treatments, one you gave in 1982 and 1983 in connection with the ex ante versus ex post question, to be discussed later, and another from 1992, which I am focusing on now (*Hammond 1982, 1983, 1992*). This latter treatment is essentially the same as Harsanyi's; he is the theorist who perhaps had most influence on your work. You encapsulate an application of Harsanyi's Aggregation Theorem in a decision-theoretic framework of your own, but if this theorem has no ethical relevance, as Sen and others have complained, how would justify your more complex construction? Do you not think that you should defend Harsanyi in the first place?*

PJH: You are surely right in remarking that Harsanyi is the theorist who had most influence on my utilitarian views. Indeed, Amartya Sen once chided me (rather mildly) for my "dangerous fascination with Harsanyi". Quite late in life, Harsanyi did not demur when I suggested that a route to rule utilitarianism, which of course he strongly endorsed, might pass through some concept of "rule consequences", appropriately defined.

But let me now turn to your question regarding Harsanyi's aggregation theorem, and the criticisms of Harsanyi's version of utilitarianism that Sen and Pattanaik, amongst others, have made of it—criticisms, by the way, that Harsanyi seemed unaware of when Claude d'Aspremont and I interviewed him in Caen for *Social Choice and Welfare* (D'Aspremont and Hammond 2001). This is rather surprising unless one recognizes that this could have been an early sign of the illness that eventually took him from us. When I was asked to contribute to the volume of *Essays in Honor of John Harsanyi* that his friend and co-author Reinhard Selten was editing, I chose to offer an alternative proof of the aggregation theorem using the theory of linear programming, and did not provide anything more than the briefest reference to my criticisms of his version of utilitarianism (Hammond 1992).

In discussing Harsanyi's concept of utility, however, let us begin with William Vickrey. In barely one—breathlessly long—sentence on page 329 of Vickrey (1945), he imagines the case where each individual has their own VNM utility function, and the benevolent (or beneficent?) social planner is asked to maximize the expected value of an even chance lottery whose risky states of the world correspond to the individuals in the society. Vickrey, after all, was interested in measuring the marginal utility of

[2]Recently, Fleurbaey and Maniquet (2018, p. 1031) seem to recognize this possibility when they write as follows: "Note that our defense of the social welfare function could even be understood as a defense of the utilitarian approach, for an ecumenical notion of utilitarianism that is flexible about the degree of inequality aversion and the definition of individual utility."

income, or more precisely, the marginal rate of substitution between different individual's incomes. So for him, each individual's VNM utility function depends only on observable "objective" circumstances such as income, educational qualifications, date of birth, marital status, etc. One sentence, however, did not offer Vickrey the scope to explore in any depth the ethical foundation of the individuals' VNM utility functions—though he did take up the subject again in his paper on "social decision rules" (Vickrey 1960). This foundation is what Harsanyi (1953, 1955, 1976, 1977) attempted to provide.

An obvious difficulty is that each individual really has at best a unique cardinal equivalence class of VNM utility functions—equivalence meaning being related by an increasing affine transformation, so that ratios of utility differences represent constant marginal rates of substitution between appropriate probability shifts. As Pattanaik (1968) in particular has pointed out, even if one fixes for each individual one particular VNM utility function within their own cardinal equivalence class, these functions will typically have to be weighted before being added. Mathematically, this is identical to the problem that arises in extending the Savage (1954) and, perhaps more precisely, the Anscombe and Aumann (1963) theory of subjective probability and subjective expected utility (SEU) to the case when there is a separate consequence domain associated with each state of the world. In my FUR 1997 conference paper (Hammond 1999a) on state-independent utility on state-dependent consequence domains, I show that this problem can be resolved if one contemplates decision problems where the agent can choose the probabilities of the different states of the world. Similarly, I think the Harsanyi problem can be resolved by contemplating the choice of generalized original positions with different probabilities of becoming different individuals in the society.

Another weakness of Harsanyi's approach, and of Vickrey's too, I suppose, comes back to my doubt that one should start by postulating a given profile of individual utility functions, rather than deriving them from suitable ethical decision problems as we try to do in Chichilnisky et al. (2018).

PM: *Am I correct in saying that this amounts to vindicating Harsanyi's Impartial Observer Theorem, as against the Aggregation Theorem we just discussed?*

PJH: What I am trying to vindicate is his use of an original position and of a VNM utility function in that position in order to arrive at the ethical decision criterion of what Harsanyi called the "impartial observer", though perhaps one should think instead of a more active "impartial benefactor".

While we are discussing Harsanyi, perhaps I may mention another significant disagreement I have with him, which concerns his view that population should be chosen to maximize average utility per head of population. Both in my Ph.D. thesis and the later *Social Choice and Welfare* paper on consequentialist demographic norms (Hammond 1973, 1988c), I have criticized this criterion as yielding dynamically inconsistent preferences—unless, that is, you keep track of the entire population of humanity all the way back to the origin of *homo sapiens*. Though one wonders if even this is far enough back given recently discovered DNA evidence that our ancestors interbred with Neanderthals, who in turn interbred with Denisovans. Anyway, with an obvious assumption that the unchangeable welfare of long forgotten ancestors should be irrelevant to contemporary decisions, one returns to classical total utilitarianism. This is even compatible with a version of the Vickrey–Harsanyi original position, where one must reckon with some possibility of never coming into existence. Here, by the way, I like to define zero utility as the level ascribed to any person who never comes into existence. This differs from Chuck Blackorby, Walter Bossert and David Donaldson who, in their book *Population Issues in Social Choice Theory* as well as in several earlier research articles, prefer to use a "critical" utility level (Blackorby et al. 2005).

To sum up, in the end I do not really feel any need to defend Harsanyi. But I am very happy to acknowledge being greatly inspired by his (and Vickrey's) appreciation of the importance of using VNM theory in a fundamental way in order first to construct an interpersonally comparable measure of utility, and then to compare social states based on their expected utilities in suitable even chance lotteries. Also, Harsanyi's (1974) famous criticism of Rawls's view of the decision-maker's extreme risk aversion in the original position remains very much to the point, as well as the subject of Hammond (1975b) on "extreme inequality aversion".

Finally, I found myself wanting to refute the claim in Myerson (1985) that "interpersonal comparisons of utility cannot be given decision theoretic significance". So, in our *Handbook* chapter on "Interpersonally Comparable Utility" (Fleurbaey and Hammond 2004), we refer to Harsanyi (1987), which opens with the example of choosing which friend should receive an opera ticket one could not use.

PM: *If I may linger on the issues here, Harsanyi's utilitarian interpretation of his two theorems runs into two separate difficulties. One is the meaningfulness of the weights in the weighted sum of individual VNM utilities that both theorems derive, and the other is the very choice of VNM utilities to represent individual preferences that are ordinal in nature. Your answer seems to address the former problem, but not the latter, which critics usually regard as being the more worrying. I believe like you that an Anscombe–Aumann reconstruction of the original position may fix the weights—I think each of us has made the point independently—but this takes the VNM representation for granted, so the other problem is still on the table.*

PJH: If I understand correctly, you made the point about fixing the weights in Mongin (2001). By contrast, in our recent working paper (Chichilnisky et al. 2018), Graciela Chichilnisky, Nick Stern and I actually circumvent the state-dependence issue involved in fixing the welfare weights in the original position. We do so by postulating an all-encompassing domain of personal consequences, each of which represents a possible individual life experience, including the possibility of no life at all. This is instead of each individual having their own specific personal consequence domain.

Our formulation allows us to treat ethical decision problems where the number of individuals, as well as their personal characteristics, are risky and affected by the decision being made. Such problems can then include even the hypothetical choice of a partial or biased original position, which implies interpersonal comparisons.

Given the relevant risks, it is natural to use what Kolm (1971, 1994) might call a "fundamental" VNM utility function defined over this all-encompassing domain. An additional equity axiom inspired by Patrick Suppes' famous paper on grading principles (Suppes 1966) then suggests that, when an impartial benefactor is choosing a world history rather than an original position, it is right to use an unbiased or impartial original position that takes into account the probability of a person coming into existence.

PM: *The decision-theoretic framework involved in our interchange is consequentialism, which we may discuss now. Among moral philosophers, this word applies to those doctrines which define the rightness of an act from the value of its consequences. In your work, especially in your reference article (*Hammond 1988a*), it both applies to an axiomatic reconstruction of decision theory and appears as one salient axiom in this reconstruction. Could I ask you to recapitulate the reconstruction and clarify the semantic connection between your consequentialism and that of the philosophers?*

PJH: Let me precede an answer, if I may, with some more personal background. I have already mentioned attending Joan Robinson's lectures on economic theory that she gave in Cambridge during the 1967/8 academic year. Most likely inspired by an argument from Jan Graaff's book *Theoretical Welfare Economics*, she once pointed out that a difficulty with long-range planning was how to specify the horizon (Graaff 1957). The point is that, however far away the chosen horizon may be, it will eventually be reached, whereupon one will be forced to revise one's original plan. I remember raising my hand and suggesting that recent work on infinite-horizon planning that I had learned about from Mirrlees' lectures showed a possible way round this objection. This interchange does much to explain why *Consistent Planning and Intertemporal Welfare Economics* became the title of my Ph.D. thesis (Hammond 1973), with several of its later chapters devoted to infinite-horizon planning models.

Sometime during the years 1969–71 that I spent at Nuffield College, Graham Pyatt recommended to me the classic paper on myopia and inconsistency by Strotz (1956). Also, at the 1970 World Congress of the Econometric Society in Cambridge, Chuck Blackorby gave a talk in which, according to my possibly imperfect recall, showed that a consumer who maximizes inconsistent preferences at different times will have intertemporal demand functions that violate the Slutsky conditions for consistent preference maximization.[3] That work helped inspire my "Changing Tastes and Coherent Dynamic Choice" (Hammond 1976b), which made a similar point in the framework of choice functions over finite sets that Arrow had used in his 1959 *Economica* article (Arrow 1959). I had also become aware of the concluding part of the second edition of Arrow's book (Arrow 1951a), which set out to justify the collective rationality postulate built into his definition of a SWF. His argument was based on an informal version of the path independence postulate that, in the theory of revealed preference, plays a role in ensuring the integrability of consumer demand functions. Meanwhile, probably while still in Oxford, my supervisor Mirrlees introduced me to the money pump argument, which typically neglects the likely effect of decreasing wealth on an agent's willingness to pay for each successive preferred alternative. Anyway, I was encouraged to look for a better argument in favour of transitivity.

[3]The published paper (Blackorby et al. 1973) uses a different approach to demonstrate inconsistency of demand behaviour.

Then, in 1974/5 at the Australian National University, I suddenly realized that I could offer a different justification for collective rationality, and for transitive preferences more generally. It drew some inspiration from Vickrey's 1964 book *Metastatics and Macroeconomics*, which emphasized that, especially after Debreu's *Theory of Value*, much of economic theory treated the consumer as if making a single lifelong consumption plan subject to just one life-time budget constraint (Vickrey 1964; Debreu 1959). As I now see it, this formulation reflects von Neumann's assertion that a game in extensive form could be fully analysed by considering only the normal form, in which each player is restricted to making a single strategy choice (Von Neumann 1928). In the book with Morgenstern, this normal or reduced form of the game is described as one where each player's only move is to announce in private to an umpire the strategy that the umpire should execute on the player's behalf (Von Neumann and Morgenstern 1944). This general claim that only the normal form matters is valid, I believe, only in a few special cases. One concerns single-person decision problems, and so team decision problems as well. A second concerns the two-person "zero-sum" (or "strictly competitive") games that were really the only ones that von Neumann was able to solve satisfactorily.

Now, my work on changing tastes showed that, except in trivial cases, this reduction to the normal form does not work when an agent has dynamically inconsistent preferences. In fact, it works for all finite decision trees if and only if the decision-maker has preferences for paths through each tree that are consistent and also transitive. Moreover, in the case of a social decision tree, an Arrow SWF allows such a reduction if and only if IIA is satisfied. Inspired by Vickrey (1964), I chose the title "Dynamic Restrictions on Metastatic Choice" for this first paper (Hammond 1977).

So far there had been no attempt to include risk or uncertainty in this approach to decision theory. The next stage comes in a paper that was never published. It bore a title something like "Some Uncomfortable Options in Welfare Economics under Uncertainty". An egalitarian welfare economist like myself, particularly after enjoying the experience of being Tony Atkinson's colleague for several years, might like to see as equal a distribution of income as possible, taking into account various incentive constraints. The paper on "Fully Progressive Taxation" written with Partha Dasgupta helped explore what might be possible (Dasgupta and Hammond 1980). Once one introduced risk and different personal probabilities, however, an ex ante Pareto efficient allocation that started out egalitarian would only be egalitarian ex post in rare circumstances. I was uneasy about this evident dynamic inconsistency between ex ante and ex post.

In 1980 Maurice Salles invited several of us to a conference he had organized with Prasanta Pattanaik to be held in Caen, immediately after the World Congress of the Econometric Society in Aix-en-Provence. Indeed, this meeting led eventually to the founding of *Social Choice and Welfare* as the leading specialist journal. My plan had been to present my "Uncomfortable Options" paper. But while travelling from Aix to Caen, I suddenly realized that the main idea of "Dynamic Restrictions on Metastatic Choice" (Hammond 1977) could be applied to decision trees with chance nodes, where it would imply the independence axiom. That led to a significantly revised conference paper emphasizing that dynamic consistency would require an ex post approach (Hammond 1983). The idea of calling this "consequentialist" decision theory only came a few years later.

After these perhaps rather lengthy preliminaries, let us now turn to the paper which you asked me about (Hammond 1988a). A paper, by the way, that had been prepared while I was visiting CORE in 1986, to which you very kindly invited me back to give some lectures on this topic in 1994. Anyway, it might actually be better to go a bit further forward to "Consequentialism, Structural Rationality and Game Theory", which appeared in the proceedings of a 1993 conference of the International Economic Association that Ken Arrow co-organized (Hammond 1996a). The later paper may offer a rather clearer statement of consequentialist decision theory, based on three axioms applied to finite decision trees in a specified domain. For simplicity, let us focus first on the case of finite decision trees without risk or uncertainty. Then there is no need to discuss chance nodes at which what Anscombe and Aumann (1963) call a "roulette lottery" is resolved, nor "natural" or "event" nodes at which a "horse lottery" is resolved.

Of these three axioms, the first is that the theory applies on an unrestricted domain—or at least a domain that is sufficiently unrestricted to accommodate the trees that are used in the proof that choice must maximize a (complete and transitive) preference ordering. Indeed, a decision theory whose scope must be limited to a restricted domain of finite decision trees seems evidently insufficient, just as majority rule is in the context of social choice theory.

To explain the second and third axioms, it may be best to introduce a little notation. Given any finite decision tree T and any decision node n of T, let $M(T, n)$ denote the non-empty set of moves that are feasible at n, and let $M^*(T, n)$ denote the set of moves at node n which our decision theory deems normatively acceptable. Since indecision at node n is not an option, as Arrow (1951a) lucidly explained in his thesis, the set $M^*(T, n)$ must be a non-empty subset of $M(T, n)$. We also let $T(n)$ denote the continuation subtree of T that starts at its initial node n. This is the result of "snipping off" $T(n)$ from T, to use the evocative terminology of Machina (1989).

Now, the second axiom is dynamic consistency. Using the notation we have just introduced, this requires that the set $M^*(T, n)$ of acceptable moves at node n in the tree T should be the same as the set $M^*(T(n), n)$ of acceptable moves at the initial node of the continuation subtree $T(n)$. This equality can be made almost tautological by recognizing that any earlier plans of what to do when the decision maker reaches node n are essentially irrelevant by the time n has been reached, so it is $M^*(T(n), n)$ that determines what the agent will actually do at node n. Thus, if we define $M^*(T, n)$ as the set of moves that the agent might actually make at node n, it must equal $M^*(T(n), n)$. Though I find this consistency axiom entirely compelling, in a talk I gave in 2017 to the conference of the European Society for the History of Economic Thought, I set out to replace it by one requiring that the agent should never face any possibility whatsoever of regretting her original decision plan.

The third axiom is intended to express Arrow's claim that an act should be judged by its consequences (Arrow 1951b). It also conforms with Savage's definition of an act as a mapping from states of the world to their consequences (Savage 1954). This "consequentialist" axiom is the one that does by far the most work, so I felt free to base the name of this approach to normative decision theory on this axiom. Indeed, I would rather call it a "pre-axiom", from which so many axioms in other versions of normative decision theory follow as logical implications. Given any finite decision tree T, one can use backward recursion to construct at each successive decision node n of T: first, the set $F(T, n)$ of all possible consequences that can result ultimately from choosing any move from the feasible set $M(T, n)$; second, the set $F^*(T)$ of all "acceptable" consequences that can result ultimately from choosing any acceptable move from $M^*(T, n)$. Evidently each $F^*(T, n)$ is always a non-empty subset of the corresponding $F(T, n)$, as one can prove by backward induction. What the "consequentialist" axiom requires is the existence of a choice function mapping non-empty finite subsets F of the consequence domain Y into non-empty choice subsets $C(F)$ with the property that, for any finite decision tree T, one has $F^*(T, n) = C(F(T, n))$ at every decision node n. Thus, the agent must behave in any decision tree as if a planned sequence of moves, or "decision strategy", can be chosen if and only if it has acceptable consequences. This "consequentialist" axiom is the application to single-person decision trees of von Neumann's principle that we discussed earlier—namely, the claim that it loses no generality to reduce a game in extensive form to its normal form (Von Neumann 1928).

The main result of this consequentialist approach to normative decision theory is that the three axioms hold if and only if the consequence choice function which maps each non-empty finite feasible set F to the non-empty choice set $C(F) \subseteq F$ is ordinal. That is, there must be a complete and transitive preference ordering defined on the consequence domain Y such that the consequence choice set $F^*(T, n) = C(F(T, n))$ at any decision node n in any finite decision tree T corresponds to choosing consequences to maximize that ordering over the finite feasible set $F(T, n)$.

PM: *This is a very helpful summary. Perhaps you could now extend it to the case of chance nodes, which permits deriving a preference ordering that satisfies the VNM independence condition. As you take this logical derivation also to be a normative justification of this notorious condition, it has led to a lively controversy with Machina.*

PJH: Where Mark Machina and I differ is indeed in the extension of the above consequentialist theory to decision trees with chance nodes. In those trees, the preference ordering is over consequence lotteries. The key axiom that $F^*(T) = C(F(T))$ would actually imply universal indifference if we were to allow random moves at chance nodes to have zero probability, so we do not allow this. Perhaps I may be allowed to mention here two papers (Hammond 1999b, c) I wrote that allow infinitesimal probabilities at chance nodes, building on my contribution to the *Festschrift* for Patrick Suppes, the "scientific philosopher" (Hammond 1994a).

Once we exclude both zero and infinitesimal probabilities, however, the existence of a preference ordering satisfying the independence axiom is necessary and, when combined with continuity with respect to changes in probabilities at chance nodes, sufficient for the preference ordering to be represented by the expected value of each utility function in a non-empty equivalence class. According to my possible misinterpretation of how his work differs from mine, Mark chose to deny what I call dynamic consistency by claiming that continuation subtrees are somehow different from decision trees. In his book that sets out his theory of resolute choice, Ned McClennen (1990) does something similar. By contrast, I argue that the key consequentialist axiom that behaviour should be explicable by its consequences should apply not only for each entire finite decision tree, but also for each continuation subtree.

Obviously, we can differ over the relative normative appeal of our differing theories. But there are two methodological considerations that worry me. The first is that if we are going to treat continuation subtrees as different from decision trees, why not do the same when considering decision trees without risk or uncertainty? Then one cannot even use my consequentialist argument for the existence of a preference ordering. This certainly troubles me, even if it might not bother Machina or McClennen, since neither hesitates to assume that there is a preference ordering. Second, faced with any decision tree, we would have to start worrying about what had preceded it, including resolutions made in the past. At least we would seem to need a fuller decision theory that includes scope not only for past resolutions but possibly a richer domain of consequences to recognize those that emerge from adhering to or departing from past resolutions.

I am not sure if Mark Machina agrees with my normative theory. But I definitely accept his theory of non-expected utility maximization as a possibly useful but refutable description of what many people—perhaps even a large majority of people—actually choose when faced, for instance, with the lotteries presented in the well-known paradox due to Allais (1953). Also, while non-expected utility may describe accurately what people do, especially in a laboratory, there is still the possibility which Savage discusses in his *Foundations of Statistics* that violating EU theory, at least in the context of the Allais paradox, seems a mistake like paying too much extra for a car that includes an option such as a fitted radio—in the era of Savage (1954).

PM: *I am coming now to another part of your work, which concerns the extension of standard, static welfare economics and social choice theory to the uncertainty case. The many problems this extension raises had been debated in the late 1960s and 1970s by* Diamond (1967a, b), Drèze (1970), Starr (1973) *and others before you tackled them in the 1980s, but you put them more sharply than these early writers, and this was perhaps because you had the comparative advantage of being a social choice theorist as well as a welfare economist. To remind the reader, when uncertainty prevails, the Pareto principle can be applied either to the individuals' ex ante preferences, thus implicitly respecting their subjective probabilities and risk attitudes, or only to their ex post preferences in each given state of the world, thus ignoring these items entirely. In the ex ante approach, the ex ante Pareto principle holds, plus the assumption that individuals satisfy SEU, whereas in the ex post approach, the ex post Pareto principle holds, plus the assumption that the social observer satisfies SEU. In a 1981 paper and two book chapters in 1982 and 1983, you emphasize the conflict between the two approaches (*Hammond 1981, 1982, 1983*)—a conflict I myself investigated much later (*Mongin 1995*), in order to put it into a proper axiomatic framework. You also express a considered preference for the ex post approach. Your ex post solution is actually also utilitarian, and it thus provides you with another line to defend this* *doctrine. I do not think you have later revisited these claims. Would you endorse them today in the same way as you did at the time?*

PJH: The topic of ex ante versus ex post is yet another to which I was introduced by Jim Mirrlees while he was supervising me in Oxford. He ascribed the idea to Peter Diamond. Indeed there is a highly relevant footnote to Peter's paper that set out a theory of efficient allocations in a model of the stock market (Diamond 1967b). The footnote explained that he had chosen to follow the ex ante approach, while recognizing that the ex post approach could be an interesting alternative.

I am not quite sure what was the immediate impetus, but eventually I took up this suggestion of Peter and Jim in the late 1970s. By the way, the ex post approach could use the individuals' own risk attitudes, provided these were judged to be ethically appropriate. It could also use their subjective probabilities, not indirectly in their subjective expected utilities, but directly through some other aggregation procedure. The paper by Starr (1973) had been interested, inter alia, in exploring cases when differing subjective probabilities would not preclude markets from achieving an ex post efficient allocation. The *Economica* paper that you kindly mention (Hammond 1981) was indeed an attempt to extend some of these results to a more general setting with many periods, etc.

One of the books you mention is *Utilitarianism and Beyond*, edited by Amartya Sen and Bernard Williams, with a number of notable contributions (Sen and Williams 1982). I think I was able to get the title changed from *Beyond Utilitarianism* when I suggested to Amartya that at least one, and eventually several, of the contributors might not want to venture all the way out there.

Anyway, after these works, my interest in the subject remained essentially passive for at least 25 years. But then in December 2008, at a seminar organized by the Paris School of Economics, I gave a talk showing that in a framework like the insurance model of Malinvaud (1972a, 1973), even if not all individuals could agree on the associated probabilities of insurable individual risks, nevertheless ex post Pareto efficiency would still require full risk pooling provided that agents could all agree that these risks were described by what De Finetti (1937) described as exchangeable random variables. These results, I fear, still need to be written up.

Finally, there was a conference in Milan in December 2015 where I had played a small part in the organization, along with Maurizio Motolese and Carsten Nielsen whom I had known as Ph.D. students at Stanford. The theme was "Welfare Evaluation under Subjective Expectations", with some emphasis on overconfidence. My talk there was some sort of survey of social choice theory in such a setting, without finding anything really new to say.

Closely related to these issues is the problem of aggregating subjective probabilities, and this leads me to share the memory of a seminar by Michael Bacharach during my two years in Oxford. This seminar introduced me to the concept of an externally Bayesian procedure for aggregating personal probabilities. A letter to Michael ensued shortly thereafter in which I pointed out that, instead of taking a weighted arithmetic mean of different individuals' probabilities, one should instead take logarithms first and consider a weighted geometric mean. That way you get a rule for which the operations of aggregation and Bayesian updating commute, which is the property that defines the externally Bayesian criterion.

PM: *I am now venturing into a territory that is less familiar to me, i.e., mechanism design, implementation and incentive compatibility. Would you explain to us how you became interested in this topic and eventually arrived at the reference paper (*Dasgupta et al. 1979*) you coauthored with Dasgupta and Maskin in 1979 on these topics?*

PJH: Again, this part of my work starts with trying to fill in some logical gaps I found while teaching a master's course in public economics at Essex. Now, the efficiency theorems of welfare economics link Pareto efficient allocations to the Walrasian equilibrium outcomes of a competitive market mechanism, provided that one introduces lump-sum wealth redistribution. Paul Samuelson in his *Foundations of Economic Analysis* had recognized that there could be problems in acquiring the information needed to implement satisfactorily such a redistribution scheme (Samuelson 1947). Actually, since the distribution of wealth typically affects aggregate consumer demands and supplies and so market-clearing prices, the redistribution scheme should really specify how each individual's wealth depends on prices, as it will when it is a rule for distributing the profits of firms in a general Arrow–Debreu economy (Arrow and Debreu 1954; Debreu 1959; Grandmont and McFadden 1972). That said, Diamond and Mirrlees' results on optimal taxation applied in a second-best world where there could be distortionary commodity taxes, but any lump-sum wealth redistribution had to take place through uniform lump-sum subsidies (or much less plausibly, taxes) that were the same for everybody, regardless of any individual circumstances. This is the kind of "imperfect" economy that Mirrlees (1997) discussed in his Nobel Prize lecture on the "economics of carrots and sticks".

Indeed, some sort of folk theorem was going around that concerned first-best optimal allocations in the model of income taxation that had been briefly explored in Vickrey (1945) before Mirrlees (1971) gave it a much fuller treatment. Assuming that leisure is a normal good, the unique first-best allocation has all workers consume the same amount, regardless of their skill; workers whose known skills were greater, however, would be expected to work more. Such an allocation seems well described by what became the Marxist slogan: "From each according to their ability; to each according to their need!" (Marx 1875). Yet trying to put this into practice would neglect incentives to reveal one's true skill and lead to this description of working in any of the Communist nations of Eastern Europe prior to 1989: "So long as the bosses pretend to pay us, we will pretend to work." Actually, as I learned from a lecture that my colleague Mark Harrison gave on the occasion of the centenary of the Russian Revolution (Harrison 2017), the Leninist Soviet constitution already suggested "to each according to their work".[4]

During the early 1970s, Leo Hurwicz's Richard T. Ely lecture to the American Economic Association, "The Design of Mechanisms for Resource Allocation", had appeared in print (Hurwicz 1973). As, of course, had Gibbard's (1973) paper on strategyproof voting schemes, though that seemed less relevant to the issues surrounding

[4]See Chap. 5, Sect. 3 of Lenin (1917), as well as Article 12 of the 1936 Constitution of the USSR.

economic allocations which interested me at that time. Hurwicz's paper, despite being largely concerned with impossibility results, like Gibbard's, did offer one positive suggestion: a competitive market mechanism did seem to be strategyproof in a large economy with many consumers.

In 1977 I had agreed to teach during Stanford's summer quarter so that I could attend the summer workshop of the Institute of Mathematical Studies in the Social Sciences, organized by Mordecai Kurz. So unfortunately I missed the conference that led to Jean-Jacques Laffont's volume *Aggregation and Revelation of Preferences* (Laffont 1979). But at the Stanford workshop I could at least present an early version of what became my paper on incentives in large economies that characterized strategyproof mechanisms, including the competitive mechanism (Hammond 1979a). What Guesnerie (1981, 1995), in his parallel work, called the "taxation principle" played a crucial role in this characterization. Smoothness conditions that excluded competitive mechanisms with redistributive lump-sum transfers were provided. In an economy with public goods, similar conditions ensured that the only way to get a strategy-proof mechanism to yield Pareto efficient allocations required the public goods to be financed by what later, in the 1980s, became the hated "poll tax" or "community charge" that Mrs. Thatcher introduced as a way of helping to finance local government expenditure. My paper, however, did point out that such a tax could be difficult to collect. I should record that Paul Champsaur and Guy Laroque did express some objections to my treatment of null sets of agents in the result that relied on smoothness conditions (Champsaur and Laroque 1981, 1982). I hope I may have answered them in the chapter (Hammond 2011) on competitive market mechanisms that Arrow, Sen and Suzumura had asked me to write for the *Handbook of Social Choice and Welfare* they edited.

Around that time, I aspired to follow a precedent that Michael Farrell and Frank Hahn had set as managing editors of the *Review of Economic Studies* when they published symposia on specific research topics — namely, the June 1962 "Symposium on Production Functions and Economic Growth", and the January 1967 "Symposium on Optimal Infinite Programmes". In 1974 Geoffrey Heal continued this tradition with an extra issue devoted to a "Symposium on the Economics of Exhaustible Resources". The subject I chose for what became the April 1979 issue was incentive compatibility, a topic for which there were already a number of good papers in the editorial pipeline.

Meanwhile Eric Maskin had been visiting "our" Cambridge while working on his Harvard Ph.D. thesis in Applied Mathematics, supervised by Kenneth Arrow. Partha Dasgupta and I had been discussing some aspects of Eric's work with him. An attempt (Dasgupta et al. 1979) to synthesize many of the results known at that time, and to place them in the context of social choice theory, seemed something worth including in the symposium. Fortunately, the referees agreed. They let us get away, however, with one rather glaring error.

PM: *The paper covers so much ground and has such a wealth of formal results that a slip in it is excusable. What was this error?*

PJH: We had a correct result, essentially due to Hurwicz (1972), showing that in the case of a two-person exchange economy, any strategy-proof mechanism guaranteeing Pareto efficient exchange would have to be dictatorial. But then we claimed that the same would hold with more than two individuals, and referred the reader to Eric's thesis for a result that was related, but failed to offer an adequate proof. Not many years later in the same journal, Mark Satterthwaite and Hugo Sonnenschein came up with a remarkably simple counter-example, which they were kind enough to conceal somewhat in their presentation of interesting new results (Satterthwaite and Sonnenschein 1981). In this counter-example, a third agent's preferences could determine which of the first two agents would be allocated the total endowment in the exchange economy, so neither is a dictator. At least it is an interesting illustration of what difference it makes to add a third player to a game with two players.

There is an important lesson that I like to draw from a particular result in our paper, similar to one which Ledyard (1978) demonstrated at about the same time. First let me start with the methodological observation that the literature on incentive compatibility is intended to deal with situations where the participants in the mechanism or game have incomplete information about each other. In my view, this calls for Harsanyi's (1967, 1968a, b) concept of a "Bayesian" equilibrium, rather than the Nash equilibrium concept that Hurwicz, Maskin and others were using rather extensively.[5] Perhaps Nash equilibrium is appropriate in something like a principal–agent setting for the very special case where all the agents happen to have complete information about each other, but the principal does not. Anyway, in general, the outcome of a game of incomplete information will be sensitive to each player's beliefs about the preferences and beliefs of other players, including their beliefs about which strategies the other players will choose. Indeed, such beliefs and the speculative activity which they engender typically play a key role in determining prices in financial markets, as well as the outcome of a first-price sealed bid auction. Yet welfare economists usually focus only on simple allocation mechanisms, without considering at all how the allocation generated by agents' behaviour in the mechanism may depend on their beliefs. Our simple result showed that the only case where this is generally justified is when one is using a strategy-proof mechanism. This may help justify why my subsequent work has largely focused on these.

Roger Guesnerie and I both moved on to consider mechanisms that remained strategy-proof even in the presence of parallel or shadow markets, where agents could conduct hidden deals on the side. Roger introduced the idea of dividing up private goods into two classes: first, goods like potatoes, cigarettes, wine, or drugs that could be exchanged on such shadow markets; second, goods like electricity, houses, and labour supplied to large firms that get noticed by tax authorities (Guesnerie 1981, 1995). The "multilateral" incentive constraints that arise not only imply that pricing schedules have to be linear, which is not too surprising; in addition, lump-sum redistribution which does not depend on publicly observable individual characteristics such as date of birth can be entirely ruled out without any need to assume smooth preferences. In effect, the possibility of shadow trading places additional constraints

[5] For notable examples, see Hurwicz et al. (1994) and Maskin (1999), as well as several contributions to the Hurwicz *Festschrift* (Groves et al. 1987).

on strategyproof mechanisms, as suggested by the title of my later article on "markets as constraints" (Hammond 1987a).

Finally, while on this subject of mechanisms and market design, it may be appropriate to mention Hammond (2017). This paper's subtitle "Twenty-Two Steps to Walrasian Equilibrium" evokes some of the practical difficulties that must be overcome if one is to get competitive markets to work as standard micro-economic theory says they should, even if one is limited only to spot markets with many traders, whose individual influence over prices is negligible.

PM: *I suppose one could argue that implementation theory provides a foundation for public economics. At least, this is the area where public economics and social choice theory have fruitfully met. Decreasing the abstraction level, we could now touch on public economics proper and discuss the contributions you made to it.*

PJH: Many of my ideas on public economics up to the late 1980s are summarized in my "provocative assessment" of theoretical progress in the subject (Hammond 1990). This was a response to the kind invitation of my Oxford friend Peter Sinclair while he was an editor of *Oxford Economic Papers*. But then, after the Berlin Wall fell in 1989, Edmond Malinvaud came to give some Schumann lectures at the European University Institute (EUI), where I spent a leave of absence from Stanford as a "professore di scienze economiche" (plural) for the two academic years 1989/90 and 1990/91. Malinvaud's lectures concerned the then topical subject of suitable programmes for liberalizing the economies of the Eastern European nations that, up to 1989, had been behind the Iron Curtain.[6] These lectures prompted me to think once again about the possible benefits of markets, especially the gains from trade and other forms of economic liberalization. This topic had interested me enough back in 1975 so that, when Murray Kemp invited me down from Canberra to the University of New South Wales in Sydney in order to give a lunch-time seminar, rather presumptuously I chose to speak about "The Gains from Trade in Imperfect Economies". Meanwhile, I became a supervisor to Jaume Sempere as he was studying for his Ph.D. at EUI. After he had had this thesis accepted, we collaborated on three articles discussing how to convert the potential Pareto gains that Kaldor and Hicks had considered in their compensation tests into ethically acceptable actual Pareto gains (Hammond and Sempere 1995, 2006, 2009).

Our first paper, which appeared in the *Economic Journal*, focused on the gains from trade. It built on a *Journal of International Economics* paper by Jean-Michel Grandmont and Daniel McFadden that extended the two standard textbook examples where trade cannot create any pecuniary externalities in the form of adverse terms-of-trade effects (Grandmont and McFadden 1972). These examples are: first, an exchange economy where the status quo is autarky; second, a "small country" whose exports and imports have no effects on world prices. There was also some discussion of the paper by Dixit and Norman (1986). This considered a Diamond–Mirrlees model of optimal taxation where one could ensure a Pareto improvement by freezing

[6]See http://cadmus.eui.eu//handle/1814/23611 for a record dated 1991 of a distinguished lecture that Malinvaud gave to the Robert Schuman Centre for Advanced Studies at the European University Institute. The title was "Macroeconomic Research and European Policy Formation".

consumer prices and then using commodity taxes or subsidies in order to create a gap between consumer and producer prices that would allow the latter to vary in order to clear markets while also providing resources to fund a uniform lump-sum subsidy that would benefit every consumer.

For our second paper in the *Journal of Public Economic Theory*, we turned attention to the additional potential Pareto gains which could be achieved from the free migration of labour. Kemp (1993) had discussed this earlier, but he treated labour as just another traded good. This ignores the non-convexities that make themselves all too evident if one tries to work simultaneously in more than one country. Or, in the example which Malinvaud presented in his *Lectures on Microeconomic Theory* in the days before high-speed trains, if one tries to eat a dinner one evening that is split between Paris and Lyons (Malinvaud 1969). We were able to overcome these non-convexities by the usual device of assuming a continuum of consumers, following the well-known major works (Aumann 1964, 1966; Hildenbrand 1974), as well as papers by Akira Yamazaki and Ali Khan on general equilibrium theory with non-convex consumption sets and dispersed consumer characteristics (Yamazaki 1978, 1981; Khan and Yamazaki 1981).

That second paper was written while I was in California. There I was well aware of Proposition 187 supported by Governor Pete Wilson and passed by Californian voters in 1994.[7] It was intended to deny the access of "illegal" immigrants to publicly provided services such as schools for their children and, in the case of the poor, access to emergency healthcare. That prompted a third paper, published in *Economic Theory*, that allowed for externalities subject to congestion. This does strike me as a legitimate cause for concern about immigration in case, as happens too often, especially in the UK leading up to the Brexit referendum fiasco, the state fails to make adequate adjustments in the provision of public resources that immigrants and their families need. This despite the additional tax revenue that immigration often generates.

Externalities subject to congestion do, of course, make it much harder to ensure Pareto gains from free migration; indeed, in order to encourage people to make efficient decisions regarding where to live and work, we found ourselves analysing residence charges that are uncomfortably close to the Thatcherite poll taxes that had emerged in my paper on incentives in large economies (Hammond 1979a). Even then Jaume and I were only able to ensure Pareto gains by assuming that migration would be limited to exchanges of population which kept fixed any congestion externalities that would have existed in the status quo allocation, in the absence of any reform. Along with David Lodge's amusing novel about an academic exchange of Professors of English Literature between "Rummidge University" in the English Midlands and "Euphoric State" on the edge of the San Francisco Bay, this inspired the "Changing Places" part of the title. It also does much to explain why the paper became the subject of seminar presentations at both U.C. Berkeley and the University of Birmingham.

PM: *Could we now turn to the vexed topic of cost–benefit analysis (CBA), another area you contributed to?*

[7]Immediately after passage, its key provisions were challenged in federal district court. They were declared unconstitutional before having time to take effect.

PJH: My introduction to it occurred while at Nuffield College, where there was a regular Friday afternoon seminar organized by Ian Little and Jim Mirrlees. That seminar concerned issues arising from trying to put into practice the methods advocated in their Manual published by the OECD (Little and Mirrlees 1969, 1974). There were the competing *Guidelines for Project Evaluation* by Partha Dasgupta, Stephen Marglin, and Amartya Sen (UNIDO 1972). Indeed, my first encounter with Amartya may have been when he came to the Nuffield seminar in order to present a talk on their alternative methodology.[8] Also, when I was at the Australian National University (ANU), I found myself gravitating to the Philosophy Department to attend a seminar on CBA that they had organized. This was most likely as an initiative of John Passmore, a great figure in Australian philosophy who in 1974, as I have since discovered, published a book on environmental ethics (Passmore 1974).

My writing on the subject of CBA began with a contribution to the then annual AUTE (Association of University Teachers of Economics) conference, when it was held in the spring of 1978 at the University of Warwick. The title of the paper is "Cost–Benefit Analysis as a Planning Procedure" (Hammond 1980).[9] The idea was that, properly conducted, a CBA test should be a tool for identifying projects or policy changes that would improve the allocation of resources. This is somewhat along the lines of Malinvaud's "Decentralized Procedures for Planning" that appeared in a 1967 conference volume that he co-edited with Michael Bacharach (Malinvaud 1967). Similar ideas arise in the MDP planning procedure, named after Malinvaud (1971, 1972b) as well as Drèze and de la Vallée Poussin (1971). Another source of inspiration for me was a seminar that I heard Peter Warr present while I was at the ANU (Warr 1977).

A key issue that arises when doing CBA is the choice of shadow prices at which to evaluate the inputs to and outputs from the project being analysed. The UNIDO manual largely advocates using consumers' demand prices. Its approach focuses ideally on how the project, together with any "balancing policies" such as financing the project that are needed to allow the economy as a whole to adjust to it, affect different consumers' well-being (Hammond 1986). If the project is small, one can use their aggregate willingness to pay, with suitable welfare weights for different individuals, to see if the benefits outweigh the costs. Doing this thoroughly, however, even for a project as small as deepening one village well by the odd metre, is likely to be very burdensome. Partha Dasgupta, in his published discussion of my paper at the AUTE conference, was energetic in defending the UNIDO approach, and in noticing some gaps in my background knowledge.

By contrast, the OECD manual by Little and Mirrlees largely favoured using producers' supply prices. In the prominent case of goods that are traded internationally by a small country facing competitive world markets, these could be the prices at which those goods cross the national border. Many non-traded goods, on the other hand, could be priced by looking at input and output relationships in the national

[8] See also Sen (1972a).

[9] See https://web.stanford.edu/~hammond/CostBenefit1978.pdf for a version where the scanned pages appear in numerical order.

economy. Some decades later, while preparing my paper "Reassessing the Diamond–Mirrlees Efficiency Theorem" for the conference whose proceedings appeared in what became the Mirrlees *Festschrift* that Gareth Myles and I co-edited, I realized how this approach based on supply prices could be justified if it was able to identify projects that would really enhance the economy's production possibilities (Hammond 2000). Then, provided that the project is accompanied by other suitable policy measures for distributing the benefits of increased production efficiency to consumers, it could generate Pareto gains. This proviso, of course, often fails to be met, as is illustrated by what has been happening over the last decades to poorer workers who have become the increasingly desperate victims of globalization.

Looking back almost 40 years, using the demand prices that the UNIDO approach emphasizes always works in principle provided that one restricts attention to genuinely small projects, if there are any, where shadow prices are fixed. Yet using demand prices seems unreasonable in requiring not only a finely detailed disaggregated analysis to determine the precise distribution of individual gains and losses, but also a full specification of what "balancing policy" gets used to re-equilibrate the national economy by, for instance, providing adequate finance for the project. On the other hand, whereas the OECD approach based on supply prices surely requires calculations that should be much easier in practice, it is really only ethically justified if one has faith that institutions which supplement the project can ensure that its gains and losses will be appropriately distributed between different individual consumers.

Finally, there is also the associated topic of welfare measurement. Some pieces I wrote during the 1980s on approximate measures (Hammond 1984, 1988d) reflected the fact that I was then rather confused. One benefit of knowing Erwin Diewert, however, was that Wolfgang Eichhorn invited me to a conference in Karlsruhe (Eichhorn 1994). By then I had finally grown to understand and appreciate the power and intuitive appeal of money metric utility functions, which are related to but more useful than Hicks's notions of compensating and equivalent variation that appear in the better textbooks.

I was also familiar with Dale Jorgenson's work on measuring social welfare that had featured in his Presidential Address to the Econometric Society and other articles (Jorgenson 1990, 1997a, b). This relied on concepts like equally or optimally distributed income equivalents, as well as hypotheses about demand behaviour close to allocations that would be reached with equal or optimal income distributions. Yet in our very unequal societies, it is hard to know what this behaviour would be. So in Hammond (1994b), following ideas due to Feldstein (1974) and Rosen (1976), I proposed a uniform money metric measure of social welfare. Its construction depends only at what demand behavior would be near allocations where all individuals have equal increases to their status quo income, which seems much more practical. Also, to support our host Wolfgang Eichhorn in what in the end was clearly his successful effort to secure the generous funding that the conference deserved, he asked us to include some discussion of issues surrounding externalities. In my case I found this relatively easy to do, though it did raise some associated issues about how to generalize the Slutsky conditions to deal with marginal willingness to pay for public goods. These issues, as far as I am aware, still remain unresolved some 25 years later.

PM: *Before we are coming to a close and I thank you for a wonderful treat of ideas and recollections, I would like to put to you a loose, but perhaps important question. Your career began, and some of your achievements indeed took place, in what was still the heyday of economic theory. In the post-war years, neo-classical economics had finally crystallized into elegant theorems on equilibrium and optimality, and entirely new theories of rational action with a strong axiomatic tone had just entered stage—game theory, decision theory, social choice theory, and their dependencies. New formal tools like convex analysis, combinatorics, linear programming or optimization theory, had emerged or become better understood. All this encouraged the mentors and senior colleagues you mentioned in this interview to expect further quick progress as well as perhaps a final unification—the expression in the singular "economic theory" testifies to this hope. You are the heir to this intellectual tradition, which your life-long work has enriched. But already before the turn of the century, the economists' academic interests shifted towards experiments, applications and new disciplinary connections, such as cognitive psychology and behavioural sciences. Today's young theorists have a hard time competing with ambitious millennials who produce more and more empirical research with shallow theoretical foundations. Do you agree with this sombre diagnosis? Do you think there is a future in the profession for the economic theory we discussed in this interview?*

PJH: Now you are asking me to do something even more dangerous than predicting future economic events, which is to predict the progress of economic thought, especially in the area of normative economic theory which has been the subject of this interview. Nevertheless, I am ready to admit that I tend to agree with your diagnosis, though in the end I may disagree with the assessment that it is sombre.

Probably I should start by confessing to some rather extreme views regarding economic methodology, though they are extreme in two opposite directions, which may be as uncomfortable as it would be for one person to try to balance a see-saw by occupying both ends simultaneously. At one end is descriptive or positive economics, which is concerned with how individuals make decisions, and how economic institutions perform, as well as who gets allocated what in an economic system. Here I am an extremist to the extent of refusing to accept, at least in the absence of convincing supporting evidence, the standard neoclassical hypothesis that economic agents behave rationally.

At the other extreme is theory that is purely normative or prescriptive, setting out principles that should guide economic policy, or decision-making more generally. Here I go to the other extreme in saying that the decision-maker should choose policies and actions that they can defend because they can be expected to have good consequences. As I claimed in an invited address to last year's meeting of the European Society for the History of Economic Thought whose theme was "rationality", one should avoid any unnecessary possibility of creating regrettable consequences. This, by the way, is intended to recognize that some regret may be necessary because of our inability to have a complete model of all the consequences that may result from those decisions that we must make right now. It turns out that planning to avoid regrettable consequences is equivalent to dynamic consistency. There is some relationship here to Gilboa and Schmeidler's view that rationality requires avoiding regret over the axioms one chooses to satisfy; I have much more concern, however, for the practical consequences of economic decisions than for the logical "consequences" (i.e., implications) of alternative axiom systems (Gilboa and Schmeidler 2001; Gilboa 2010, 2015). Anyway, this kind of rationality based on avoiding regrettable consequences leads rather directly to subjective expected utility as a decision criterion, with some modifications that I am still exploring in an attempt to recognize that any model we use to guide our decision-making involves bounds that prevent our noticing possibilities whose neglect we may later regret.

When discussing the future of our discipline, we should admit the possibility of diminishing returns to some kinds of further research. Certainly young researchers trying to establish themselves in academic economics may be well advised to focus on better descriptive models. So I welcome progress in behavioural economics that attempts to improve our understanding of how real people form their tastes and beliefs, and of how they make decisions, including in laboratory experiments. Of course, the search for a coherent theoretical framework is important, as is looking for psychologically realistic ways of advising people how to make better decisions. Indeed, one reason that my move to Warwick in 2007 has proved personally beneficial is that it created more scope for me to explore some of these topics.

I would like to end, however, with what may be a salutary lesson. One reaction to the financial meltdown in 2007–8 was that economists had neglected to construct appropriate models that build in some understanding of what really happens in messy real financial markets.[10] The hope seemed to be that such models could help us predict financial markets better, and perhaps head off the next crisis. Yet to me, thinking as a normative economist, this may be fruitless. Instead, rather than trying to model the existing mess better, we should at least start thinking about how to regulate and even re-design financial markets so that their behaviour becomes much easier to understand. Then, perhaps, we can hope to control them well enough to ward off unnecessary threats of disaster.

Finally, let me conclude this last response by expressing my profound thanks to you for the wonderfully well-informed and probing questions that you have asked me. Attempting to answer them has been a most welcome intellectual challenge.

Acknowledgements The two of us thank Marc Fleurbaey for his initiative, encouragement, and forbearance, as well as for helpful editorial comments on an earlier draft.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica*, *21*(4), 503–546.

Anscombe, F. J., & Aumann, R. J. (1963). A definition of subjective probability. *Annals of Mathematical Statisics*, *34*(1), 199–205.

[10] See especially the papers in *Nature* by Bouchaud (2008) as well as Farmer and Foley (2009).

Arrow, K. (1950). A difficulty in the concept of social welfare. *Journal of Political Economy*, *58*(4), 328–346.

Arrow, K. J. (1951a). *Social choice and individual values*. New Haven: Yale University Press, 2nd edn. 1963.

Arrow, K. J. (1951b). Alternative approaches to the theory of choice in risk-taking situations. *Econometrica*, *19*(4), 404–437.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, *26*(102), 121–127.

Arrow, K. J. (1977). Extended sympathy and the possibility of social choice. *American Economic Review*, *67*(1), 219–225.

Arrow, K. J., & Debreu, G. (1954). Existence of an equilibrium for a competitive economy. *Econometrica*, *22*(3), 265–290.

Aumann, R. J. (1964). Markets with a continuum of traders. *Econometrica*, *32*(1), 39–50.

Aumann, R. J. (1966). Existence of a competitive equilibrium in markets with a continuum of traders. *Econometrica*, *34*(1), 1–17.

Blackorby, C., Bossert, W., & Donaldson, D. (2005). *Population issues in social choice theory*. Cambridge: Cambridge University Press.

Blackorby, C., Nissen, D., Primont, D., & Russell, R. R. (1973). Consistent intertemporal decision making. *Review of Economics Studies*, *40*(2), 239–248.

Bouchaud, J. P. (2008). Economics needs a scientific revolution. *Nature*, *455*(7217), 1181.

Broome, J. (1990). Fairness. *Proceedings of the Aristotelian Society*, *91*, 87–101.

Broome, J. (1991). *Weighing goods*. Oxford: Blackwell.

Champsaur, P., & Laroque, G. (1981). Fair allocations in large economies. *Journal of Economic Theory*, *25*(2), 269–282.

Champsaur, P., & Laroque, G. (1982). A note on incentives in large economies. *Review of Economic Studies*, *49*(4), 627–635.

Chichilnisky, G., Hammond, P., & Stern, N. H. (2018). Should we discount the welfare of future generations? Ramsey and Suppes versus Koopmans and Arrow. In *Warwick economics research papers series (WERPS)(1174)*, http://wrap.warwick.ac.uk/107726/.

Dalton, H. (1920). The measurement of the inequality of incomes. *Economic Journal*, *30*, 348–361.

Dasgupta, P. S., & Hammond, P. J. (1980). Fully progressive taxation. *Journal of Public Economics*, *13*, 141–154.

Dasgupta, P. S., Hammond, P. J., & Maskin, E. S. (1979). The implementation of social choice rules: Some general results on incentive compatibility. *Review of Economic Studies*, *46*, 185–216.

D'Aspremont, C., & Gevers, L. (1977). Equity and the informational basis of collective choice. *Review of Economic Studies*, *44*(2), 199–209.

D'Aspremont, C., & Hammond, P. J. (2001). An interview with John C. Harsanyi. *Social Choice and Welfare*, *18*, 389–401.

Debreu, G. (1959). *Theory of value: An axiomatic analysis of economic equilibrium*. New Haven, CT: Yale University Press.

De Finetti, B. (1937). La prévision: ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, *7*, 1–68.

Deschamps, R., & Gevers, L. (1978). Leximin and utilitarian rules: A joint characterization. *Journal of Economic Theory*, *17*(2), 143–163.

Diamond, P. A. (1967a). Cardinal welfare, individualistic ethics, and interpersonal comparison of utility: Comment. *Journal of Political Economy*, *75*, 765–766.

Diamond, P. A. (1967b). The role of a stock market in a general equilibrium model with technological uncertainty. *American Economic Review*, *57*(4), 759–776.

Diamond, P. A., & Mirrlees, J. A. (1971). Optimal taxation and public production I: Production efficiency. *American Economic Review*, *61*(1), 8–27.

Dixit, A. K., & Norman, V. (1986). Gains from trade without lump-sum compensation. *Journal of International Economics*, *21*(1–2), 99–110.

Drèze, J. (1970). Market allocation under uncertainty. *European Economic Review*, *2*, 133–165.

Drèze, J., & de la Vallée Poussin, D. (1971). A tâtonnement process for public goods. *Review of Economic Studies*, *38*(2), 133–150.
Eichhorn, W. (Ed.). (1994). *Models and measurement of welfare and inequality*. Berlin: Springer.
Farmer, J. D., & Foley, D. K. (2009). The economy needs agent-based modeling. *Nature*, *460*(7256), 685–6.
Feldstein, M. S. (1974). Distributional preferences in public expenditure analysis. In H. Hochman & G. Peterson (Eds.), *Redistribution through public choice* (pp. 136–161). New York: Columbia University Press.
Fleurbaey, M., & Hammond, P. J. (2004). Interpersonally comparable utility. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory* (Vol. 2, pp. 1181–1285)., Extensions Boston, MA: Kluwer Academic Publishers.
Fleurbaey, M., & Maniquet, F. (2011). *A theory of fairness and social welfare*. Cambridge: Cambridge University Press.
Fleurbaey, M., & Maniquet, F. (2017). Fairness and well-being measurement. *Mathematical Social Sciences*, *90*, 119–126.
Fleurbaey, M., & Maniquet, F. (2018). Optimal income taxation theory and principles of fairness. *Journal of Economic Literature*, *56*(3), 1029–1079.
Fleurbaey, M., & Mongin, P. (2005). The news of the death of welfare economics is greatly exaggerated. *Social Choice and Welfare*, *25*, 381–418.
Gibbard, A. S. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*, 587–601.
Gilboa, Itzhak. (2010). Questions in decision theory. *Annual Reviews in Economics*, *2*, 1–19.
Gilboa, I. (2015). Rationality and the Bayesian paradigm. *Journal of Economic Methodology*, *22*(3), 312–334.
Gilboa, I., & Schmeidler, D. (2001). *A theory of case-based decisions*. Cambridge: Cambridge University Press.
Graaff, J. V. (1957). *Theoretical welfare economics*. Cambridge: Cambridge University Press.
Grandmont, J.-M., & McFadden, D. (1972). A technical note on classical gains from trade. *Journal of International Economics*, *2*(2), 109–125.
Groves, T., Radner, R., & Reiter, S. (1987). *Information, incentives, and economic mechanisms: Essays in honor of Leonid Hurwicz*. Minneapolis: University of Minnesota Press.
Guesnerie, R. (1981). *On taxation and incentives: Further reflections on the limits of redistribution*. Discussion Paper No. 89, Sonderforschungsbereich 21, University of Bonn.
Guesnerie, R. (1995). *A contribution to the pure theory of taxation*. Cambridge: Cambridge University Press.
Hammond, P. J. (1973). *Consistent planning and intertemporal welfare economics*. Ph.D. thesis, University of Cambridge.
Hammond, P. J. (1975a). Agreeable plans with many capital goods. *Review of Economic Studies*, *42*, 1–14.
Hammond, P. J. (1975b). A note on extreme inequality aversion. *Journal of Economic Theory*, *11*, 465–467.
Hammond, P. J. (1976a). Equity, Arrow's conditions, and Rawls' difference principle. *Econometrica*, *44*, 793–804.
Hammond, P. J. (1976b). Changing tastes and coherent dynamic choice. *Review of Economic Studies*, *43*, 159–173.
Hammond, P. J. (1977). Dynamic restrictions on metastatic choice. *Economica*, *44*, 337–350.
Hammond, P. J. (1979a). Straightforward individual incentive compatibility in large economies. *Review of Economic Studies*, *46*, 263–282.
Hammond, P. J. (1979b). Equity in two-person situations: Some consequences. *Econometrica*, *47*, 1127–1135.
Hammond, P. J. (1980). Cost-benefit analysis as a planning procedure. In Currie, D. A., & Peters, W. (Eds.), *Contemporary economic analysis (vol. 2,pp. 221–250), Proceedings of the conference of the association of university teachers of economics*, 1978. London: Croom-Helm.

Hammond, P. J. (1981). Ex-ante and ex-post welfare optimality under uncertainty. *Economica*, *48*, 235–250.
Hammond, P. J. (1982). Utilitarianism, uncertainty and information. In A. K. Sen & B. Williams (Eds.), *Utilitarianism and beyond* (pp. 85–102). Cambridge: Cambridge University Press.
Hammond, P. J. (1983). Ex-post optimality as a dynamically consistent objective for collective choice under uncertainty. In P. K. Pattanaik & M. Salles (Eds.), *Social choice and welfare* (pp. 175–205). Amsterdam: North-Holland.
Hammond, P. J. (1984). Approximate measures of social welfare and the size of tax reform. In D. Bös, M. Rose, & C. Seidl (Eds.), *Beiträge zur neueren Steuertheorie* (pp. 95–115). Berlin: Springer.
Hammond, P. J. (1986). Project evaluation by potential tax reform. *Journal of Public Economics*, *30*, 1–36.
Hammond, P. J. (1987a). Markets as constraints: Multilateral incentive compatibility in continuum economies. *Review of Economic Studies*, *54*, 399–412.
Hammond, P. J. (1987b). Social choice: The science of the impossible? In G. R. Feiwel (Ed.), *Arrow and the foundations of the theory of economic policy* (pp. 116–131). Macmillan: New York University Press.
Hammond, P. J. (1988a). Consequentialist foundations for expected utility. *Theory and Decision*, *25*, 25–78.
Hammond, P. J. (1988b). Consequentialism and the independence axiom. In Munier, B. R. (Ed.), *Risk, decision and rationality*, Proceedings of the 3rd international conference on the foundations and applications of utility, risk and decision theories (pp. 503–516). Dordrecht: D. Reidel.
Hammond, P. J. (1988c). Consequentialist demographic norms and parenting rights. *Social Choice and Welfare*, *5*, 127–145.
Hammond, P. J. (1988d). Principles for evaluating public sector projects. In P. Hare (Ed.), *Surveys in public sector economics* (pp. 15–44). Oxford: Basil Backwell.
Hammond, P. J. (1989). Some assumptions of contemporary neoclassical economic theology. In G. R. Feiwel (Ed.), *Joan Robinson and modern economic theory* (pp. 186–257). Macmillan: New York University Press.
Hammond, P. J. (1990). Theoretical progress in public economics: A provocative assessment. *Oxford Economic Papers*, *42*, 6–33.
Hammond, P. J. (1991a). Interpersonal comparisons of utility: Why and how they are and should be made. In J. Elster & J. E. Roemer (Eds.), *Interpersonal comparisons of well-being* (pp. 200–254). Cambridge: Cambridge University Press.
Hammond, P. J. (1991b). Independence of irrelevant interpersonal comparisons. *Social Choice and Welfare*, *8*, 1–19.
Hammond, P. J. (1992). Harsanyi's utilitarian theorem: A simpler proof and some ethical connotations. In R. Selten (Ed.), *Rational interaction: Essays in honor of John Harsanyi* (pp. 305–319). Berlin: Springer-Verlag.
Hammond, P. J. (1993). Credible liberalization: Beyond the three theorems of neoclassical welfare economics. In D. Bös (Ed.), *Economics in a changing world* (Vol. 3, pp. 21–39)., Public policy and economic organization London: Macmillan.
Hammond, P. J. (1994a). Elementary non-archimedean representations of probability for decision theory and games. In P. Humphreys (Ed.), *Patrick Suppes: Scientific philosopher* (Vol. I, pp. 25–59)., Probability and probabilistic causality New York: Kluwer Academic Publishers.
Hammond, P. J. (1994b). Money metric measures of individual and social welfare allowing for environmental externalities. In W. Eichhorn (Ed.), *Models and measurement of welfare and inequality* (pp. 694–724). Berlin: Springer.
Hammond, P. J. (1995). Social choice of individual and group rights. In W. A. Barnett, H. Moulin, M. Salles, & N. Schofield (Eds.), *Social choice, welfare, and ethics* (pp. 55–77). Cambridge: Cambridge University Press.

Hammond, P. J. (1996a). Consequentialism, structural rationality and game theory. In K. J. Arrow, E. Colombatto, M. Perlman, & C. Schmidt (Eds.), *The rational foundations of economic behaviour* (pp. 25–42). London: Macmillan.
Hammond, P. J. (1996b). Consequentialist decision theory and utilitarian ethics. In F. Farina, F. Hahn, & S. Vannucci (Eds.), *Ethics, rationality, and economic behaviour* (pp. 92–118). Oxford: Clarendon Press.
Hammond, P. J. (1997). Game forms versus social choice rules as models of rights. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Social choice re-examined* (Vol. II, pp. 82–95). London: Macmillan.
Hammond, P. J. (1998a). Objective expected utility: A consequentialist perspective. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of Utility Theory* (Vol. 1, pp. 145–211)., Principles Boston, MA: Kluwer Academic Publishers.
Hammond, P. J. (1998b). Subjective expected utility. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory* (Vol. 1, pp. 213–271)., Principles Boston, MA: Kluwer Academic Publishers.
Hammond, P. J. (1999a). Subjectively expected state-independent utility on state-dependent consequence domains. In M. J. Machina & B. Munier (Eds.), *Beliefs, interactions, and preferences in decision making* (pp. 7–21). Dordrecht: Kluwer Academic.
Hammond, P. J. (1999b). Consequentialism, non-archimedean probabilities, and lexicographic expected utility. In C. Bicchieri, R. Jeffrey, & B. Skyrms (Eds.), *The logic of strategy* (pp. 39–66). Oxford: Oxford University Press.
Hammond, P. J. (1999c). Non-archimedean subjective probabilities in decision theory and games. *Mathematical Social Sciences*, *38*, 139–156.
Hammond, P. J. (1999d). *Roberts' weak welfarism theorem: A minor correction*. Working Papers 99021, Stanford University, Department of Economics; available at https://ideas.repec.org/p/wop/stanec/99021.html.
Hammond, P. J. (2000). Reassessing the Diamond-Mirrlees efficiency theorem. In P. J. Hammond & G. D. Myles (Eds.), *Incentives, organization, and public economics: Papers in honour of Sir James Mirrlees* (pp. 193–216). Oxford: Oxford University Press.
Hammond, P. J. (2011). Competitive market mechanisms as social choice procedures. In K. J. Arrow, A. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare* (Vol. II, pp. 47–151). Amsterdam: North-Holland.
Hammond, P. J. (2017). Designing a strategyproof spot market mechanism with many traders: Twenty-two steps to Walrasian equilibrium. *Economic Theory*, *63*(1), 1–50.
Hammond, P. J., & Mirrlees, J. A. (1973). Agreeable plans. In J. A. Mirrlees & N. H. Stern (Eds.), *Models of economic growth* (pp. 283–299). London: Macmillan.
Hammond, P. J., & Sempere, J. (1995). Limits to the potential gains from economic integration and other supply side policies. *Economic Journal*, *105*, 1180–1204.
Hammond, P. J., & Sempere, J. (2006). Gains from Trade versus gains from migration: What makes them so Different? *Journal of Public Economic Theory*, *8*, 145–170.
Hammond, P. J., & Sempere, J. (2009). Migration with local public goods and the gains from changing places. *Economic Theory*, *41*, 359–377.
Hare, R. (1952). *The language of morals*. Oxford: Clarendon Press.
Hare, R. (1963). *Freedom and reason*. Oxford: Clarendon Press.
Hare, R. (1981). *Moral thinking: Its levels, method, and point*. Oxford: Clarendon Press.
Harrison, M. (2017). The soviet economy, 1917–1991: Its life and afterlife. https://warwick.ac.uk/fac/soc/economics/staff/mharrison/public/2017_independent_review_preprint.pdf.
Harsanyi, J. C. (1953). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, *61*(5), 434–435.
Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*(4), 309–321.
Harsanyi, J. C. (1967). Games with incomplete information played by 'Bayesian' Players, I-III: Part I. The basic model. *Management Science*, *14*(3), 159–182.

Harsanyi, J. C. (1968a). Games with incomplete information played by 'Bayesian' Players, Part II. Bayesian equilibrium points. *Management Science*, *14*(5), 320–334.

Harsanyi, J. C. (1968b). Games with incomplete information played by 'Bayesian' Players, Part III. The basic probability distribution of the game. *Management Science*, *14*(7), 486–502.

Harsanyi, J. C. (1974). On some problems arising from Professor Rawls conception of distributive justice. *Theory and Decision*, *4*(3–4), 325–344.

Harsanyi, J. C. (1976). *Essays on ethics, social behaviour, and scientific explanation*. Dordrecht: Reidel.

Harsanyi, J. C. (1977). *Rational behavior and bargaining equilibrium in games and social situations*. Cambridge: Cambridge University Press.

Harsanyi, J. C. (1987). Interpersonal utility comparison. In Eatwell, J., Milgate, M., & Newman, P. (Eds.), *The new palgrave dictionary of economics*. London: Macmillan. (Theory and Decision Library by John C. Harsanyi (2013-10-04) Paperback – 1722).

Hildenbrand, W. (1974). *Core and equilibria of a large economy*. Princeton: Princeton University Press.

Hurwicz, L. (1972). On informationally decentralized systems. In C. B. McGuire & R. Radner (Eds.), *Decision and organization*. Amsterdam: North Holland.

Hurwicz, L. (1973). The design of mechanisms for resource allocation. *American Economic Review*, *63*(2), 1–30.

Hurwicz, L., Maskin, E. S., & Postlewaite, A. (1994). Feasible nash implementation of social choice rules when the designer does not know endowments or production sets. In J. O. Ledyard (Ed.), *The economics of informational decentralization: Complexity, efficiency, and stability: Essays in honor of Stanley Reiter* (pp. 367–433). New York: Springer Science + Business Media.

Jorgenson, D. W. (1990). Aggregate consumer behavior and the measurement of social welfare. *Econometrica*, *58*, 1007–1040.

Jorgenson, D. W. (1997a). *Welfare, Vol. 1: Aggregate consumer behavior*. Cambridge, MA: MIT Press.

Jorgenson, D. W. (1997b). *Welfare—Vol. 2: Measuring social welfare*. Cambridge, MA: MIT Press.

Kaplow, L., & Shavell, S. (2001). Any non-welfarist method of policy assessment violates the Pareto principle. *Journal of Political Economy*, *109*(2), 281–286.

Kaplow, L., & Shavell, S. (2002). *Fairness versus welfare*. Cambridge, MA: Harvard University Press.

Kaplow, L., & Shavell, S. (2003). Fairness versus welfare: Notes on the Pareto principle, preferences, and distributive justice. *Journal of Legal Studies*, *32*, 331–362.

Kemp, M. C. (1993). The welfare gains from international migration. *Keio Economic Studies*, *30*(1), 1–5.

Khan, M. A., & Yamazaki, A. (1981). On the cores of economies with indivisible commodities and a continuum of traders. *Journal of Economic Theory*, *24*(2), 218–225.

Kolm, S.-C. (1971, translated 1987). *Justice et Équité*. Paris: CEPREMAP; *Justice and equity*. Cambridge, MA: MIT Press.

Kolm, S.-C. (1994). The meaning of 'fundamental preferences'. *Social Choice and Welfare*, *11*(3), 193–198.

Laffont, J.-J. (Ed.). (1979). *Aggregation and revelation of preferences*. Amsterdam: Elsevier Science.

Ledyard, J. O. (1978). Incentive compatibility and incomplete information. *Journal of Economic Theory*, *18*(1), 171–189.

Lenin, V. I. (1917). *The state and revolution*.

Lipsey, R. G., & Lancaster, K. J. (1956). The general theory of second best. *Review of Economic Studies*, *24*(1), 11–32.

Little, I. M. D., & Mirrlees, J. A. (1969). *Manual of industrial project analysis for developing countries*. Paris: OECD Development Centre.

Little, I. M. D., & Mirrlees, J. A. (1974). *Project appraisal and planning for developing countries*. London: Heinemann Educational Books.

Machina, M. J. (1989). Dynamic consistency and non-expected utility models of choice under uncertainty. *Journal of Economic Literature*, *27*(4), 1622–1668.
Malinvaud, E. (1967). Decentralized procedures for planning. In E. Malinvaud & M. O. L. Bacharach (Eds.), *Activity analysis in the theory of growth and planning* (pp. 170–208). London: Macmillan.
Malinvaud, Edmond. (1969, translated 1972). *Leçons de théorie microéconomique*. Paris: Dunod; *Lectures on microeconomic theory*. Amsterdam: North-Holland.
Malinvaud, E. (1971). A planning approach to the public good problem. *Swedish Journal of Economics*, *73*(1), 96–112.
Malinvaud, E. (1972a). The allocation of individual risks in large markets. *Journal of Economic Theory*, *4*(2), 312–328.
Malinvaud, E. (1972b). Prices for individual consumption, quantity indicators for collective consumption. *Review of Economic Studies*, *39*, 385–405.
Malinvaud, E. (1973). Markets for an exchange economy with individual risks. *Econometrica*, *41*(3), 383–410.
Marx, K. (1875). *Critique of the Gotha Program, Section I.*
Maskin, E. S. (1999). Nash equilibrium and welfare optimality. *Review of Economic Studies*, *66*(1), 23–38.
McLean, I., & Hewitt, F. (1994). *Condorcet: Foundations of social choice and political theory*. Cheltenham: Edward Elgar.
McLean, I., & Urken, A. (1995). *Classics of social choice: Pioneering contributions to social choice and voting from Pliny to Lewis Carroll*. Michigan: University of Michigan Press.
McLennen, E. F. (1990). *Rationality and dynamic choice: Foundational explorations*. Cambridge: Cambridge University Press.
Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, *38*(2), 175–208.
Mirrlees, J. A. (1997). Information and incentives: The economics of carrots and sticks. *Economic Journal*, *107*(444), 1311–1329.
Mirrlees, J. A., & Stern, N. H. (Eds.). (1974). *Models of economic growth*. London: Macmillan.
Mongin, P. (1995). Consistent Bayesian aggregation. *Journal of Economic Theory*, *66*(2), 313–351.
Mongin, P. (2001). The impartial observer theorem of social ethics. *Economics and Philosophy*, *17*, 147–80.
Murakami, Y. (1968). *Logic and social choice*. London: Routledge & Kegan Paul.
Myerson, R. (1985). Bayesian equilibrium and incentive compatibility: An introduction. In L. Hurwicz, D. Schmeidler, & H. Sonnneschein (Eds.), *Social goals and social organization: Essays in memory of Elisha A. Pazner* (pp. 229–259). Cambridge: Cambridge University Press.
Passmore, J. (1974). *Man's responsibility for nature: Ecological problems and western traditions*. New York: Scribner.
Pattanaik, P. K. (1971). *Voting and collective choice*. Cambridge: Cambridge University Press.
Pattanaik, P. K. (1968). Risk, impersonality, and the social welfare function. *Journal of Political Economy*, *76*(6), 1152–1169.
Pigou, A. C. (1920). *The economics of welfare*. London: Macmillan.
Rawls, J. (1958). Justice as fairness. *Philosophical Review*, *67*(2), 164–194.
Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.
Robbins, L. (1938). Interpersonal comparisons of utility: A comment. *Economic Journal*, *48*(192), 635–641.
Roberts, K. W. S. (1980a). Possibility theorems with interpersonally comparable welfare levels. *Review of Economic Studies*, *47*(2), 409–420.
Roberts, K. W. S. (1980b). Interpersonal comparability and social choice theory. *Review of Economic Studies*, *47*(2), 421–439.
Rosen, H. S. (1976). A methodology for evaluating tax reform proposals. *Journal of Public Economics*, *6*, 105–122.
Samuelson, P. A. (1947). *Foundations of economic analysis*. Cambridge, MA: Harvard University Press.

Satterthwaite, M. A., & Sonnenschein, H. (1981). Strategy-proof allocation mechanisms at differentiable points. *Review of Economic Studies*, *48*(4), 587–597.

Savage, L. J. (1954). *Foundations of statistics*. New York: Wiley.

Sen, A. K. (1970a). *Collective choice and social welfare*. San Francisco: Holden Day.

Sen, A. K. (1970b). The impossibility of a Paretian liberal. *Journal of Political Economy*, *78*(1), 152–157.

Sen, A. K. (1970c). Interpersonal aggregation and partial comparability. *Econometrica*, *38*(3), 393–409.

Sen, A. K. (1972a). Control areas and accounting prices: An approach to economic evaluation. *Economic Journal*, *82*(325), 486–501.

Sen, A. K. (1972b). Interpersonal aggregation and partial comparability: A correction. *Econometrica*, *40*(5), 959.

Sen, A. K. (1973). *On economic inequality*. Oxford: Clarendon Press.

Sen, A. K. (1985). *Commodities and capabilities*. Amsterdam: North-Holland.

Sen, A. K., & Williams, B. (Eds.). (1982). *Utilitarianism and beyond*. Cambridge: Cambridge University Press.

Starr, R. (1973). Optimal production and allocation under uncertainty. *Quarterly Journal of Economics*, *87*, 81–95.

Strasnick, S. (1976). Social choice and the derivation of Rawls's difference principle. *Journal of Philosophy*, *73*(4), 85–99.

Strotz, R. H. (1956). Myopia and inconsistency in dynamic utility maximization. *Review of Economic Studies*, *23*(3), 165–180.

Suppes, P. C. (1966). Some formal models of grading principles. *Synthese*, *16*, 284–306.

UNIDO. (1972). *Guidelines for project evaluation*. Dasgupta, P., Marglin S., & Sen, A. K. New York: United Nations.

Vickrey, W. S. (1945). Measuring marginal utility by reactions to risk. *Econometrica*, *13*, 319–333.

Vickrey, W. S. (1960). Utility, strategy, and social decision rules. *Quarterly Journal of Economics*, *74*(4), 507–535.

Vickrey, W. S. (1964). *Metastatics and macreconomics*. New York: Harcourt, Brace & World.

Von Neumann, J. (1928). Zur Theorie der Gesellschaftsspiele. *Mathematische Annalen*, *100*, 295–320.

Von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.

Warr, P. G. (1977). On the shadow pricing of traded commodities. *Journal of Political Economy*, *85*(4), 865–872.

Yamazaki, A. (1978). An equilibrium existence theorem without convexity assumptions. *Econometrica*, *46*, 541–555.

Yamazaki, A. (1981). Diversified consumption characteristics and conditionally dispersed endowment distribution: Regularizing effect and existence of equilibria. *Econometrica*, *49*, 639–645.

Prasanta K. Pattanaik

Taradas Bandyopadhyay and Yongsheng Xu

The interview was conducted on April 30, 2015, in the office of Prasanta Pattanaik (PKP) at the University of California, Riverside, by Taradas Bandyopadhyay and Yongsheng Xu (BX).

BX: How did you get into economics?

PKP: There were no clear reasons for my choosing economics. In my time, economics was not being taught in Indian schools. So when I completed my school education, I had no clear idea about what economics was like. When I joined the college, I first opted for mathematics, logic, and Sanskrit. But, following a suggestion of my maternal grandfather, I switched from Sanskrit to economics.

BX: So you declared major in Economics.

PKP: We did not have the system of choosing a major in the first two years of college. In those two years, we had to study three different elective subjects in addition to the mandatory courses on English literature and Odia literature (Odia is my language). It is only in our 3rd and 4th years of college that we had to choose a subject to specialize in. Mathematics, logic, and economics were my elective subjects in the first two years of college. In the 3rd and 4th years of college, I chose to study economics as my subject of what we called the "Honours" program.

BX: So you had some training in logic in your earlier education.

PKP: Yes, I had a good education in basic logic. It was all traditional logic and not modern logic. But Gopal Patnaik, the professor, who taught me logic, was a wonderful teacher, and I learned much from his course on logic.

T. Bandyopadhyay
University of California Riverside, Riverside, USA
e-mail: taradas.bandyopadhyay@ucr.edu

Y. Xu (✉)
Georgia State University, Atlanta, USA
e-mail: yxu3@gsu.edu

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_13

BX: Did that help you in future when you engaged in research in axiomatic social choice theory?

PKP: Actually, before I started studying social choice theory, I had taught myself some logic beyond what I had studied in my undergraduate classes, though I did not know its usefulness for my future research. After I enrolled in the Masters program in the Delhi School of Economics, I came across a copy of Alfred Tarski's *Introduction to Logic and to the Methodology of Deductive Sciences* (Oxford University Press, Oxford, 1941) in a used book shop. I found it very interesting and studied it. So I had some understanding of modern logic when I started my work on social choice theory. Subsequently, I studied several other books on logic, including Patrick Suppes' *Introduction to Logic* (D. Van Nostrand Company, New York, 1957).

BX: So basically apart from that class, you taught yourself logic by reading different books.

PKP: That is right.

BX: Did you take any classes in philosophy during your undergraduate studies?

PKP: I never took a formal course in philosophy, but I had strong interest in philosophy. Until I went to the Delhi School of Economics, I had read mainly books on Hindu and Buddhist philosophy. My father had a fairly large collection of books which included many volumes on Indian philosophy. As an undergraduate, I read most of them, sometimes without understanding them well. I recall one of the books - *Eastern Religions and Western Thought* (Clarendon Press, 1940) by S. R. Radhakrishnan, the second President of India. It had several chapters on early western philosophical thinking. That was probably my earliest exposure to western philosophy.

BX: So, you did not take any philosophy course in college. You came to Delhi in 1963 and in 1968 JPE paper you demonstrated your knowledge and appreciation of moral philosophy. Did you take any class in moral philosophy during your graduate studies at the Delhi school?

PKP: I did much of my study of moral philosophy as a graduate student in Delhi. It is there that I read books such as R. M. Hare's *Language of Morals* (Oxford University Press, Oxford, 1952) and *Freedom and Reason* (Oxford University Press, Oxford, 1963), P. H. Nowell-Smith's *Ethics* (Penguin Books, Harmondsworth, Middlesex, England, 1954), and C. J. Stevenson's *Ethics and Language* (Yale University Press, New Haven, 1944). In particular, I was fascinated by how many moral philosophers analyzed the structure of the moral language involving terms such as "good", "right", etc.

BX: This was when you were doing your masters in the Delhi school.

PKP: I read most of these books when I was doing my Ph.D. As a part of my M. A. course on microeconomic theory, I had studied I. M. D. Little's *A Critique of Welfare Economics* (Oxford University Press, Oxford, 1950), where Little discusses the language of welfare economics. That first aroused my interest in the structure of moral language.

BX: And the structures that you had in the Delhi School helped. You didn't have semester or quarter system and so you had more flexibility in arranging your time.

PKP: Absolutely. It was more like the old British system which gave you a lot of freedom to read whatever you liked. There were examinations that you had to pass, but those examinations were not in every quarter or anything like that.

BX: You did not have to take classes?

PKP: We had to take classes for our M.A. But we were not continuously under the pressure of examinations, such as a quiz every week, two examinations for each subject in each term, and things like that. We did not have that pressure.

BX: That helped, actually.

PKP: That helped a lot.

BX: After doing your masters, you taught in an undergraduate college for a few months.

PKP: Yes, for two months or so.

BX: How come you decided to do a Ph.D. in welfare economics?

PKP: It came as a mixture of choice and accident. I was interested in three subjects. They were microeconomic theory, trade theory, and portfolio choice. We were lucky to have some outstanding teachers. One was Amartya Sen. He joined the Delhi School of Economics in 1963, the year in which I joined the Delhi School of Economics as an M.A. student. Another outstanding teacher was Jagdish Bhagwati. I was keenly interested in both trade theory and microeconomic theory. I did not have any specific interest in welfare economics to start with. It was under Amartya Sen's direction that I developed my interest in welfare economics. My choice of welfare economics was partly accidental and partly a matter of conscious choice.

BX: So you taught for two months. And within those two months, you made up your mind to go to the Ph.D. program or you already knew?

PKP: Even before leaving my teaching job in an undergraduate college, I had registered for the Ph.D. program. What happened was that I took up this job as a fallback to support myself as a Ph.D. student in case I would not get a fellowship to do my Ph.D. There was only one fellowship for Ph.D. students and I could not be sure of getting it. As soon as I was given a fellowship in the Delhi School of Economics, I resigned from my teaching job. But, since I had not followed certain administrative procedures in applying for the fellowship, I faced some amount of trouble in the college where I was first teaching. Apparently, there were some rules that, if you were employed by any college of the university, then, to apply for a fellowship, you needed to get permission from the college first. But I had no idea about those rules and had applied for a fellowship without any permission from my college. There was some problem about that, but it got resolved.

BX: So when you applied for the Ph.D. program at that time you knew that you would work in welfare economics?

PKP: Yes. By that time, I had made up my mind.

BX: So, for the Ph.D., you did not have to take any classes and just to write some papers on the areas that you had interests, right?

PKP: That is right. After your M. A., you had to choose a topic for your Ph.D. dissertation and you had to work on your own chosen topic. There was no course work for Ph.D.

BX: So you started to write your dissertation once entering into the program.

PKP: I had to do some reading in welfare economics first because, by the time I decided that I would work in this area, I had not read much beyond contributions such as Hla Myint's *Theories of Welfare Economics* (Harvard University Press, Cambridge, 1948), the first essay in T. C. Koopmans' *Three Essays on the State of Economic Science* (McGraw-Hill, 1957), and I. M. D. Little's *A Critique of Welfare Economics* (Clarendon Press, Oxford, 1950).

BX: And Graaff?

PKP: Yes, I had also studied J. de V. Graaff's *Theoretical Welfare Economics* (Cambridge University Press, Cambridge, 1957) and various articles on Pareto optimality, the two optimality theorems, compensation principles, externalities, public goods, and so on.

BX: At that stage, (modern) social choice was relatively new, was born just a few years earlier.

PKP: Yes, in 1965 it was still quite new. There were some papers on social choice theory but not very many.

BX: Arrow's book was not available at that time in Delhi?

PKP: It was available. All the journals that I was interested in were also available. But Arrow's theorem was yet to get into the standard curriculum. So, when I did my M.A., there was little about Arrow's theorem in our curriculum. We studied mainly the two optimality theorems, the compensation criteria, externalities, public goods, etc.

BX: Bergson-Samuelson social welfare function was there introduced in Graaff?

PKP: Yes, it was also in our course. The Bergson-Samuelson social welfare function was there, but not Arrow's impossibility theorem.

BX: So who helped you to choose the problems for your dissertation?

PKP: Amartya Sen. He was a wonderful supervisor. Many people encouraged me to go from Delhi to some university in the U.K. or the U.S.A. to do my Ph.D. That was the standard practice among Indian students at that time. Somehow I decided to stay on in the Delhi School of Economics. I think that was one of the best decisions that I have taken in my life. I doubt that I could have got as good an academic supervisor anywhere else, especially since I did not have much idea about British and American universities and was in no position to make my choice of a university wisely.

BX: So, did he guide you to read the literature before you started working in actual problem?

PKP: He first asked me to read K. J. Arrow's *Social Choice and Individual Values* (Wiley, New York, 2nd edition, 1963). So I read the book, and, for nearly three months, I read it again and again.

BX: And Blau, who made the corrections?

PKP: Yes, I also read J. Blau's "The existence of social welfare functions" (*Econometrica*, 25, 1957).

BX: What was your first research problem?

PKP: My first paper was a small note, "A note on Leibenstein's 'Notes on welfare economics and the theory of democracy"' (*Economic Journal*, 77, 1967), and my second paper was "Risk, impersonality and the social welfare function" (*Journal of Political Economy*, 76, 1968). As I have mentioned, I was reading Arrow's book repeatedly for the first three months or so, but it was difficult to read only Arrow's book for three months. So, during that time, I also read a few papers, such as J. Harsanyi's two papers, "Cardinal utility in welfare economics and the theory of risk-taking" (*Journal of Political Economy*, 61, 1953) and "Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility" (*Journal of Political Economy*, 63, 1955), and some books, such as K. Popper's *The Logic of Scientific Discovery* ((Hutchinson, London, 1959) and *The Open Society and Its Enemies*, Volumes 1 and 2 (Routledge & Kegan Paul, London, 1945).

BX: So you had quite a bit philosophy while doing your dissertation. That must have helped you to understand Harsanyi's paper.

PKP: Yes. In particular, R. M. Hare's notion of universalizability of moral judgments (see his *Language of Morals*, Oxford University Press, Oxford, 1964) helped me quite a bit to understand Harsanyi's concept of "impartiality" and Rawls' notion of an "initial situation", where nobody knows anything about his/her specific position in the society.

BX: Rawls did not have a book at that time so presumably you read his papers.

PKP: Yes, I read his papers, "Outlines for a decision procedure for ethics" (*Philosophical Review*, 60, 1951) and "Justice as fairness" (*Philosophical Review*, 67, 1958).

BX: You spent three months reading Arrow. Anything coming out from reading the book?

PKP: My reading of Arrow led me subsequently to work on the restricted preference approach. But that was a little later. I took quite some time to understand and appreciate the intuition underlying Arrow's problem. The formal part was difficult, but was not as difficult as understanding the intuition. Why is the problem of interest? Where does it exactly fit in welfare economics? Questions such as these perplexed me quite a bit. One thing that puzzled me much was I. M. D. Little's paper ("Social choice and individual values," *Journal of Political Economy*, 60, 1952) on Arrow's theorem. Little claimed that Arrow's impossibility theorem did not have much to do with welfare economics. As a student starting to work on welfare economics and trying to read Arrow's book as thoroughly as I could, I was completely baffled when I came across this claim of Little, a highly respected welfare economist. Only later, after reading Amartya Sen's "Social choice theory: A re-examination", *Econometrica*, 45, 1977), I saw clearly what Little's own conception of welfare economics was. About 40 years later after I completed my Ph.D., I also wrote a paper ("Little and Bergson on Arrow's concept of social welfare", *Social Choice and Welfare*", 25, 2005) on the intuition underlying Little's conception of welfare economics.

BX: At that time, you had Bergson, Little, and Arrow. Did it strike you that they were talking about very different things?

PKP: At that time, the received wisdom was that, underlying every Bergson-Samuelson social welfare function, there was an Arrow-type social welfare function. When Bergson and Little were claiming that Arrow's aggregation was not welfare economics but something else, the standard counterargument to that was that, underlying every Bergson-Samuelson social welfare function, there must be an Arrow aggregation procedure. Only later, I could see the intuitive distinction between the two problems. Actually, I think Little's 1952 paper was a profound paper. Arrow's impossibility theorem was so new at that time. Even at that very early stage, Little (1952) saw exactly what Arrow was doing and the distinction between Arrow's problem of aggregation of opinions or preferences and Little's own conception of welfare economics as dealing with the social welfare judgments of individuals and the ethical bases of such judgments. The distinction was lost sight of until Sen stated it much more sharply in his *Econometrica* (1977) paper.

BX: So, on that aspect, how do you see Arrow's contribution fitting into welfare economics at that point?

PKP: As I have indicated, it was quite some time after I completed my Ph.D. that I saw clearly the distinction between the conception of social welfare of Little (1952) and A. Bergson ("On the concept of social welfare", *Quarterly Journal of Economics*, 68, 1954) and that of Arrow (1951). The distinction, as I see it, is this. Little (1952) and Bergson (1954) are talking about the values underlying the welfare judgments of individuals while Arrow is concerned with how social decisions are to be taken, given the preferences/opinions/judgments of the individuals in the society. As I see it, both the issues are important and have their legitimate places in welfare economics. It is not clear to me why Little thought that Arrow's concern was not a part of welfare economics.

BX: In the early years, people seemed to say that they understood each of those independent conditions (considered in Arrow's book) and combining of those conditions have some implications that seemed a big jump, not immediate. Did that present any problem to you when you were working on your dissertation?

PKP: The impossibility result, by itself, did not cause me too much of a problem. While the formal proof of the result was difficult, once one mastered it, the difficulty was over. But where I had much more difficulty was in figuring out how the different aspects of welfare economics and social choice theory that I had studied fitted together.

BX: you took a very short time (a year or so) to complete your dissertation. So what did you do after that?

PKP: I continued to work in the same area. Also, I had a fascination for the moral language, and I did some more reading on that. Once I knew that the formal requirement for a Ph.D. was met, I could afford to read other related material.

BX: So you were not worried about getting a job?

PKP: At that time, the job pressure was not that great, at least in the Indian system. If you did a Ph.D. from the Delhi School of Economics, then you could be reasonably sure about getting some teaching and research position in an Indian university.

BX: So there was no pressure.

BX: While you were waiting for your Ph.D., you were teaching in the Delhi School of Economics, teaching courses for graduate students?

PKP: I was not teaching many classes formally; I was mainly tutoring and giving a few lectures.

BX: Since you were not teaching formally, you were free to "teach" anything. Did you introduce anything you just learned?

PKP: In the Indian system, the curriculum was rather rigid. One had some freedom, but not to introduce a completely different subject. A well-known professor might be able to do it, but not a Ph.D. student.

BX: How were the students?

PKP: My students in the Delhi School of Economics were brilliant. Some of the very best students of India came to the Delhi School of Economics. You could throw at them just about any challenging idea and they could take it.

BX: After your degree, you were working in the similar areas. Then there was a big change, you moved to strategic or sincere voting. How did you get interested in those issues?

PKP: You know Y. Murakami's book, (*Logic and Social Choice*, Dover Publications, London, 1968). It had a few pages (pages 74–80) on strategic voting. It was published in 1968, to be followed a year later by R. Farquharson's *Theory of Voting* (Yale University Press, New Haven, 1969), where strategic voting was discussed in a very elegant fashion. Prior to Murakami (1968), Arrow's book (*Social Choice and Individual Values,* 1951, 1963; p. 7) had, indeed, mentioned the possibility of strategic misrepresentation of preferences, but Arrow did not pursue the theme. In Murakami (1968), however, there was a more extended discussion of the problem of strategic voting; Murakami called it the problem of stability of outcomes of "sincere" individual decisions. Shortly after I received my Ph.D., I read Murakami's book, and, subsequently, I also read Farquharson's book. I recall that I wrote to Amartya Sen asking him whether he thought that strategic voting might be a good problem to work on. I don't exactly remember his answer but, as far as I recollect, he wrote a very encouraging reply, saying that, if I felt interested, I should go ahead and work on the problem. He was absolutely wonderful. He spent time and effort to read through everything that I wrote at that stage and gave detailed comments. After I received his reply to my letter, I started to work seriously on the problem of strategic voting.

BX: By that time, you were in Harvard, right?

PKP: Yes.

BX: Did you get a chance to talk to anyone on those topics?

PKP: I had a few conversations with Kenneth Arrow. But, as someone who had just completed his Ph.D., I was too diffident to approach Kenneth Arrow and have a conversation. I came from the Indian system, and I was too awed to do it. Later I felt that I could have got a lot more benefit from my time at Harvard if I had more courage. But approaching Arrow and asking him whether he would be able to spare time to talk with me about some problem was too daunting a task for me.

BX: Allan Gibbard was a student at Harvard at that time.

PKP: That is right.

BX: Did you get to meet or know him?

PKP: No. I listened to his first presentation of his famous paper on quasi-transitivity and oligarchy. It was presented in a seminar course. I was there. But I did not have any interaction with him.

BX: So you knew of him, but never met him there.

PKP: No, I did not meet him at that time. I met him a few years later.

BX: After Sen's positive result, non-dictatorship is consistent with quasi-transitivity, he (Gibbard) came back to say that there is a link to oligarchy. Right?

PKP: I don't know the exact sequence. My impression is that he was already thinking of the problem. It was not necessarily because Sen had shown that non-dictatorship was consistent with quasi transitivity.

BX: Was that true Sen was visiting Harvard around the same time? Perhaps Sen talked about the positive result then?

PKP: That is right. But Allan Gibbard probably thought of the main ideas of his oligarchy result independently. He might have read Sen's paper; I do not really know.

BX: For that paper, he wrote for the class jointly taught by Arrow and Sen. It was just published in *Economics and Philosophy* (Social Choice and the Arrow's Conditions, *Economics and Philosophy* 30(3), 269–284, 2014).

PKP: That is the seminar course I mentioned earlier - the joint course taught by Arrow, Sen, and Rawls.

BX: It was cool.

BX: When you started working on sincere voting, later became known as manipulation, were you already thinking in terms of counter threat or just threat?

PKP: I was thinking in terms of Murakami's (1968) notion of stability. Later I found exact the same concept of strategic voting in Farquharson's (1969) book. Farquharson's discussion was in terms of the notion of "equilibrium", but the basic idea was the same as Murakami's.

BX: And Mark Satterthwaite was working in the area.

PKP: Yes. I believe that the contributions of Murakami (1968) and Farquharson (1969) preceded those of Gibbard and Satterthwaite. As you know, the publication of Farquharson's (1969) book was held up for a long time for some reason. I do not know why it was held up. As I mentioned earlier, I started working on strategic voting only after I read the books of Murakami (1968) and Farquharson (1969).

BX: These came as natural reactions to the Arrow type question, or were they looking more on the positive side? Some argue voting is more on the positive side than Arrow's impossibility theorem.

PKP: Let me put it this way. Basically, Arrow's theorem told us that, even if we knew people's true preferences or opinions, we would still have a difficulty in aggregating them to reach a social ranking of the various options for the society. Arrow was not concerned with the difficulties that we might have in getting hold of the true preferences or opinions of people. The literature on strategic voting tells us that asking people to give us their true preferences or opinions may not work sometimes because people may find it beneficial for them to mis-reveal their preferences or opinions. It seems to me that both these aspects are relevant for the problem of reaching social decisions on the basis of the preferences of the individuals in the society.

BX: You came to know Gibbard's and Satterthwaite's papers later.

PKP: Yes. My first paper on strategic voting ("On the stability of Sincere voting situations", Journal *of Economic Theory*, 6, 1973) was published in 1973 (it had been received by the journal in August 1972). There I mentioned that, after the completion of my paper, I had come to know about a yet-unpublished paper by Gibbard on a similar theme. When writing my paper, I was not aware of Gibbard's paper. Gibbard also was not aware of my paper when he wrote his paper. We were asking the same question but in different frameworks.

BX: Your paper came out in 1973, Gibbard's paper came out in 1973.

PKP: Yes. I presume that he also submitted his paper for publication in 1972 or earlier.

BX: Mark Satterthwaite's paper came out in 1975.

BX: The next major idea you worked on, of course you worked in a big way, was relaxing rationality condition in Sen's Paretian liberal issues. When did you learn of his result? His paper and book both came out in 1970. Presumably he was working on it while in Delhi school. What was your reaction?

PKP: I think I came to know about his result in 1968. At first, I thought that the main point of the paper was that Sen was relaxing Arrow's condition of rationality to acyclicity of the social strict preference relation and was still deriving an impossibility result. Subsequently, I realized that there was another aspect of what Sen was doing which was perhaps far more important: Sen was providing the first formal formulation, in welfare economics, of the notion of rights of individuals. I remember a conversation between Stephen Marglin and Amartya Sen, where I was present. As far as I recall, it was in late 1968. Marglin was arguing that what Sen was doing in his paper on the impossibility of a Paretian liberal ("The impossibility of a Paretian liberal", *Journal of Political Economy*", 78, 1970) was intuitively similar to what he (Marglin), Sen, and others had done in the literature on the social rate of discount. The similarity that Marglin was pointing out was that, in both contexts, the freedoms or rights[1] of individuals were coming into conflict with the requirement of Pareto optimality of the social outcome, given the presence of externalities. In the literature on the social rate of discount, the individual freedom under consideration was the individual's freedom to decide how much she would save while, in Sen's problem of the Paretian liberal, the freedom or right under consideration was the individual's freedom or right with respect to his own private life. Many years later, looking back on that conversation, I realized that Marglin's arguments contained deep insights.

BX: It seems Stephen Marglin was the first person to articulate the rights structure different from Sen.

PKP: It was only when Wulf Gaertner, Kotaro Suzumura, and I worked on the issue of an appropriate formal formulation of individual rights that I remembered the conversation between Marglin and Sen to which I have alluded above and realized that most probably Marglin had in mind a conception of individual rights similar to what Wulf, Kotaro, and I were advocating.

BX: So you had quite different impression of Sen's result initially.

PKP: Initially I agreed with Sen's formal formulation of individual rights without any reservation whatsoever.

[1] I use the terms "freedom" and "right" interchangeably here though there are important differences between them.

BX: So what made you to change your position and to be different from Sen's later? Although Sen was talking about liberalism, he didn't emphasize rights that much in that particular article.

PKP: I believe that that is correct. Initially, Sen didn't use the term "right"; instead he used the terms "liberal" and "liberalism".

BX: But it was Gibbard who coined that word rights more directly.

BX: What do you think about the state of the literature? Could you talk about how your thinking on this problem had evolved?

PKP: Sen's result on the impossibility of a Paretian liberal had several striking features. First, it proved an impossibility result using a collective rationality condition (namely, acyclicity of the social strict preference relation), which was much weaker than the rationality condition postulated by Arrow. Second, for the first time in modern literature on social choice and welfare, it sought to provide a formal formulation of some essential features of J. S. Mill's (On *Liberty*, 1859, reprinted by Liberal Arts Press, New York, 1956) concept of individual rights to liberty in private matters. Finally, it posed in a dramatic fashion the tension between individual rights and the requirement of Pareto optimality for the final social outcome. Much of the later debate about Sen's result concentrated on the last two features I have just mentioned. Some writers raised the issue of whether Sen's condition of minimal liberalism was consistent with our intuition about individual rights, and several writers went further and claimed that, if individual rights were viewed in an "appropriate" fashion, then the conflict between individual rights and Paretian values would disappear. Of course, from a purely formal point of view, one can delink Sen's condition of minimal liberalism from its interpretation in terms of individual rights and consider it as a constraint on the aggregation procedure, which has nothing to do with rights and liberties. But that was not how Sen intended the condition to be viewed, and, in my opinion, that would also rob the condition of much of its interest and significance.

To start with, I did not have any problem with the condition of minimal liberalism, but later on I developed concerns about whether the condition was consistent with our intuition about individual rights (I hasten to add that I do not think that any reasonable reformulation of individual rights will provide us an escape route from the basic tension between individual rights and Paretian values to which Sen has drawn our attention). In an earlier question, you wanted to know why I changed my position regarding Sen's condition of minimal liberalism. Because of an oversight, I did not respond to your question at that stage. Let me respond to it now.

The way I changed my position regarding Sen's condition of minimal liberalism was a little peculiar. Bob Sugden had criticized Sen's formulation of individual rights and liberties in terms of his condition of minimal liberalism (see, for example, R. Sugden, "Liberty, preference, and choice", *Economics and Philosophy*, 1, 1985). As far as I can recall, sometime in the late eighties, Wulf Gaertner, Kotaro Suzumura, and I felt that Bob's criticisms were too strong. Given that Wulf, Kotaro, and I were all sympathetic towards Sen's formulation, we thought of writing a paper to respond to the criticisms of Bob. But, in the course of working on this paper (W. Gaertner, P. K. Pattanaik, and K. Suzumura, "Individual rights revisited", *Economica*, 59, 1992), we basically came round to Bob's point of view. So far as I am concerned, Bob

Sugden's writings did influence me strongly, but my position shifted decisively only after I had the opportunity of interacting with Stig Kanger, who was Professor of Theoretical Philosophy at Uppsala University at that time. Stig Kanger, some of his colleagues at Uppsala university, and a few people, including me, from the United Kingdom met several times in Uppsala University, University of Birmingham (U.K.), and London to discuss various problems relating to social choice and rational choice (as far as I can recall, Bob Sugden participated in some of those meetings). Those meetings profoundly influenced my ideas about rights and liberties. I still recall the excitement with which, following one of those meetings, I read S. Kanger and H. Kanger's paper "Rights and parliamentarism" (*Theoria*, 32, 1966). To me, that paper is a splendid example of how much intuitive insight one can gain through a rigorous analysis of the linguistic structure of words, such as "rights". Reading that paper convinced me that there were important incompatibilities between Sen's formulation of the notion of individual rights and liberties and the way in which we use the term "rights" in our daily language. Wulf Gaertner, Kotaro Suzumura, and I have discussed in detail the nature of these incompatibilities in our *Economica* (1992) paper; elsewhere, we have also separately elaborated the arguments developed in that paper. So there is no need for me to repeat those arguments here. Apart from minor modifications and refinements, my position about the appropriate formal formulation of individual rights continues to be the same as the position taken in the Gaertner-Pattanaik-Suzumura paper (1992).

BX: What do you think of the present state of the literature?

PKP: I am not sure that that much work is being done on individual rights presently in social choice theory.

BX: Let's go back to Sen's result. Suppose we don't call his notion as a notion for rights and stick to his minimal liberalism condition. How do you locate his contribution to the social choice literature?

PKP: Suppose one completely delinks the condition of minimal liberalism from Sen's original interpretation of it in terms of individual rights. Then one cannot criticize it on the ground that it conflicts with some essential features of individual rights which it was seeking to capture. But the examples that critics have used to argue that there are tensions between the condition of minimal liberalism and one's intuition about individual rights can still be used to argue that the condition, without its interpretation in terms of individual rights, would conflict with one's respect for an individual's rights to liberty in private matters. While delinking the condition of minimal liberalism from its original interpretation in terms of rights is a formal possibility, it does not, however, seem to be an attractive idea, given that the condition constitutes a historic attempt to break out of the traditional welfaristic structure of welfare economics. Let me clarify two points here. First, I believe that it is completely beyond dispute that Sen's (1970) paper is a path-breaking contribution, which, for the first time, brought into modern welfare economics and social choice theory "non-welfaristic" considerations, such as individual rights and liberties. As far as I know, no one in social choice theory has ever questioned the tremendous conceptual impact of Sen's contributions on individual rights. Second, I believe that Sen's basic insight about the tension between individual rights and Paretian values remains completely intact even if one switches from his formulation of individual rights to the "game form formulation" advocated, among others, by Sugden, Gaertner, Suzumura, and me.

BX: It seems this particular right issue that economists speak about gives a bigger context in economic analysis.

PKP: Yes, it does. The broadening of the scope of welfare economics to include the so called "non-welfaristic" considerations has its origin in the contributions of two economists. The first economist was John Hicks. In the preface to his book *Essays on World Economics* (Clarendon Press, Oxford, 1959) he expressed strong reservations about what he called "economic welfarism". That was back in 1959. As far as I know, no one picked up that theme until Sen's (1970) paper introduced individual rights and liberties into welfare economics and opened up the path for the exploration of non-welfaristic values in general.

BX: Another important area that you worked on is choice under complete uncertainty. How did you get into that area?

PKP: My interest in this area came from a paper by Y. Kannai and B. Peleg ("A note on the extension of an order on a set to the power set", *Journal of Economic* Theory, 32, 1984). The interpretation that many people, including me, put on the Kannai–Peleg framework ran in terms of decision-making under uncertainty.

BX: That was to compare different sets.

PKP: Yes, it involved the comparison of different sets of outcomes.

BX: That took you to the area of complete uncertainty.

PKP: Yes. Among the questions that both of you sent me earlier for the interview, there was a question about when I started working on the link between the problem of comparing sets of outcomes to the problem of measuring freedom. I don't want to embarrass you, Yongsheng, but the idea of linking the two problems came to me from you. You visited Birmingham in 1989. If I recall correctly, you mentioned to me that Nick Baigent had suggested to you that, when visiting Britain, you might look me up in Birmingham. In any case, I am deeply grateful that you came to see me in Birmingham. Without our discussions in Birmingham, probably I would never have been seriously interested in the problem of measuring freedom.

YX: That's very kind of you. But there were some crucial axioms needed proper motivation and I was not quite sure at that stage.

TB: You came to give your "job" seminar at UC-Riverside and presented the paper on freedom of choice. My reaction was "Wow, I wouldn't have thought of that way."

BX: Sugden wrote you that he and Jones had a paper dealing with freedom as well. After reading their paper, what was your reaction to their approach?

PKP: When Bob was visiting me here in Riverside, I told him about the paper ("On ranking of opportunity sets in terms of freedom of choice", *Recherches Economiques de Louvain*, 56, 1990) by Yongsheng and me. He then told me that, in a paper ("Revaluating choice", *International Review of Law and Economics*, 2, 1982) by him and P. Jones, there are axioms similar to certain axioms in the paper by Yongsheng and me. After reading their paper, I felt very guilty that I was unaware of their paper published many years before the paper by Yongsheng and me. Subsequently, when Salvador Barbera, Walter Bossert, and I wrote a survey of the literature on the ranking of sets, I took care to make completely clear the priority of the paper by Jones and Sugden. I admire the paper of Jones and Sugden. They go straight to the intuitive core of the problem and deal with it very elegantly and without much formalism. I admire that quality.

BX: Some people have been saying that this (ranking sets of outcomes) is not the right way of modeling freedom. In particular, freedom is more complex than this simple approach. What is your view on this, especially on interdependence of individual actions?

PKP: As you know, Kaushik Basu ("Achievements, capabilities and the concept of well-being, A review of Commodities and Capabilities by Amartya Sen", *Social Choice and Welfare*, 4, 1987) was the first to draw attention to the problem that arises when, in a world of interdependent individuals, we seek to conceive an individual's freedom as the opportunity to choose an outcome from a set of outcomes available to the individual concerned (it may be recalled that, in the literature on functioning and capability, outcomes are functioning bundles). Subsequently, Yongsheng and I have raised the same problem on different occasions. In the world as we know it, often an individual does not choose an outcome. In many situations, the agent chooses an action and the outcome for her is determined jointly by her action and the actions chosen by other agents. In such situations, the conception of an individual's freedom in terms of a set of available outcomes from which she can pick up whatever outcome she chooses to have turns out to be problematic. It seems to me that this is an important issue that deserves much further attention than it has received so far. It is not that I am denying the importance the notion of opportunity or freedom of choice for the well-being of an individual. All that I am saying is that the notion, currently prevalent in the literature on functioning and capability, of an individual's freedom as the ability to choose any functioning bundle from a set of available functioning bundles may be of limited usefulness in a multitude of situations where the interdependence of individuals and strategic interactions constitute an essential feature.

BX: For people working in the area of functioning and capability, perhaps the major reason that they take the earlier approach is its operational simplicity. Once you view the notion of freedom in a more complex way, it would create lots of problems in measuring or quantifying freedom. There seems to be a tradeoff between the simple way of approaching the problem and the "right" way.

PKP: I agree with that. We know that the problem of measuring freedom is very difficult even in a relatively simple framework where an individual's freedom of choice is taken to be reflected in a set of outcomes from which the individual can freely choose any outcome. If we assume that the individual can choose her action but the outcome for her is determined by her own action together with the actions of several other individuals, then the problem becomes much more complex. Yongsheng, you and I have struggles with this more difficult problem for several years without any conspicuous progress to show for our labor. Here, as in many other areas of economics, one may be compelled to work in a tractable but limited framework because the problem becomes too intractable when we try to go beyond that limited framework. At the same time, however, it may be important to acknowledge and keep in mind the many limitations of the conceptual framework which, despite its limitations, we continue to use because of its tractability.

BX: One of the arguments for the functioning and capability approach to human well-being is that it brings welfare economics closer to real life. What's your opinion?

PKP: There is a valid point in that argument. After all, attributes such as nourishment, education, health, and housing figure all the time in our economic and political debates, and they also constitute some of the most important functionings or dimensions of well-being that figure in the functioning and capability approach to well-being. Even most people who would identify an individual's well-being with her happiness would probably readily agree that being well-nourished, being healthy, being protected from the elements, etc., are conducive to sustained happiness, and that, if one lacks a significant number of these attributes to a significant extent, then one cannot be happy. Also, even if one admits the logical possibility that a person, who is malnourished, unhealthy, homeless, and so on, can be happy, one can still plausibly argue that empirically such instances are extremely rare and can be ignored. Thus, the central concerns of the functioning and capability approach overlap with the central concerns in many other areas. Viewed from this perspective, the difference between the conception of well-being based on sustained happiness, as distinct from transient pleasures, and the conception of well-being in the functioning and capability approach may not be as great as what it is sometimes thought to be. There is a rapidly expanding body of scholarly research on happiness and the determinants of happiness. I suspect that, despite the differences between their terminologies and analytical tools, those working on happiness and those working on the functioning and capability approach may have much to contribute to each other.

BX: You have been working on multi-dimensional well-being, deprivation. Is the idea of research coming out because of your work in welfare economics or because of your deeper interest in development economics where you've seen poverty, deprivation, hunger?

PKP: I think it came mainly from my interest in welfare economics. I feel deeply concerned about poverty, deprivation, hunger, etc. in India as well as in other parts of the world. I am also very much interested in the different conceptions of well-being and deprivation, including the ones advocated by the functioning and capability approach. But it is not clear to me that the distinctions between these alternative conceptions are all that important for tackling the immediate and urgent task of reducing poverty and deprivation in India and many other countries.

BX: So, when you work on the Odisha[2] project, are you interested just in the question like what it is, what the state is, or in knowing what it is and what you can do about it?

PKP: Both. We need to know much more about the economy of the state to say what can be or should be done. Incidentally, may I please clarify something here? I presume that by the "Odisha project" you are referring to the forthcoming book *The*

[2]Odisha is a state of India, located in the eastern coast, with a population of about 44 million.

Economy of Odisha: A Profile, to be published by Oxford University Press, Delhi.[3] I am afraid that I cannot claim the entire credit for this volume. The volume is edited by Pulin B. Nayak, Santosh C. Panda, and me, and it contains contributions on the economy of Odisha by 26 authors including the editors.

BX: What is your view of the present state of development economics in general and in Odisha in particular?

PKP: I am not really well qualified to comment on the state of development economics. First, though I have strong interest in development economics, my reading of the huge literature on economic development has not been as extensive as it should be. Second, I have done very little research on issues relating to economic development. I do, however, find recent developments in the area quite exciting. In particular, the experimental approach to development economics is fascinating; it has added an entirely new dimension to the subject. Of course, analytical modeling and traditional econometric analysis of observational data continue to be very important for development economics. One cannot claim a priori superiority for any of these approaches over the others. We need all of them. In a rich and complex area such as development economics, we need every bit of help that we can possibly get from all the tools that are available to us.

BX: We need the entire menu.

PKP: Yes, we need the entire menu.

BX: Returning to an earlier point. You think the literature on sustained happiness poses an important challenge to the functioning-capability approach. Is this right? How should they respond to each other, and is there anything that they can learn from each other?

PKP: I am not sure that the research on happiness and its determinants done by psychologists and economists poses a challenge to the functioning and capability approach. But I believe that all of us working in the framework of the functioning and capability approach can learn much from the recent research on happiness. Consider the following scenario. Suppose the psychologists give us a reasonable way of measuring happiness. Further, suppose they find that the functionings that we usually talk about in the functioning and capability approach promote happiness, that a person with fairly low achievements in terms of many these functionings is highly unlikely to be happy, and that happiness also depends on the freedom to choose one's life style. I am not suggesting that research on happiness has already corroborated all these suppositions. But, if indeed these suppositions turn out to be true, then, at least for the purpose of public policy, it may not matter all that much whether our conception of an individual's well-being runs in terms of happiness or in terms of functionings and capability.

BX: What's your view of the role of a theorist in general and a theoretical welfare economist and social choice theorist in particular in the economics profession? What's your view on a theorist taking on applied and empirical work?

[3]The book, which was unpublished at the time of the interview, has been subsequently published in 2016 by Oxford University Press, Delhi.

PKP: The role of theorists in general and theoretical economists and social choice theorists in particular is a rather big issue and I really do not have much to say about it. It seems to me that the choice between working on theory and working on empirical problems is ultimately a matter of one's personal taste. I have mainly done theoretical research and I have done empirical research only occasionally. But that is entirely because of my aptitude for theory and not because of any judgment on my part about the relative value or importance of the two types of research.

BX: You are one of the founders of the Society for Social Choice and Welfare, and have been involved with the journal, Social Choice and Welfare, from its inception. Could you tell us how you feel about the evolution of the Society, the Journal, and the social choice field in the past 30 + years? Do you see any challenge for the Society, the Journal, the field, and the people working in this area in the future? Would you like to offer your advice to our young colleagues?

PKP: I believe that the journal, as well as the Society, is doing very well. A fairly large number of young scholars have chosen to work in the field of welfare economics and social choice theory over the last 30 years or so. That is a good indicator of the robust health of the subject, and I am very optimistic about the future of the subject. I am afraid that I do not have any advice for young colleagues. In fact, it seems to me that young scholars rarely need general advice (as distinct from suggestions about specific research problems) from their senior colleagues!

BX: You "officially" retired a few years ago, but continue to be very active in research. What has kept you busy? Have you got any time to further your other interests since your retirement, say for example, Sanskrit, poems, etc.?

PKP: I have continued my research after retiring in 2007 partly because of a habit formed over several decades. But it is also partly because I quite enjoy doing research, especially when there is no pressure on me to do any research. My university (University of California, Riverside) has helped me much by providing me, even after my retirement, with all the support that I need for my research.

Retirement, however, has given me the opportunity to pursue my other interests. I have been always keenly interested in literature, especially poetry. Now that I do very little teaching, I can read literature much more extensively than I could when I had to devote most of my time to teaching.

BX: For those who are graduating and are looking for academic jobs, they might want to know your experience. How did you land your first job? Was the job market then vastly different from current? How did you manage to be productive after getting the very first job?

PKP: I had my entire education in India, and, leaving aside a very short stint of teaching in a college after my M.A., I seriously searched for an academic position only after I completed my Ph.D. in India in 1968. The circumstances were so very different then that my initial job search experience can hardly be of much interest to young scholars graduating now. I would, however, like to say that I was singularly lucky to have the strong support of Amartya Sen, who was the supervisor of my Ph.D. dissertation, Stephen Marglin, and K. J. Arrow and Frank Hahn, both of whom were examiners of my Ph.D. dissertation. I am deeply grateful to all of them for their help and support at various stages of my academic career.

John E. Roemer

Roberto Veneziani and Marc Fleurbaey

1 On Being an Economist

1. *Economists (both theoretical and applied) often like to portray themselves as neutral analysts, who possess a kit of scientific tools that can be applied to a number of issues and are value-neutral. In the definition of "social scientist", they tend to emphasise the noun, and set the adjective aside. Your work seems to be inspired instead by a stronger, broad social motivation. A similar emphasis on the social role of the economist can be identified in the founders of social choice theory. Is there a social responsibility of the economist? Do you view the economist as an "intellectuel engagé"?*

Clearly this is a loaded question; the answer is yes. I do believe economists are scientists, or should be. Some (Deidre McCloskey) argue that economics is merely rhetoric, a language with which one argues for one's pre-conceived ideas. Indeed, there is nothing wrong with that: a common language with clear rules enables communication and argument. Much economics, however, is true discovery: we use it to understand how an aspect of the world works. It's a mistake to think that economics is either pure science or pure rhetoric, as it is some of both. Honesty requires that the economist warn his/her audience what assumptions or judgments are built into the premises of an argument, and may therefore escape notice. The first theorem of welfare economics is an obvious example: it can be used to argue that a market

R. Veneziani (✉)
Queen Mary University of London, London, USA
e-mail: r.veneziani@qmul.ac.uk

M. Fleurbaey
CNRS, Paris, France
e-mail: marc.fleurbaey@gmail.com

Paris School of Economics, Paris, France

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_14

system with private ownership of assets arranges things nicely, or, by pointing out how demanding the premises of the theorem are, it can be used to generate scepticism about the harmonious outcome of such a system.

To the extent that any scientist is searching for truth, he or she has a social responsibility. Since the questions that much of economics addresses concern issues that affect the distribution of income, the social consequence of economic research may be more partisan than, say, the research of astrophysicists. I believe that much of economics takes sides, and the economist should be forthright about this.

2. *Although some of your contributions focus on empirical issues, and much of your work has a clear relevance in policy terms, your overall intellectual efforts are probably best characterised as "theoretical" rather than "empirical". What, in your view, and in the light of current trends in the discipline, should be the relation between theoretical efforts and empirical analysis?*

I believe economic theory has become too detached from reality. Most, or all, of the great economic theorists of the past had thorough empirical knowledge, but many, even most, of the top theorists today do not. The ambience among Ph.D. students in economics is more like that in an applied mathematics department than a history department, and I think that is a shame. Much of contemporary game theory has gone off the deep end: it has become a scholastic exercise, with very little applicability to real social situations. I do not have the answer for how to control this. Mathematics gives us a wonderful way of thinking, and it is seductive, especially for young people, who desire simplicity and elegance.

I think behavioural economics is a refreshing development. To be somewhat more precise, the premise of hyper-rationality has rendered much game theory scholastic and irrelevant; what behavioural economics does is relax that premise, but it should not relax the modelling standards. It's the elegance of the hyper-rationality model that is seductive. With some work, we will develop elegant theories without hyper-rationality. We can depend on the deep proposition that good science is always beautiful (which has something to do with the simplicity of truth). When aspects of a model become too complex, such as some strategies in game-theoretic analysis, one must question their relevance. Just as it is difficult to diagnose a bubble in asset prices, it is difficult for those working in game theory to see when the models are becoming too complex to be good descriptions of human behaviour.

I think that the problem with economic theory, of excessive abstraction, will be resolved over time. To the extent that economic theory fails to be applied, it will die. And if it's based on fantastic premises, it will not be applied. In macroeconomics, I think this will happen to the rational expectations model. The rational expectations model is elegant, but its failure to understand or even *admit the existence of* the 2008 crisis is its death knell. It may take time, but macroeconomics will be rebuilt.

My own path was to focus on economic theory and political philosophy. I was in part seduced by the elegance of abstract reasoning, but it was also in part due to where my talents lay. I have always tried to read history, but I have never been good at it. I don't seem to have the synthetic ability to put together many different historical observations and see the common thread, the general story, as good historians can do. Based on my own experience, I think we will always need a division of labor between theoretical and applied economists. Theory, however, has to be held to a high standard of relevance. Excessive formalization is a distraction, as is the trivialization of economics known as 'freakonomics' (Steve Levitt).

3. *One important and constant feature of your theoretical work is the constant dialogue with other disciplines ranging from political philosophy to political science to, more recently, climate science. How important is interdisciplinarity in your research? And more generally, in an era of increasing specialisation in research, how important is it for economists (and especially welfare economists) to be able to cross disciplinary boundaries?*

I believe interdisciplinarity has to come naturally. In my own case, I was led to political philosophy in my attempt to understand why Marx's concept of exploitation diagnosed an instance of injustice. It was also serendipitous: I met the political philosopher G. A. Cohen at the time I was writing my book on class and exploitation (published in 1982), and he was also trying to put classical Marxist questions into a modern form. Without Cohen's influence, I would probably not have engaged with political philosophy.

It is clear that global warming is one of the greatest challenges to human society of this century, perhaps the greatest challenge. It involves intergenerational ethics, politics, inter-regional ethics, and economics. I have not become a climate scientist, as your question tends to imply, but it seems to me natural to be drawn to this huge question, and to attempt to help solve it.

4. *How important has your affiliation to the so-called "September group" been both in terms of encouraging an interdisciplinary approach to economics, and more generally in your intellectual development?*

The September Group began when a few young social scientists, historians, and philosophers, on the left and predominantly Marxists at the time, found they were engaged in a common project: to re-phrase Marxist questions using modern language, and to address them with contemporary tools of social science and analytical philosophy. G. A. Cohen had published *Karl Marx's Theory of History: A Defence* and Jon Elster *Logic and Society*, in 1978; in 1980 I was working on the manuscript of *A General Theory of Exploitation and Class*, using general equilibrium and game theory to study classical issues of Marxism, and read these books. In other words, analytical Marxism was in the air. Others were doing analogous work. The common project lasted through the 1980s, during which time analytical Marxism became an important school. Not surprisingly, we were attacked by many more traditional (dogmatic?) Marxists, who said that our use of 'bourgeois' tools guaranteed that our theories would be bourgeois, not Marxist. By the end of the 1980s, I think that our task had largely been completed, in the sense that we had a pretty good idea what parts of Marxist theory should be preserved and what parts junked. (We had differing views on the fraction of the Marxist corpus that was worth preserving—one extreme view being Elster's that it should all be junked.) Some members left the group after the Eastern-European and Russian revolutions of 1989–1991, but the group stayed together, and we continue to meet, approximately annually, although it could not be said that the themes continue to be Marxist. One does not easily cast aside a group of intellectual comrades with whom one has discussed and debated for over thirty years.

Jerry Cohen was probably my greatest intellectual influence as an adult. (My socialist parents influenced me tremendously as an adolescent.) In particular, Cohen gave me confidence that doing rigorous economics focusing on left-wing topics was a worthwhile way to spend a lifetime. That confidence was built by observing his fine example in political philosophy. I was also strongly influenced by Jon Elster, who possesses a wonderful synthetic talent, which, as I said, I lack. I consider Elster to be the anchorman of the social sciences. His evolution to a point where he holds a deeply agnostic view about the possibility of constructing grand theory in the social sciences has been fascinating to watch. Elster has the highest standards of intellectual honesty, and that takes courage.

5. *Unlike in the past, economics is nowadays largely a journal-based discipline. In contrast to the widespread orientation towards journals, a significant proportion of your academic contributions have been published in books. Why have you chosen to publish books? And what are the differences between the two types of outlets?*

I have always tried to publish my work in journals prior to writing a book on a topic. Sometimes I have failed: my work on equality of opportunity was rejected by many economics journals, and so I published a short book on the topic in 1998. (The first version of that theory was stated in an article in *Philosophy & Public Affairs*, not an economics journal.) In a book, one has more freedom to present ideas informally. One also has the freedom to be more interdisciplinary than is permissible in the top economics journals—in my case, to discuss both economic and philosophical sides of a question. In addition, in a book, one can be more conjectural. If one attempts to offer a conjecture in a paper for a journal, a referee will ask for the evidence. And you may not have much to offer…yet.

2 On Economics

1. *What do you consider to be the main intellectual influences in your formation and then in the evolution of your thinking (both in economics and more generally)*?

I've said my main intellectual influence was Jerry Cohen, due to confluence of our political ideas, our penchants for abstract thinking, and his powerful mind. Although I was an intellectual socialist as an adolescent, I was further radicalized by the student movement around the Vietnam war. In economics, I came of age during the heyday of general equilibrium theory. I took Gérard Debreu's courses on linear programming and general equilibrium theory at Berkeley, in 1967, after transferring into the economics department from the mathematics department, as a graduate student. I was suspended from Berkeley in 1968, due to participation and arrest in a demonstration and an occupation, lost my draft deferment, and taught math in San Francisco secondary schools for five years, before being readmitted to Berkeley (unlike the more dangerous Mario Savio) to write a dissertation on a totally applied, political topic. (At that time, draft deferments were also given to teachers of math and science in inner city schools.) In the thesis, I did not use the economic theory that I had learned, most of which I had forgotten, due to atrophy, over the five year hiatus. My defender among the senior Berkeley faculty, without whom I would not have been awarded the Ph.D., was the iconoclast Benjamin Ward.

After being appointed an assistant professor at UC Davis in 1974, several mathematical economists there suggested that I teach a seminar on Michio Morishima's *Marx's Economics*, which had been recently published (1973). I spent a summer reading the book, and was taken with it: this was the first time I had seen 'mathematical Marxism,' something developed by Japanese economists during the 1960s. At this point, I started to relearn mathematical economics, and began on the research program that I described earlier. I had met Andreu Mas-Colell, then a tenured professor at Berkeley, and about my age, who was most solicitous of me, due to his own leftist past, and my own mathematical history. Through Andreu, I met Joaquim Silvestre, a fellow Catalan, who has been my tutor in economic theory since then. Joaquim put up with my very naïve questions about economic theory with great patience, probably because we were, so to speak, on the same side of the barricades, both politically and methodologically. I eventually graduated to being his co-author, leading to the most important collaborations of my career.

In terms of the evolution of my thinking, it's fair to say I hesitate at this point to call myself a Marxist. There are too many parts of Marx's economics that are clearly wrong, and hence to continue to maintain that label is misleading. I continue to think that historical materialism is a good way to look at the evolution of society. There are, of course, many caveats. There does seem to be a fairly strong tendency for the 'forces of production,' in Marx's words, to develop, and I think Marx was right to say that property relations would have to adjust to that development, rather than the other way around. (Capitalists, for example, may try to impede the development of those forces in order to maintain their particular profitable niches, but they will not succeed in doing so.) Unlike some on the left, I did not cheer when the Soviet Union dissolved, because I still feel that the experiment with socialism that took place during the twentieth century came out of an eternal human desire for creating a more communal way of living, based on a solidaristic ethos. I need not belabour here how the good intentions of many nineteenth and twentieth century socialists came eventually to nought, due to fatal design errors, political and economic, in the centrally planned economies and societies. Still, we do not know the extent to which social democracy, and the welfare states in all the advanced capitalist countries would have developed had it not been for foil of the Soviet Union. One of the great accomplishments of the twentieth century was the extensive socialization of income in the advanced countries, as represented, most simply, by the fraction of gross national product that is collected and allocated by the state.

I am, much more than I was forty years ago, a social democrat: I think the best societies in the world are the Nordic ones, and despite their small size, they provide a beacon for all to follow. One can argue that they are very special cases—small, and with very homogeneous populations during the period when their welfare states expanded. Nevertheless, they provide a counterexample to right-wing claims about human nature, which has important corollaries for economic design.

One of the most damaging of right-wing shibboleths is the permanence of narrow self-interest as the main motivator for human beings. This has been absorbed into neoclassical economics in the attention paid to the incentive problem. I don't regard the incentive problem as unimportant today: what I deny is that it is a fact of nature, rather than a consequence of behaviour that emerges in a class society. Even Joe Stiglitz, whom I greatly admire, has said that the incentive problem should be treated as having the same permanence as the production function (I don't recall his exact words, and may be doing him an injustice). I would never say that. I do not think it is crazy to envision a time when economic agents will not be primarily motivated by self-interest, and then the incentive problem as we know it will disappear. There will surely be new problems of economic design if and when that time comes; production functions will continue to be important constraints on economic feasibility, but the incentive problem will not. Even if people will have a desire to cooperate with each other, however, it is not trivial to know what action each should take—how can that decision be decentralized in the presence of asymmetric information? I am trying to think about this in my slow effort to formulate an economics of cooperation, as opposed to one of competition.

2. *Your initial theoretical efforts have been within the so-called Analytical Marxist school and Analytical Marxism has been your entry point into normative economics. How much of your earlier analytical Marxist theorisation has inspired your later work and which insights, if any, remain relevant?*

I consider the work I did on exploitation and class to be, at this point, mainly of interest to the historian of economic thought. It is, I think, worthwhile to know that much of Marxian economics can be derived using modern concepts of equilibrium and optimization—in particular, the relationship between (Marxian) exploitation and class, and between wealth and class (where class is defined not in terms of income, but in terms of relation to the labor process). The class-wealth correspondence is a durable idea. But I think that the Marxian notion of exploitation as the unequal exchange of labor is neither ethically useful, nor does it provide an essential insight about income distribution. Exploitation as a concept of injustice has been replaced by the egalitarian concepts created by late twentieth century political philosophers, principally John Rawls, Ronald Dworkin, and G. A. Cohen. In my view, exploitation's injustice is a special case of injustice due to unequal ownership of productive assets, and therefore the general question of when the distribution of productive assets is unjust becomes the thing to study. For this reason, Robert Nozick is a key figure, as he gave the most serious argument attempting to justify the unequal ownership of productive assets. In terms of exploitation as an explanation of the emergence of a surplus, and economic growth, I would say that the more fundamental explanation has to do with the conjunction of two institutions, one ethos, and one fact: private and unequal ownership of productive assets, the use of markets (in particular, a labor market), the profit-making ethos, and the scarcity of capital relative to the available supply of labor. Exploitation emerges from these things, and they are all necessary for its emergence.

I do think analytical Marxism was an important school, and infinitely superior to the post-modern Marxism that is now flourishing in some comparative literature and humanities departments, which turns science on its head.

3. *Setting aside issues about the efficiency of allocations, normative economics and social choice theory have traditionally focused on distributive issues, focusing on variables like income or wealth, and more recently (also thanks to your work) on basic goods, opportunities, functionings and capabilities. Significantly less attention has been paid, however, to other notions, such as power and force, that are potentially of relevant normative interest and that are central in political philosophy. Is this because these notions are too complex to be captured in standard mathematical models?*

Power and force are the subject of political science, while the four terms you mention above are words in the language of distributive justice. These are different topics. I suppose many of us on the left who live in advanced democracies focussed on distributive justice, rather than power and force, in the last third of the twentieth century, because it was somewhat of a mystery how such a clearly unjust economic distribution could emerge with universal suffrage and an extensive system of civil liberties. (I believe that Marx or Engels said that once universal suffrage came, socialism would not be far behind.) In other words, the main explanation of capitalist hegemony appeared not to be in the control of guns by capitalists, but rather, due to the massive influence of their ideology. It therefore became of primary importance to counter that ideology (a philosophical task), and to understand how, despite universal suffrage and civil liberties, capitalist power was so clearly maintained.

There are some who have talked a good deal about power and force: the calculations that Bowles and Gintis have made about the importance of ‘guard labor’ are a case in point. I think, however, that power and force come to fore in autocracies, and had the most advanced societies of the recent period been such, left-wing social scientists in the ‘first world’ would have focussed more on power. Mao Zedong’s aphorism about power flowing from the barrel of a gun is not the most useful one for liberal democracy. Guns and force are used to control the most rebellious elements in society, but in advanced democracies, ideas are the main weapon in the class struggle.

4. *Looking at the development of your theorisation, it is possible to identify major themes that you have focused on in various periods of your career. Roughly (and without doing full justice to the complexity of your work), one can say that in 1970s and up to the mid 1980s you have focused on Analytical Marxism and*

exploitation theory; from the mid-1980s to the early 1990s the main theme of your research has been market socialism; from mid 1990s to the 2000s you have developed your theory of equality of opportunity as well as your analysis of politics; while from the mid 2000s to this day, the normative aspects of climate change have been at the centre of your interests. This is a broad and diverse set of issues. Is there a thread common to all (or most) of them?

As I said, I began as a Marxist, as a result of my parents' influence, and the radicalization of students during the Vietnam war when I was in college. I was led to analytical philosophy in my attempt to understand the foundations of the Marxian condemnation of capitalism based on exploitation. I became interested in political science because 'elections are the democratic form of class struggle,' as Seymour Martin Lipset wrote in 1960. And I explained earlier how my interest in climate change was natural, due to the conjunction of economic and ethical questions that it involves. This was not, however, a self-consciously chosen path. Perhaps the most important decision I made was to leave pure mathematics, as a graduate student, and to look for a more political way to use mathematics.

5. *The role of mathematics in economics is a vexed issue and it has been at the centre of recent debates on the financial crisis and the inability of economists to predict it. In your theoretical work, you have privileged an analytical approach, but you have not exclusively adopted the standard formal modelling tools, and have almost equally often opted for philosophical analysis. What is your view on the use of formal models in economics and especially in welfare economics?*

I've said earlier that I don't think the problem is with mathematics, it is with the premises of models, that are often taken for reasons of elegance or habit rather than realism. The rational expectations hypothesis is a case in point. Hyper-rationality is another (well, this may be the general category of which rational expectations is an instance): like Tolstoy's families, there's one way to be rational, and an infinite number of ways to be non-rational. Elegance and simplicity pull one towards the rational hypothesis. At some point in the future, there will probably emerge a 'standard model' of what we now call 'irrational' behaviour, and that will replace hyper-rationality.

I think welfare economics—more specifically, social-choice theory—has also entered a cul-de-sac. People construct ever more complicated models based on a few powerful and path-breaking ideas, such as the Arrow Impossibility Theorem, and Eric Maskin's implementation theorem for social-choice rules. We can observe this degeneration of a science, while not knowing how to combat it. I think Thomas Piketty's *Capital in the 21st century* is a path-breaker, both in its substance and in its demonstration that a great deal can be said without the use of complicated models or econometrics. I am hoping his book will set an example for economics students of how to do social science. Of course, Piketty has a mind in the 'top 1%' of economists' minds, and he made his *tour de force* look like child's play. Left-wing economists today, like Piketty, do not have to struggle to escape the Marxist methodological straight-jacket, as my cohort did, a generation earlier.

6. *You have often argued that the advent of information economics in the 1970s has produced a significant change in our views about the economy, shifting the focus of attention from the role of prices (and markets) as coordination devices to their role for incentivising economic agents. How do you see the relative importance of coordination and incentives in social organization?*

This, to me, is a fascinating and unsolved question. In the post-war period until around 1970, the price system, in economic theory, was viewed mainly as a coordination device. Decentralization of resource allocation was the focus. I took a field in 'economic planning' as a graduate student at Berkeley, in 1967, taught by Roy Radner and Bent Hansen, and this was the focus. The principal-agent problem was formulated (formally) around 1970, and quite quickly the focus turned to viewing prices as providing economic agents with incentives. General equilibrium, concomitantly, was dethroned and replaced by game theory.

The distinction can be illustrated with a thought experiment. Imagine an economy with a firm that produces the economy's single consumption good, and whose inputs are labor of many different occupations. Each worker has preferences over income (derived from wage labor in the firm), and the bundle of traits characterizing occupations, which has two elements: an element describing the nature of the occupation (say, its status in society, or the nature of the work), and the educational qualification necessary to perform the occupation. Facing a vector of occupational wages, each worker chooses an occupation to maximize her utility, balancing off these three attributes (wage, the nature of the work, education). A wage vector is an equilibrium vector if all occupational labor markets clear when the firm maximizes profits, taking the wage vector and the price of output (which is normalized to one) as given.

Now suppose that workers care very little about income, but a great deal about the bundles of characteristics of different occupations. There is an equilibrium. Suppose taxes are introduced to redistribute income. Then the occupational labor supplies will hardly change: they are inelastic with respect to income. The wage vector performs an important role: it aligns the profit-making behaviour of the firm with the *inelastic* occupational labor supplies, so that all labor markets clear. Thus taxes have no effect on production: they are purely redistributive, and impose no efficiency cost on the economy.

On the contrary, suppose, more conventionally, that workers care a great deal about income. Then taxation will alter occupational labor choices, and we will have the usual efficiency costs of taxation.

The first story is one where prices have mainly a coordination role, and the second is one where they have largely an incentivizing role. Which story is a more accurate parable for our economy?

To the extent that the coordination role of prices is important, distribution can be separated from efficiency, and that was more or less the background view of the market socialists, like Oskar Lange, in the earlier part of the twentieth century. To the extent that the incentivizing role of prices is primary, we get the principal-agent problem, and conservative arguments against state intervention, because taxes induce inefficiencies. I believe quite a lot hinges on the true roles of prices. There has been a shift in how economists view prices; this came about probably in part because of the failures of planning. I do not think economists have devoted the study to this question that it deserves. The first step would be to formulate the distinction more clearly.

7. *Your thorough exploration of the concept of Marxian exploitation led you to conclude that the key focus of normative interest in exploitation theory is the institution of private property and the distribution of productive assets. In particular, you abandoned the notion of exploitation as the unequal exchange of labour (UE) in favour of a game-theoretic definition based on property relations (PR). In some later, and less well-known contributions, however, you have reconsidered this conclusion. To be precise, you have acknowledged the limits of PR and proposed that 'an agent is exploited in the Marxist sense, or capitalistically exploited, if and only if PR holds and the exploiter gains by virtue of the labor of the exploited' (Roemer 1989, p. 96). Would you like to explain, and perhaps tell us if there is a connection with your work on the proportional solution?*

I'm not particularly interested anymore in puzzles about exploitation. I do think the proportional solution is an important discovery; although I was the co-author on the article that presented it (in 1993), the idea is due to Joaquim Silvestre. We were thinking about how one might formulate, in a general way, what the natural allocation of labor and commodities would be in an economy with public ownership of firms, in the sense that the Walrasian allocation is the natural one in a private-ownership market economy, and we came up with several proposals, the most important of which is the proportional solution.

Think of an economy with one publicly owned firm, producing one good, with many workers. According to the old socialist mantra, there should be 'work according to ability and pay according to work.' The natural idea, then, is that income should be proportional to labor time (or perhaps efficiency units of labor time). Question: Is there an allocation with this property that is also Pareto efficient? We proved that in very general economies, there is. The idea generalizes to an economy with many firms, many commodities, and many kinds of labor.

I don't advocate this as the *just* distribution, because one has to inquire into the process by which skills were formed—were opportunities equal for the formation of skill, and so on. But it is, nevertheless, an obvious question that arises out of socialist economic theory, and it is quite amazing that the question was not asked until 1993! I think this tardiness in addressing the question was due to neoclassical economists' lack of interest in payment in proportion to labor, and Marxist economists' ignorance of neoclassical economics, including the idea of Pareto efficiency, which they may have mistakenly viewed as unimportant.[1] Still, one wonders why Oskar Lange did not ask the question, as he knew both approaches well.

The proportional solution has an important relation to the tragedy of the commons. Think of a community that extracts a common pool resource (CPR), such as fish from

[1]For instance, Maurice Dobb once argued that the idea that markets bring about Pareto efficiency is obvious.

a lake owned by the community. Each fisher has preferences over labor performed and fish consumed. If each fisher keeps his catch, then the allocation of fish will be proportional to the efficiency units of labor performed by fishers (random variation aside). The tragedy of the commons is often illustrated by the fact that (under standard assumptions of decreasing returns to scale in production and convexity of preferences), the Nash equilibrium of the game among fishers, whose strategies are labor supplies, is Pareto inefficient. Fishers fish too much. The proportional solution for the economy is the solution to aim at: it preserves the feature that each fisher keeps her catch, *and* is Pareto efficient. Such an allocation is locally unique.

Question: Is there a way the fishers can access the proportional solution? It turns out that there is, by using the protocol of 'Kantian optimization.'

8. *Looking at the articles published in general journals, as well as at the rankings and impact factors of field journals, it seems that the field of social choice and welfare has lost much ground, especially as compared to its heyday in the 1970s and 1980s. In your opinion, what are the determinants of this evolution, both outside of the field and within it? Is there something that social choice theorists and welfare economists should reconsider of their field in order for them to reach a broader audience?*

I think the new interest in inequality, due to what has been happening to income and wealth distributions in the advanced world, will eventually change what people work on. The most dramatic example to date is Piketty's book. I hope social-choice theorists take a lesson from this fine book.

9. *A significant impact on your worldview seems to derive from General Equilibrium Theory. As you noted in one of your earlier works, "Like many economists of my generation, I am strongly influenced by the power of the equilibrium method: of examining a model when it is at rest, so to speak, in the sense that all the rules that describe how its parts work are simultaneously fulfilled." (Roemer 1981, p. 10). Is this influence still relevant? And what are the advantages of analysing the economy in general, and distributive justice in particular through the Walrasian lens? What are the limitations, if any, of the Walrasian, general equilibrium outlook on the economy and more specifically on welfare economics?*

A general equilibrium model does not have to be Walrasian. One might model economies with some kinds of non-private ownership. The general equilibrium model is the model of income distribution. It leaves some important things out, such as the formation of preferences, and the origin of endowments, but it focuses on something of great importance.

We have done almost nothing to understand what an equilibrium would look like in an economy where people are motivated by more public or cooperative concerns. Surely, we can do far better than equip people with 'altruistic preferences' and then continue with utility maximization as before. (In particular, we immediately face a free-rider problem: if the economy is large, then the welfare of others, if I am an altruist, is a public good, and the free-rider problem bites. Any contribution I make to the welfare of others has only a trivial effect, while its cost to me may be non-trivial.) This suggests that cooperation, if it comes about, must involve some other *optimization protocol*. In reality, people often overcome free-rider problems, which is to say, they do not play the Nash equilibrium in the relevant game, but behave 'collectively,' somehow. In Britain during World War II, many people 'did their bit,' although this was not Nash behaviour. I do not think this is explained by the 'warm glow' they got from 'doing their bit,' but rather from some kind of cooperative thinking. Can we use examples like this to conceptualize what optimizing behaviour might look like in a more solidaristic society? Can we then use *that* to understand what an economic equilibrium would look like? What would be the welfare properties of that equilibrium? I have proposed recently that cooperation may involve people engaging in 'Kantian optimization.'

10. *A key principle of the theory of equality of opportunity is that "there is … a 'before' and an 'after': before the competition starts, opportunities must be equalized, by social intervention if need be, but after it begins, individuals are on their own." (Roemer 1998, p. 2) Robert Sugden (2004) has criticised the modern theory of opportunity precisely because of this aspect (which he attributes to its Walrasian roots). In his view, one of the key principles of the theory of equality of opportunity—namely that equal efforts should yield equal rewards—is incompatible with the allocation of resources through markets because of imperfect information, and in particular the division of knowledge. Therefore, argues Sugden, equality of opportunity cannot really reconcile egalitarianism with principles of freedom and responsibility. What is your answer to this criticism?*

I am not familiar with this paper of Sugden. It is true that markets reward skill and not effort, and this is one important reason that it is difficult to equalize opportunities in a market economy. Probably James Heckman's focus is the best one from a policy viewpoint: upon intensive pre-school education to compensate children from disadvantaged families for the paucity of family resources. Nevertheless, taxation and social insurance remain an important tool for repairing unjust disadvantage among adults. I am heartened by recent work of some public economists arguing that the optimal tax rate on high incomes in the United States is over 80%. The low tax rates on high incomes that we currently have in the US are, in my view, purely due to successful conservative propaganda, and are not justified for incentive reasons. The idea that high marginal tax rates on high incomes would destroy innovation is a big lie. How does Sugden respond to the successful redistribution of the Nordic economies, using not only taxation but also the solidaristic wage?

11. *You have argued that "welfarism is a methodologically unacceptable premise in studying distributive justice because it forecloses the possibility of modeling property rights" (Roemer 1994a, p. 199). What is your view of welfarism and what is the additional information you now consider most relevant to consider in addition to welfare?*

A theory of justice is welfarist if the only information it uses to rank social alternatives are the welfare levels of individuals in those alternatives. Any theory of justice that involves knowledge of rights or rights violations is therefore not welfarist, because rights are always specified in non-welfare terms. In the theory of equal opportunity, it does not suffice to know welfare levels, or outcomes (more generally); one must know the *efforts* that people expended, and their *circumstances*. This is non-welfare information.

As I believe the theory of equal opportunity is the most useful theory of distributive justice at present, I think knowledge of the circumstances of individuals is of key interest. To know whether an income distribution is reasonably just, it does not suffice to know the Gini coefficient, or any statistic that can be computed from knowing only the distribution of income: one must also know the extent to which individuals had equal opportunities for earning income. The term of art for influences on one's income-forming capacity that are important and beyond one's control is 'circumstances.'

12. *Much of your earlier thinking (including Roemer* 1981, 1982, 1994a, b*; Roemer and Silvestre* 1993*) was focused on the efficiency and fairness characteristics of systems of public ownership of (external) resources (but private property of self including skills) as opposed to the capitalist regime with private property. Is this still an important topic today (e.g. after the fall of the Berlin wall and the extension of the capitalist system in the whole globe)? And would you still agree that "The private property system is just one possible way of organizing economic activity; it may have been the best way for a certain period but is probably not the best way today, nor will it be in the future" (Roemer* 1988*, p. 11)?*

I very much adhere to this view. I believe that property relations are in flux, and it is hard to know what forms will be important in a century or two. The socialist experiments of the twentieth century were important. The evident obsolescence of certain forms of property that has occurred because of the information revolution is interesting. I am still a believer in Einstein's *bon mot*, that socialism is mankind's attempt to overcome its predatory phase, and I believe that attempt will continue. The move to the right that we have seen in the last forty years is still, historically speaking, a short moment. It is very hard to predict what will happen. Perhaps we will succeed in addressing climate change with global cooperation—of course, I hope so—and in so doing, a more cooperative period will be inaugurated. Perhaps there will be a century of increasingly disruptive financial crises, owing to democracy's inability to regulate finance capital, and this will create social movements against capitalism. Perhaps the increasing concentration of income and wealth, predicted by Piketty, will bring about social movements that will inaugurate fundamental economic changes. Given the social history of our species, it would be myopic to predict that massive institutional changes are a thing of the past.

13. *Much of the recent economics literature on inequality has focused on earnings differentials, emphasising how the increase in inequalities observed since the*

early 1980s in most advanced countries derives from differences in skill and talents. The recent empirical work by Piketty questions this view and suggests that the lion's share of current inequalities derive from unequal wealth holdings. What are the implications of this finding from the viewpoint of welfare economics and distributive justice?

Piketty predicts that the twenty-first century will be one in which inheritance comes to be the primary factor in the distribution of wealth. Indeed, he thinks the twenty-first century will look more like the nineteenth century than the twentieth century in terms of class. However, these predictions would be upended if there are major crises, or world wars, which destroy capital, as occurred in the period 1914–1944. One of the main lessons of Piketty's analysis is that that period comprised a reversal in the historical trend of the increasing power of capital, but a temporary reversal. I view Piketty as making predictions of increased wealth concentration under 'business as usual:' but I think it is unlikely that the century will be characterized by business as usual. Piketty's warning—that without social and political organization, we are headed for a more unequal capitalism than ever—is of great importance.

14. *In your book on Equality of Opportunity (Roemer 1998), you present your formalisation of Equality of Opportunity as "an algorithm which will enable a society (or a social planner) to translate any such view about personal accountability into a social policy that will implement a kind or degree of equal opportunity consonant with that view. If my algorithm is generally accepted as a reasonable one, then the political debate over what equality of opportunity requires can be reduced from one over social policy to a more fundamental debate about the proper realm of individual accountability" (p. 3). What is your view about such accountability? Where should the cut between circumstances and responsibility be placed?*

I don't think it is my job to answer that, nor is it particularly important that we do. I did not propose the equal-opportunity theory as a complete theory of justice, but rather as a useful algorithm. The title of my 1993 article on the topic contained the phrase 'pragmatic planner.' My view is that in most societies, implementing policies that take as circumstances attributes of a person's environment such as the education of his or her parents, will lead to dramatically more equal outcomes. In other words, we do not have to solve the metaphysical problem of what in fact people are responsible for. For the time being, we can stick to circumstances for which people are obviously not responsible, yet have great impact on their chances in life. Besides the education of one's parents, race or ethnicity and sex are obvious circumstances with respect to which we should equalize opportunities for wage earning capacity.

I would personally also include measures of a person's native talents as circumstances. And I would not consider children responsible for anything below an 'age of consent.' If a society sets the age of consent at, say, age 14, then the accomplishments of the child at age 14 should be taken to be circumstances in looking at his or her opportunities as an adult.

But these are more radical positions. I think we will have made a great advance towards justice if we compensate persons for the disadvantages due to their social environments only—even only for the disadvantage inherent in coming from a poorly educated family.

15. *Various strands of your research put a significant emphasis on culture, preferences and education. For example, in Roemer (1988) your discussion of the morality of wealth inequalities leading to exploitation focuses on savings behaviour and you note that this is not necessarily exogenous. Education and preferences are also central in your discussion of personal responsibility in the context of the modern theory of equality of opportunity. You have not really formalized the formation of preferences. Why?*

I have occasionally tried to think about endogenizing preferences, but I have never succeeded in saying anything interesting about the problem. In terms of equality of opportunity, there are some difficult questions about responsibility. Ronald Dworkin believed that a person should be held responsible for his preferences, as long as he comes to identify with them. Thus, if a woman forms preferences for a life with truncated goals, because of being brought up in a sexist environment, and she comes to identify with these preferences (the 'contented housewife'), Dworkin would not believe she is due any compensation. Or at least that is what his theory says. My own view is that a person should not be held fully responsible for preferences that were formed as a natural reaction to an unjust environment. Of course, this is paternalistic. The 'nudge' approach to paternalism, of Thaler and Sunstein, is certainly one step, but it is not sufficient. I do not have a well worked out solution to this problem, but I do not think Dworkin's solution is the right one.

16. *One of the key innovations of the modern theory of equality of opportunity is its reconciliation of the ideals of equality and responsibility, traditionally considered as antagonistic. The theory of equality of opportunity is now one of the main contenders in political philosophy. However, two criticisms have been raised against it. First, the reconciliation of equality and responsibility has come at the cost of marginalising other normative values and ideals that have traditionally played an important role in egalitarian approaches such as community and self-actualisation. (See, for example, Cohen's (2009) emphasis on community and your own emphasis on self-actualisation in your theory of market socialism.) Another criticism often moved against equality of opportunity is that the emphasis on responsibility has somewhat crowded equality out. Some have argued that an allocation consistent with the equality of opportunity algorithm would likely be less egalitarian and also significantly less forgiving than, for example, existing welfare states. What is your reaction to these objections? Should considerations about outcomes be put back into the picture? As acknowledged by Richard Arneson, the theory of equality of opportunity "is blind to results once equal opportunities have been provided... [but] in some circumstances the refusal to tender more resources is unfair" (Arneson 1994, p. 225).*

Let me deal with the second issue first. The EOp theory involves both an equalizing and a disequalizing component. We want to *equalize* outcomes for those who expend comparable effort, but who have different circumstances, and we want to permit *inequality* for those with the same circumstances who expend different efforts. I put the emphasis on the former. Some philosophers, such as Elizabeth Anderson, put the emphasis on the latter. I believe that if we equalized opportunities for wage-earning capacity in a population, with respect to the circumstances I listed above, we would have dramatically more equal distributions of income. One need only look at the societies which have done the most to equalize opportunities, according to my definition: they are the Nordic ones, and have the most equal income distributions in the world. Some philosophers focus too much on cases that affect a small number of people—examples such as irresponsible motor cyclists who crash while not wearing helmets. I am interested in the massive economic inequalities associated with children's being raised in families with massively different resources.

I agree with Jerry Cohen that we have other important values (than equality of opportunity), such as community, and community will be compromised if income differences are too great. I am sympathetic to Cohen's view that therefore the goal of community should constrain degrees of inequality that might otherwise be forthcoming in an opportunity-egalitarian society.

Let me give another example of competing values. In healthcare, we have a strong norm of horizontal equity, which here means that two people with identical illnesses should be treated in the same way by the hospital, regardless of their circumstances. (We should devote no less care to the lung cancer victim who acquired the disease as a result of smoking than to the victim who never smoked.) One reason for such a norm is that we do not want healthcare personnel to make distinctions between patients that will possibly harm the provider-patient relationship. This means we will not be holding people as responsible for their outcomes as we possibly could. These norms are, however, important, and they should *constrain* the set of feasible policies. I have shown how to do this in several articles.

An equal-opportunity problem is specified by naming the *objective* function, the *circumstances*, the *policy space*, and the *kinds of effort* that are relevant. There is sensitivity required to other social values in specifying all these components of the problem. Critics of the approach often do not understand this. It was therefore, perhaps, unwise of me to call my formulation of equality of opportunity as an algorithm, as that word implies that the policy application is more cut-and-dried than, in fact, it may be.

17. *Another aspect of the theory of equality of opportunity that is at first sight slightly at odds with some of your earlier work is the fact that in designing the optimal distributive benchmark you take the fundamental organisational structure of the economy essentially as given. This seems in contrast, for example, with your work on public ownership or market socialism. Thus, it appears that, from an equality of opportunity perspective, a society with a very questionable (perhaps feudal) social structure in which however people are assigned to the various social positions randomly is in principle as just as one in which the social*

structure itself (and the resulting outcome inequalities) is less unequal. To put it differently, can the theory of equality of opportunity discriminate between a situation in which every child has equal chances of living in a slum and one in which slums are eliminated?

As I said, I intend the EOp theory as a pragmatic approach to responsibility, that can be applied by any society to equalize opportunities with respect to its own conceptions of responsibility. Today, most people, even in the advanced democracies, believe that individuals should be rewarded for their natural talent. Thus, talent would not be a circumstance in EOp policy for these societies. I believe talent *should* be a circumstance—in my view, those valuable attributes that persons possess as a matter of luck of birth are ones for which they do not *deserve* reward. I would be happy, however, if our societies equalized opportunities with respect to educational and wealth attributes of families into which persons are born. Equalizing opportunities with respect to differential native talent will probably have to wait. This is what it means to say that the EOp approach that I advocate is political, not metaphysical, to plagiarize J. Rawls.

The example of children in slums seems to me a clear one where the specification of the policy space is key. A reformist approach would look at policies that could improve the living conditions of poor families without massive social interventions that would prevent poverty, or even the poverty of these families. A more radical approach would take the policy space to include these massive interventions. My contention is that even specifying a fairly restricted set of circumstances, and a fairly restricted policy space, will lead, according to the theory, to quite significant policy reform.

18. *Compared to more general analyses in political philosophy, theories of justice in economics tend to focus on issues of distribution. This implies that in the main normative economics tends to neglect other important normative values and ideals, such as self-realisation, community, dignity, republican citizenship, and so on, but also issues of power or force that are central in debates in political philosophy. Does this make our reflection unduly narrow, or can it be rationalised as a simple matter of disciplinary specialisation? And what may explain this neglect?*

I talked about power and force in question 3. The EOp theory addresses the issue of economic distribution. I have not attempted to formulate a general theory of justice. Of course I believe these other issues are important. I continue to believe, however, that economic distribution is of paramount importance. I do not think one can have a well functioning democracy or republic, or a society in which all have dignity, in the face of massive economic inequality. For instance, I question Rawls's placing lexicographic priority on liberty over economic equality. In principle, I am willing to consider trade-offs between liberty and the economic level. If some autocracy was necessary in China to raise 300 million people out of poverty, as occurred in the last thirty years, was that possibly not an acceptable trade-off? I am not saying this *was* the case, but to exclude consideration of such questions ex ante strikes me as incorrect.

19. *In your views, what are the main research questions that one should study in normative economics now? In economics more generally?*

I do not like to make such prescriptions. I see how my own work has proceeded in a meandering way, and I do not think we humans possess the imagination to plan a research program for the next 100 years. We face problems as they develop, one step at a time. It is idealistic to think we can do much better than this. The particular problem that now excites me is how to think about cooperation. I maintain that if you ponder this a bit, you see how little economic theory has said about it, yet how important a topic it is. The consensus among biologists and psychologists is that the ability to cooperate is the key distinguishing feature of homo sapiens in the family of great apes. But our economic theory is only about competition. It treats cooperation as a special case of competition (in the sense that cooperative behaviour can sometimes be sustained as Nash equilibria of a complicated game). I find this an unnatural explanation, one that is probably fundamentally wrong.

20. *In your work on politics, you have explored the way in which democratic parties seek the vote of the electorate for a greater variety of reasons than just getting into office, and you have also studied the internal factions within parties. (A) What motivated your interest in this topic? (B) How do you relate this work to your work on the economy? (C) What is your view of the best way of organizing collective decisions? How does democracy fit into your conception of a good society?*

My work on political parties and elections has been positive, not normative. I found the existing formal theories of political competition, exemplified by the Hotelling-Downs model, to be apolitical and internally inconsistent. (If both candidates propose the same policy, and neither has any policy preferences, why should any concerned citizen vote for one or the other?) I wanted to put partisanship and representation into the formal models of political competition, which I did with the concepts of endogenous party Wittman equilibrium (in collaboration with Ignacio Ortuño-Ortin) and of party unanimity Nash equilibrium (PUNE). The Hotelling-Downs model, as well as being very special (it possesses equilibria only with unidimensional policy spaces) is apolitical. It was a bad model that dominated the field because formal theorists proposed no alternative.

Democracy is, I think, the only mechanism we know for holding politicians accountable. Of course, accountability is imperfect. Roughly speaking, democracy is increasingly compromised to the extent that the income and wealth distributions become more unequal. The Citizens' United decision by the US Supreme Court concerning campaign finance is a fairly clear move away from democracy, as it empowers the political 'voice' of money. It is interesting to see that a fair number of political scientists say that the decision will not have much effect. Even if this turns out to be true, the point is irrelevant: as a matter of principle, the role of moneyed interests (although not necessarily of money) should be sharply limited in politics.

21. *If only one of your contributions could be passed to the next generations and be influential, which one would you like to prioritize?*

Let history be the judge.

22. *For the first time in history, people across the globe have marched to underline the urgency of measures to tackle climate change. One of the slogans of demonstrators worldwide referred to "climate justice". What is "climate justice" and what kind of normative issues does climate change raise?*

There are two principal issues in climate justice. The first is intergenerational: what is the just way to share the scarce global resource of a clean biosphere across generations? The second is intragenerational: what is the fair way to share the responsibility of reducing greenhouse gas emissions across regions of the world today? In particular, how should the sacrifice of economic growth that will be necessary to reduce greenhouse gas emissions be spread across the global North and global South? My co-authors and I believe that a politically feasible solution to the second problem requires that the dates at which developing countries (China, India) catch up to the global North in income per capita not be delayed from what they otherwise would have been. We argue that these dates can be preserved, economic growth can continue at a reduced rate, and global emissions of greenhouse gases can be constrained to stay below a reasonable limit. However, the rates of economic growth in the global North and South must be reduced from what they would have been under 'business as usual.' The political task is to achieve international agreement—a cooperative solution!

Bibliography

Arneson, R. (1994). What do socialists want? In E. O. Wright (Ed.), *Equal shares*. London: Verso.

Cohen, G. A. (1978). *Karl Marx's theory of history: A defence*. Princeton: Princeton University Press.

Cohen, G. A. (1999). Marxism after the collapse of the Soviet Union. *The Journal of Ethics, 3,* 99–104.

Cohen, G. A. (2009). *Why not socialism?*. Princeton: Princeton University Press.

Hurley, S. (2001). Luck and equality. *Proceedings of the Aristotelian Society, 75*(Suppl.), 51–72.

Piketty, T. (2014). *Capital in the 21st century*. Cambridge, MA: Harvard University Press.

Rawls, J. (1971). *A theory of justice*. Cambridge, MA: Harvard University Press.

Roemer, J. E. (1981). *Analytical foundations of Marxian economic theory*. Cambridge, MA: Harvard University Press.

Roemer, J. E. (1982). *A general theory of exploitation and class*. Cambridge, MA: Harvard University Press.

Roemer, J. E. (1988). *Free to lose. An introduction to Marxist economic philosophy*. Cambridge, MA: Harvard University Press.

Roemer, J. E. (1989). What is exploitation? Reply to Jeffrey Reiman. *Philosophy & Public Affairs, 18,* 90–97.
Roemer, J. E. (1994a). *Egalitarian perspectives. Essays in philosophical economics.* Cambridge: Cambridge University Press.
Roemer, J. E. (1994b). *A future for socialism.* Cambridge, MA: Harvard University Press.
Roemer, J. E. (1996). *Theories of distributive justice.* Cambridge, MA: Harvard University Press.
Roemer, J. E. (1998). *Equality of opportunity*. Cambridge, MA: Harvard University Press.
Roemer, J. E. (2010). Kantian equilibrium. *Scandinavian Journal of Economics, 112,* 1–24.
Roemer, J. E., & Silvestre, J. (1993). The proportional solution for economies with both private and public ownership. *Journal of Economic Theory, 59,* 426–444.
Roemer, J. E., & Wright, E. O. (Eds.). (1996). *Equal shares*. London: Verso.
Sugden, R. (2004). Living with unfairness: The limits of equal opportunity in a market economy. *Social Choice and Welfare, 22,* 211–236.

William Thomson

Youngsub Chun and Christopher P. Chambers

Questions asked by Youngsub Chun and Chris Chambers

Q: Most of us know the story, but would you explain how a Frenchman ended up with the name William Thomson?

A: Having a Scottish paternal grandfather has a lot to do with it.

Q: How did you get interested in fairness?

A: How can one not be interested in fairness? Open any newspaper, turn on your television! Most conflicts around the world are caused by unfair treatment of certain individuals, populations, or by perceptions of unfair treatment. It is the lack of interest in distributional issues among much of the economics profession that needs to be explained.

I conjecture that an important reason why economists shy away from the topic is the fear that value judgments would have to be made, and value judgments are unscientific. Yet, understanding the logical implications of value judgments is value-free scientific work. Value judgments will ultimately have to be made but the economist can leave them to the policy maker, the bankruptcy judge, the school administrator, …, the citizenry at large.

Q: In your view, how does the theory of fair allocation relate to welfare economics and public economics?

A: I am not sure what the canonical definitions of welfare economics and public economics are. Perhaps public economics has traditionally been focused on taxation, and it is usually studied in the context of market economies, where prices are used to allo-

Y. Chun (✉)
Seoul National University, Seoul, South Korea
e-mail: ychun@snu.ac.kr

C. P. Chambers
Georgetown University, Washington, DC, USA
e-mail: cc1950@georgetown.edu

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_15

cate resources. Taxation and the financing of public goods are the essential concern. The theory of fair allocation, as I perceive the subject anyway, encompasses many problems of distribution of privately appropriable resources when these resources are described in a concrete way.

Consider the following array of problems. A collection of indivisible resources, often called "objects," has to be distributed to a group of people, say offices to coworkers in an office building. An interval of time has to be partitioned into subintervals during which a service is provided to a group of people; only one person can be served at a given time, and each person values subintervals differently, their lengths being one but not the only consideration. A facility has to be located on a road network; its users are distributed at different points of the network, and each user values its proximity to the facility differently. The cost of a service has to be shared among a group of people when their uses of it are nested; it could be a cab that coworkers hire to return home and they are dropped off at their respective homes; or an irrigation ditch that farmers use jointly and the costs of bringing water to each of them separately differ, due to their greater or lesser proximity to the main channel.

These are just a few examples of situations where fairness issues arise, but fairness actually come up in almost any type of allocation problem. These types of microeconomics problems, specified at this level of detail—I should perhaps speak of nanoeconomics—are not within the purview of standard welfare economics and public economics.

Much of the theory of fair allocation has been recently written from the design viewpoint, and the axiomatic method has been a principal mode of investigation. Again, this approach is not typical in traditional welfare economics and public choice. When engaged in this kind of work, we start from scratch. We try not to be too influenced by common practice, although of course we draw lessons from it.

Q: What was the environment like at Ecole Polytechnique? At Stanford? How did your education shape your way of thinking about economic problems?

A: At that time, economics education was very limited in French schools in general. At the high-school level, there was no such thing as economics.

At Ecole Polytechnique, the offerings amounted to very little. One course was compulsory, on national accounting. A few very specialized seminars were also offered. I attended one on general equilibrium, in which I discovered G. Debreu's *Theory of value*. I was surprised that economics could be treated in such a formal way. The rigor appealed to me.

In the early 70s, Stanford was one of the rare places where you could learn about game theory, but it wasn't part of the regular Ph.D. program yet. I am not sure it was part of any Ph.D. program anywhere in the world at that point actually. A main attraction was the IMSSS workshops, which lasted a significant part of each summer. They were devoted to general equilibrium and game theory. Some of the most prominent people in these fields attended them. K. Arrow, F. Hahn, and R. Aumann were regular participants. G. Debreu also made appearances.

Q: As one of the pioneers of the axiomatic approach to economic environments, we were hoping you would be able to explain your ideas behind the basics of the approach, how it should be implemented, and why it is useful for economic science.

A: I published an essay about the axiomatic method (Thomson 2001a), and I am currently at work on a larger, more ambitious and comprehensive, pedagogical exposition of it (Thomson 2021). I present my view of what the axiomatic program is about, and how to conduct axiomatic work. I also discuss the pitfalls of axiomatic analysis, how *not to do* axiomatic work.

Why it is useful to economic analysis should be clear. I cannot think of a better way of addressing questions of design. Economic institutions are not God-given, they are man-made. The axiomatic method provides complete answers to design questions. It starts with the axioms, which are the reasons, written in mathematical form, why we are interested in allocations and allocation rules. An axiomatic study addresses the compatibility of social objectives embodied in a list of axioms, when imposed in various combinations. When axioms in a list are compatible, it usually offers a description of all the rules satisfying them. The goal of the axiomatic program as a whole is to trace out, for each class of problems that one may encounter, the boundary that separates those lists of properties that can be met together from those that cannot, and when compatible, to obtain as explicit as possible a description of the rules that satisfy them all.

Q: You have demonstrated a particular interest in "variable population axioms", that is, in axioms that relate the behavior of economic systems across populations. Can you explain us a bit of the ideas of these ideas. For example, where does the idea of population monotonicity come from?

A: It is the idea of solidarity that underlies most of the monotonicity axioms. Put abstractly, it says that when a parameter of an economy changes and if no one is responsible for the change—no one deserves any credit when it is socially beneficial (that is, permits a Pareto improvement) or any blame when it is not (that is, when no such improvement is possible)—everyone's welfare should be affected in the same direction: if someone is made better off, no one should be made worse off. The parameter that changes can be resources, technologies, preferences, population…In each of these applications, solidarity takes the form of a separate axiom, often a monotonicity axiom if the space to which the parameter belongs is equipped with an order structure. Population monotonicity is such an axiom.

Solidarity is enshrined in most religions and in the moral codes of most societies. It is part of our ambient ideology. Passing from the general idea of solidarity to our specific solidarity axioms requires only a small step. I don't see a significant conceptual difference between its application to variations in resources, technologies, or preferences, and its application to variations in populations. It is true that the fact that population is a discrete variable makes it somewhat special because it is a little easier to imagine applying different rules for different populations than to doing so for different endowments of unproduced resources or to different technologies. Mathematically, this has significant implications, however, because one has to work in a model in which the dimension of the space of allocations is not fixed. It depends on the population size. This certainly makes it more challenging.

Q: You are known for having spent long hours advising your students. Did you apply any of your ideas on allocating time to this particular problem? How would you suggest scientists balance their time between students and research?

A: I wish I could say that I had invoked the axiomatic method to solve this kind of problems, or any of the problems I have faced in my daily life, but there probably would have been too many variables to take into account. The axiomatic method does have limitations, and one of them is that the problems to be solved should not be too complex; their description should not involve too many parameters.

As for how a young professor should balance teaching and research, I would say focus on your research. Time goes very fast, the publication process is very slow and subject to a lot of randomness. Soon, requests for letters evaluating your work will be sent out by your department to the best scholars in your field. So, make yourself visible in the profession: circulate your research, attend conferences, do a good job with your refereeing assignments, although here I would say, do not feel obliged to always say yes. Getting too deeply involved in teaching is dangerous, and advising graduate students should not be your priority. If you can engage in some joint research with the better ones, you may do so however.

Q: Do you have any advice for young researchers interested in the axiomatic approach?

A: Earlier, I mentioned two pedagogical pieces on the subject, one journal article and the other a book size manuscript. In both, I discuss how I think axiomatic work should be carried out. I also address my pet peeves about the way it is sometimes done. So here, I will only emphasize one point, which is that the focus should be on the economic content of axioms. Will the man on the street understand the idea that underly them? Will he endorse it? The answers should be yes. At least, there should circumstances in which he would. The relevance of each axiom does depend on the context.

However, one should also be ready to accept the technical help that some axioms provide. Occasionally, in order to make progress on some otherwise intractable model, one has to invoke properties that may not be as compelling as one would like. I find that to be a reasonable research strategy. With time, as one' s understanding of the subject progresses, one may be able to do without these crutches.

Q: If there is a student interested in what you are doing, which papers do you recommend to read? Why?

A: I was fascinated by Nash (1950)'s paper on the bargaining problem when I was a graduate student and I suspect that this paper remains an excellent introduction. Arrow (1951), Sen (1970) were important to me, and I recommend the more recent books by Moulin (1988) and Young (1994). Several areas of economic theory have developed in the axiomatic mode and showcase the power of the method: two-sided matching, with applications to the assignment of medical students to residency programs; "school choice," which has to do with the assignment of students to the various schools in a school district when each student—rather his or her parents—has in mind a ranking of the schools and each school has its own ranking over students; allocating scarce organs to patients waiting for a transplant. In none of these applications money changes hands, but in the next two it does: assigning workers to jobs and specifying their salaries, and allocating frequency bands to telecommunication companies. All of these areas have been the object of important design literatures, and axiomatic reasoning has been critical in their development.

These are many situations where making allocation decisions through prices is not a good option, or is not an option at all. Sometimes markets are considered unethical; sometimes they are even illegal. Alternatives to markets have to be invented. Researchers in these areas refer to their program as "market design" but routinely open their seminar presentations or their research papers by pointing out those multiple situations where markets do not have the good properties that they enjoy in our textbooks or are simply unavailable, and that something had to be done about it. An expression that would better describe what they contribute to is "alternatives-to-markets design."

It is not only resource allocation that is the object of interesting axiomatic work. Take the study of ranking methods for example. Rankings are more and more pervasive in our daily life. An ever widening range of important decisions are being made on the basis of rankings, of and by individuals, organizations, societies. That certainly includes academia: journals are ranked, universities are ranked, economics programs are ranked, researchers are ranked. Whether we like it or not, rankings play important roles, implicitly or explicitly, in promotion and tenure decisions (think of the profession's obsession with the "top-five"[1]), and in the allocation of research funds. The axiomatic literatures on ranking and voting are undergoing very interesting developments.

Q: In your book, you give guidance to young economists on how to write papers. Can you go one step further and tell young economists how to choose a topic and find interesting papers? Can you also make suggestions as to which field young economists should work on?

A: I don't have a recipe to find a good topic, but I would recommend to develop early on the habit of reading papers not to find out what's in them, but rather what's not there, what's missing. Because what's missing can be the subject of one's own paper.

Also, we should not ignore the discomfort we sometimes experience about some of the assumptions underlying a model when we are first introduced to it. We may feel that an author or a speaker's approach to a problem suffered from some important limitations; yet, we soon forget about our hesitation and go along with the approach. We should challenge it instead.

But the best topics, certainly the ones that are the most exciting to work on, emerge spontaneously and randomly, in the course of a conversation, when attending a seminar, when raking one's brain for homework or exam questions; also, when engaging in activities that are unrelated to research, for example, when dealing with bureaucracies to enroll our kids in school or standing in line to buy a concert ticket or a service.

Q: What do you think of the future of social choice? How about economic theory in general?

A: It is often said that economic theory is in decline. The power of new computational techniques and data availability has greatly increased the attractiveness of empirical work. I do not see a fundamental conflict between these approaches. Design work can be complemented with empirical and experimental work. The allocation rules that

[1] I should say "unhealthy" obsession.

we design have to be implementable on computers and opportunities for computer scientists and economists to interact have multiplied considerably in recent years. These rules should also be put to the test in actual practice. There may well be things that we worry about in our theoretical work but are not perceived as much of a problem in actuality, and conversely, problems that we dismissed as not being that important and that people care a lot about even though they occur rarely.

"Abstract" Arrovian social choice theory has often been criticized for mostly delivering negative results. But investigators should not be held responsible for the answers to the questions they ask if the questions are legitimate. It turns out that whether one ends up with positive or negative results depends in important ways on the structure of the set of alternatives, and on the properties that one can assume preferences satisfy. For the unstructured sets of abstract social choice, one is much more likely to end up with incompatible properties of social choice mappings than for the highly structured models that arise in the study of concretely specified resource allocation problems. These structures are reflected in restrictions on preferences that would not be meaningful otherwise; they allow us to formulate requirements of good behavior of allocation rules that one could not even discuss without these structures, and to define rules that could not be defined. Some rules often enjoy various combinations of these properties that would have no counterpart in unstructured models or whose counterparts would be incompatible. Of course, the introduction of "economic" structures into the model of traditional social choice theory do not necessarily lead to positive results. But the opposite would have been surprising. So theory allows us to understand that it is not good enough to speak about the desirability of social objectives in some abstract way.

Q: You have produced more than 50 Ph.D. students. What is the best way of teaching Ph.D. students? Can you give your advice to young professors?

A: I do not want to presume that I know enough about this to give advice to young professors. I have had the privilege of working in a top-notch department with a reputation for high-level theory and I will comment instead on the role that this has played in recruiting good students.

Lionel McKenzie started it all. He set the tone, established the standards of rigor that are the hallmark of our department. Thanks to him, generations of well-trained students have come to us to pursue Ph.D.'s. I have been the beneficiary of his pioneering work.

My book of pedagogical essays (Thomson 2001b) has a chapter entitled "Being a grad student in economics," and many of the things I collected there were lessons I drew from the mistakes I made when I was a graduate student. I should simply refer to that chapter, but keywords are "initiative" and "follow-up".

You don't have to be invited to show up at your advisor's door to tell them about the ideas you just got from reading an exciting paper; however, to make the best use of your advisor as a sounding board, be ready to present these ideas in as clear a way as possible: write down formally the definitions you will need; work out an example to illustrate a conjecture your formulated; refresh your memory of papers that seem relevant; prepare what you will write on the board. Also, don't forget that very soon you will be the expert on your subject, and you will have to guide your advisor as much as your advisor will guide you. These conversations do not have to be very long by the way. It is more important that they be frequent. And even more important, follow-up on these exchanges: read the suggested papers; try to prove the result that you discussed, to construct a counterexample to a conjecture, or look for the data to support some hypothesis you formulated.

Here are a few things that I feel students do not do enough of: reading papers that are not in their area of research; auditing classes when they are done with their course work; meeting with outside speakers and asking questions during their seminars; interacting with students in other fields. Also, faculty will certainly be very grateful for any help hosting visitors, organizing workshops and special events.

I said it before, but it bears repeating that as a young professor, you shouldn't get swallowed up by teaching, departmental service, the various committees on which you may be asked to serve, and student advising. It is more important for you to make yourself visible in the profession at large. Advertise your findings in your webpage and by posting them on the internet. Participate in conferences and perform your refereeing duties with diligence. However, learn to say no. Don't accept too many assignments; in particular, don't accept refereeing assignments that are too far afield from what you know.

Don't expect that every decision made about your submitted papers will be fair. There is much randomness to the publication process. Your papers will be rejected when they deserve to be accepted more often that the opposite will be true: as a young and unproved author, you won't be given the benefit of the doubt. Don't be discouraged. Your paper will eventually be accepted but it will take much longer than you expect.

Q: In recent years, you have spent a lot of time on the bankruptcy problem. Is there any special reason why you like this problem so much?

A: There are many special reasons.

First, let me use a slightly more general term than bankruptcy because bankruptcy is just an application of the formal model. I will talk about a group of people having claims on a resource when there isn't enough of the resource to fully honor all of their claims, and use the expression "claims problem."

It concerns an issue that all societies have to face, always have had to face. It is discussed in antiquity and medieval literature. How can one not be intrigued by Aristotle's view of it, the "contested garment problem," the "three-wives problem," and the recommendations the Talmud makes for them, by the "four-son inheritance problem" of Ibz Ezra, what he proposed as a solution for it, or by Rabad's generalization of this proposal?

Moreover, the data of a claims problem can be interpreted in the context of taxation, and taxation is also an issue that all societies have to deal with, that affects every citizen, from the man on the street to the investment banker.

Mathematically, claims problems constitute an extremely simple class: for an n-claimant problem, $n + 1$ numbers, a list of claims and an amount to divide, plus an inequality: the sum of the claims should exceed the endowment. Less than a line suffices to define it. You do not need any background in economics to understand what it's about. Everyone immediately gets it, economists and non-economists alike. And everyone has an opinion about it.

Next, and in part because of the simplicity of the model, many interesting rules can be defined to solve claims problems that enjoy a great variety of appealing properties. The richness of the inventory of available rules makes it in fact particularly critical, more critical than for other types of problems, that one should find ways of differentiating among them. The axiomatic method is the ideal way of evaluating them. And as always, it allows us to go further than what is done in practice, to find out if something better can be done, to identify those maximal lists of desirable properties that can be satisfied together; and when the properties are parameterized, to understand the tradeoffs between these parameters.

One could of course argue that the base model that I described is too simple, but it does cover a situation that is common enough. Again, think about bankruptcy and taxation. Besides, it has been enriched in countless interesting ways. I cannot think of a class of problems that is so important and so amenable to axiomatic analysis. I started working on claims problems 15 years ago, and if someone had told me that I would end up writing a book on the subject, I would have laughed: a book about $n + 1$ numbers related by an inequality?

Many positive results have been obtained but of course there are incompatibilities too. For some models, it turns out that very elementary axiom systems either lead to impossibilities or to very unpalatable conclusions. In other contexts, we often find that all the decision power should be given to a single person, leading to what we call "dictatorships." Monotonicity properties are particularly demanding in most contexts—strong domain restrictions often have to be imposed for positive results—but such properties are easily satisfied in the context of claims problems. Of course, it has to do with the model being described in terms of a small number of real-valued parameters. But the modeling is natural, and it does allow a large number of compatibility results. It is rewarding to work with a model that is blessed with so many good news.

So, claims problems are important; there is mystery, interesting mathematics, and the entry cost for researchers is extremely low. In fact, claims problems are an extremely powerful vehicle to introduce the axiomatic method to neophytes: they can only be described as a little pedagogical paradise [2].

Q: In the early days of your career, you focused on the axiomatic theory of bargaining. What motivated you to the study this class of problems. You do not seem to have done much research on the bargaining problem recently. What is the reason?

A: What is appealing about Nash's (1950) formulation of what we call Nash's bargaining problem is that it is very general, so in principle, any solution to the bargaining problem can be used to solve any type of allocation problem that one is interested in. It suffices to map the allocation problem into a bargaining problem, apply the chosen bargaining solution to get an outcome in utility space, and then return to the physical space of the allocation problem by taking its inverse image under the profile of utility

[2] For a survey see Thomson (2019)

functions. The study of the bargaining problem has been the object of considerable axiomatic analysis, so we have a very good understanding of which social objectives can be met jointly in this context.

There are several difficulties with that approach however. One is that the mapping from allocation problems to bargaining requires that choices be made and there are often more than one, especially as pertains to the disagreement point. And the eventual decisions to which we are led will often depend on these choices.

Another is that even if a solution to bargaining has been characterized in terms of a list of axioms expressing ideas that one finds particular compelling in one's particular application, the class of problems obtained as the image of the class of allocation problems under investigation may not be large enough to guarantee that this characterization holds.

A third is that the informational basis of allocation theory is quite different from what it is in Nash bargaining. First, in the passage from allocation problems to bargaining problems, the concrete "economic" structure of the space of alternatives is lost. As a result, two instances of the allocation problems under consideration may be mapped into the same bargaining problem. Yet, we may feel that the specific way in which they differ is quite relevant; these differences have to be ignored when working in the space of bargaining problems.

On the other hand, because Nash's bargaining problem is formulated in utility space whereas for most allocation problems, only the rankings of alternatives by the participants are available or deemed relevant or practical, some "cardinalization" of preferences need to be introduced to convert an allocation problem into a bargaining problem, that is, cardinal numerical representations of preferences have to be chosen. Obviously, the allocation that is eventually chosen will depend on that choice, injecting some amount of arbitrariness in the final choice.

Q: Has any graduate student not been able to complete the degree with you? What was the reason? Any recommendation to a beginning student?

A: I recall only one student who started working with me and finished with someone else. After a while, it became apparent that the match wasn't good.

Q: What are the good conditions for a good economics department?

A: I already mentioned the role played by Lionel McKenzie and the intellectual climate that he fostered, so let me mention more mundane things. How a department is physically configured is not to be neglected. The way ours is organized, on a single floor, with all offices opening on a circular corridor, and with two large rooms in the middle that we use for seminars and classes, is most conducive to people bumping into each other all the time. It would be even better if all of the graduate students' offices were also on that floor, but that would probably be impractical. Unfortunately, there are not enough office for that.

Having an easily accessible lounge with decent furniture, a coffee machine, and a blackboard is also a must. I have often been surprised to discover, when I visit other schools, that they only have a small coffee room, with no or little furniture, and that people have to stand in front of the coffee machine waiting for their turn; sometimes only a few people fit in the room.

Congeniality is of course the most important thing. People have to get along. The culture of the department has to be one of mutual respect. Exchanges between faculty and students have to be encouraged and facilitated. All of these things are true at Rochester.

References

Arrow, K. (1951). *Social choice and individual values*. New York: Wiley. 2nd edition in 1963.

Moulin, H. (1988). *Axioms of cooperative decision making*. Econometric society monograph. Cambridge: Cambridge University Press.

Nash, J. F. (1950). The bargaining problem. *Econometrica*, *18*, 155–162.

Sen, A. (1970). *Collective choice and social welfare*. Cambridge: Harvard University Press.

Thomson, W. (2001a). On the axiomatic method and its recent applications to game theory and resource allocation. *Social Choice and Welfare*, *18*, 327–387.

Thomson, W. (2001b). *A guide for the young economist* (2nd edition in 2011). Cambridge: M.I.T. Press.

Thomson, W. (2019). *How to divide when there isn't enough: From Aristotle, the Talmud, and Maimonides to the axiomatics of resource allocation*. Monograph of the Econometric Society. Cambridge: Cambridge University Press (forthcoming).

Thomson, W. (2021). *The Axiomatics of Economic Design*, Book manuscript (forthcoming)

Young, P. (1994). *Equity: In Theory and practice*. Princeton: Princeton University Press.

John A. Weymark

Felix Bierbrauer and Claude d'Aspremont

This interview of John Weymark (**JW**) was conducted on October 13, 2015, in Cologne, Germany, by Felix Bierbrauer (**FB**) and Claude d'Aspremont (**CdA**). The text of the interview has been edited to improve its readability, to clarify some of what was originally said, and to provide bibliographic details for the works cited.

CdA: Let me start with a biographical question. In your CV, you mentioned that you were born in Moose Jaw, Saskatchewan. Did you spend all your childhood and schooling there, and when did you move to Vancouver?

JW: I moved to Vancouver when I was eleven, in the middle of the school year. I did all of my schooling from what's called sixth grade—except for one year in Winnipeg—in a suburb of Vancouver, and then went off to the University of British Colombia (UBC).

CdA: So you were in Vancouver for most of your school education. Then you did your B.A. at UBC. Did you meet some future colleagues like David Donaldson there, and how did you decide to go to Penn for your Ph.D. afterwards?

JW: In response to the first part of your question, I met David Donaldson my first day as an undergraduate. At UBC at the time, most courses were a year long. It was a semester-based university, but they didn't split courses up a semester at a time. I took a special program called Arts 1. It was essentially a great books program that

F. Bierbrauer (✉)
Center for Macroeconomic Research, University of Cologne, Albert-Magnus Platz, 50923 Cologne, Germany
e-mail: bierbrauer@wiso.uni-koeln.de

C. d'Aspremont
Center for Operations Research and Econometrics, Voie du Roman Pays, 34, 1348 Louvain-la-Neuve, Belgium
e-mail: claude.daspremont@uclouvain.be

M. Fleurbaey and M. Salles (eds.), *Conversations on Social Choice and Welfare Theory - Vol. 1*, Studies in Choice and Welfare,
https://doi.org/10.1007/978-3-030-62769-0_16

was thematically oriented, interdisciplinary, and it counted for 60% of my credits. We read Plato, Dostoyevsky, Darwin, and Freud, among others. My section's theme was freedom and authority. The instructors came from a variety of disciplines. There was a group of 120 split into small groups of 20. One semester I had one group leader, and the second semester, I had a second group leader. We met once a week as a whole group.

To get us going, we went on a retreat where we read Conrad's *Heart of Darkness*, and that's when I met David. He was my group leader for the second term. David is a remarkable person and very intellectually engaging. It's because I found him so interesting that I decided in my second year to actually take an economics course. That's how I ended up in economics.

David Donaldson subsequently became a co-author, and then later when I was on the faculty at UBC, of course, a colleague. He's the person from whom I first learned about Arrow's Theorem (Arrow 1951) as an undergraduate. There were many faculty who were very influential, like Chris Archibald and Erwin Diewert. The course I took from Erwin is where I learned duality theory as an undergraduate.

There is another connection that, who would have known, played a big role in my career. As an undergraduate, you can't really understand a research seminar. But somebody—it might have been Erwin—told me that Peter Diamond was coming and what he would be talking about is really important, so I might want to go hear him. The Diamond and Mirrlees optimal commodity tax papers (Diamond and Mirrlees 1971) had just come out and Peter talked about them. I went to the seminar. I didn't understand a whole lot, but I got some of the basics because I'd learned some basic welfare economics from Chris Archibald.

So I had learned about Arrow's Theorem and social choice from David and went to my first research seminar, and it turned out to be on optimal commodity taxation. Also, David Donaldson had me read John Rawls' first paper (Rawls 1958) where he started working out his views on original positions and a theory of justice. All that was as an undergraduate.

CdA: Wow, that's impressive! And did the decision to go to Penn for your Ph.D. come later, or did you decide immediately to do a Ph.D.?

JW: Well, I was very academically inclined. I don't know at what point I decided to actually do a Ph.D., but it was pretty early on as an undergraduate. The choice was partly dictated by who would give me money to go.

FB: Was there a well-defined alternative to being an academic?

JW: Not really. I hadn't really thought about it. When it came to deciding to go to grad school, my two best choices were going to Northwestern or to Penn. I was admitted with some support at other places. I wrote an honors paper as an undergraduate with Terry Wales. Terry had been on the faculty at Penn and was able to give me lots of good advice. He was influential in getting me support that didn't have any service requirements. My initial offer was to be a research assistant. I think that was the final deciding factor. It was a good choice.

FB: In terms of your family background, was this an expected path for somebody having your background—becoming an academic—or was it a surprise to your parents that you chose this path?

JW: I don't think it was a surprise to my parents because I was very bookish and liked math and sciences. None of them went to university. I'm the first generation to do that. My father, who was the youngest of five, was the only one of his siblings that didn't have post-secondary education—he would have, if it hadn't been for the war. His father was a stonemason; I don't even think he went to high school. My parents went to high school, but that was it.

CdA: You completed your thesis in 1977 under Karl Shell's supervision, *Essays in Public Economics*. Were parts of your thesis published?

JW: Yes. I left Penn in '76. There was just a tiny bit of my thesis left to do. I finished it in the first couple of weeks after I left, but I wasn't able to schedule a time when all my examiners could meet until the following spring. That's why the degree is '77.

My thesis had two parts. I initially started working on congestible goods, goods that are intermediate between private and public. I had a way of modeling them as a technological relationship where you have some inputs, and the outputs are how much of this good each person gets. In the case of pure private goods, one unit more for one person is one unit less for another, so it's a kind of sharing technology. If it's a pure public good, there is a fixed maximum consumption for each person possible with given inputs, with free disposal permitted. It's like having a production surface that is a reverse Leontief. There is a ray, and the frontier is kinked at that ray pointing downwards. You could also have convex technologies in between the polar cases.

There were two essays on that. One of them was looking at optimality conditions, welfare optimality. In one polar case, these conditions reduced to the public goods Samuelson rule; at the other extreme, they are the standard conditions for private goods. The second chapter provided a general equilibrium analysis of this problem. I developed a concept of equilibrium which at one extreme is Walrasian and at the other extreme is Lindahl.

At Penn, there was a theory topics course every semester and Karl Shell taught it one term. In that course, there was a mix of topics that were kind of hot at the time that he was interested in. We studied optimal commodity taxation and modeling transaction costs in general equilibrium because Karl was working with Walt Heller on trying to pull these two issues together (Heller and Shell 1974). I started thinking about commodity tax problems as a result of that. The other half of my thesis dealt with tax issues.

There is the famous Diamond–Mirrlees Production Efficiency Theorem (Diamond and Mirrlees 1971). Peter Diamond and Jim Mirrlees had sufficient conditions on properties of the excess demand functions that would guarantee that it was optimal in their environment to have aggregate production efficiency. One of my chapters developed a necessary and sufficient condition for that. It was also related to some work that Frank Hahn (Hahn 1973) had done. The final chapter was basically a note. There was an obscure paper by Diamond and Mirrlees (1976) dealing with shadow pricing, and I extended it to get some partial results when you have a non-convex technology.

The material in the first half of the thesis had been a really active area of research, but it kind of died at the time I finished my thesis. I extracted a short note out of it (Weymark 1979a) and just put the rest of this part aside. The chapter with the necessary and sufficient conditions for the efficiency theorem I published in the *Journal of Economic Theory* (JET) (Weymark 1978a). I didn't do anything with three of chapters initially, but over the years I've ended up publishing two of them. People had heard about my thesis—local public goods and congestible goods came back in vogue. I got invited to a conference and eventually published the chapter on optimality conditions in its proceedings (Weymark 2004). I have still never published the general equilibrium one, although it gets cited sometimes. I published the shadow pricing chapter in a festschrift (Weymark 2005b).

Let me say one more thing about my graduate education. My research career has been very interdisciplinary. This was a period—I think primarily because of the influence of Rawls' *A Theory of Justice* (Rawls 1971), which came out just before I went to grad school—when economists got excited about Rawls' work. There were lots of interdisciplinary courses, and I took advantage of them. I took a course that was actually run out of the law school jointly by Bob Pollak and Bruce Ackerman, who is a legal scholar. We read parts of *A Theory of Justice* and considered other issues related to that, but once a week Bob had an extra session for the economists. We worked through Sen's collective choice book (Sen 1970), which was only a couple of years old at that time. I also went to the philosophy department and took a course in decision theory from Dick Jeffrey. I actually did that before I did the course I just mentioned. That was my first introduction to John Harsanyi and his work on decision theory.

CdA: That's an impressive variety of things that you did. Did you do that by yourself? Did your supervisor suggest that you do these things, or was that your own choice?

JW: We didn't really have a supervisor until we got to the thesis stage. It was my choice. When I went to the philosophy department to take this course on decision theory, Dick Jeffrey was kind of taken aback. He had taught it for years and said, "You're the first economist who's come over and wanted to take this course." He didn't get many students.

There is something we may come back to later. I wrote a term paper in either that course or the other one I mentioned where I tried to make sense of Harsanyi (1955). Harsanyi is very imprecise in formulating his decision theoretic arguments for utilitarianism. There are gaps in his proof. He has an implicit assumption that he uses in the proof which is not in the theorem statement. The first person to actually prove Harsanyi's theorems precisely is a philosopher of science named Zoltan Domotor. I never met him, but he was at Penn too. When we were talking about Harsanyi's work in Dick Jeffrey's course, Dick told me that his colleague had a manuscript (Domotor 1979). It was a handwritten manuscript, and Dick gave it to us. It uses functional analysis, and at the time I could hardly read it. It turned out to be something that was influential later on when I finally understood what was in it. My first attempts at trying to make sense of Harsanyi's work as a graduate student weren't completely successful, honestly.

FB: Did your supervisor, Karl Shell, in the end have only a minor role, in the sense that he just came in at the very end? How was your relation with him, and how were your relations with other colleagues you met at the time?

JW: The normal way that the graduate programs in the US worked, and still work now, is that you spend a couple of years doing course work first. Karl was on leave my first year, so I didn't meet him until the second year, but there was a personal connection. Karl and David Donaldson were housemates in graduate school and very good friends, and it's actually because of David that Karl met his wife. So there were these close personal connections.

I got to know Karl when he came back. That was the semester that he did this course on optimal commodity taxation, transaction costs in general equilibrium, and non-convexities in general equilibrium. It was a mix of topics. But faculty weren't giving advice on what courses to take. I'd taken a course on public economics with Bob Inman. Bob had devoted part of his thesis (Inman 1971) to issues related to impure public goods, and that's what got me started thinking about that. As a result of Karl's course, I'd started thinking about commodity taxes, so he was a natural person to want to work with. Bob Inman was on my committee as well. I had a two-person committee and they were both very accessible.

FB: Let's move on. The next step was that you went as an Assistant Professor to Duke University. What is your memory of that time? There are various aspects that are interesting. What did you work on, but also, how did you handle the pressure of being an Assistant Professor and having to qualify for tenure?

JW: I actually didn't feel pressure. I was probably being very naïve because Duke at the time actually had a terrible reputation. They had gone years without giving tenure to anybody. I just did what I could. When I went out on the market, my thesis—although it was theoretical—was all public economics, and that's how I presented myself. And that's essentially what Duke hired me for.

But I was also attracted by the fact that there was an active group of people working in micro theory, including Roy Weintraub and Dan Graham. Wes Magat in the business school was working on mechanism design issues. Dan was away the first year, and he normally taught the first graduate micro course. They needed someone to teach that, so I got put into it. Then the next year he came back and took his course back, but Roy was on leave, so I taught his course. That was the whole graduate micro sequence. In retrospect, this was one of the best things that could have happened to me because you just learn material so much better by teaching it. I think a lot of the precision that's in my papers came from having to be careful teaching. I never ended up actually teaching public economics when I was at Duke.

We had a nice reading group that first year. This was a period where people were trying to integrate money into general equilibrium models, people like Jean-Michel Grandmont, Jean-Pascal Benassy, Yves Younès, and so on. Roy Weintraub ended up writing a book on micro-foundations of macro based on this literature (Weintraub

1979), and then he drifted off into the history of modern economic thought. Dan worked on other topics, so there was nobody working on theory anymore.

I was on the faculty at Duke for five years, but I only actually spent three of them there. At the end of my second year, I had an opportunity to go to UBC as a visitor for a year, which I did. One of the things I worked on in that period was tax reform problems. I was looking at a particular kind of dynamic process, and it was very close to some work that Henry Tulkens at CORE was doing in a different context (Tulkens and Zamir 1979). I sent him the paper (Weymark 1981b) and he liked it, and he asked me whether I would consider coming to CORE for a postdoc. The people at Duke were very gracious. They let me have a second year's leave to do that, and then I went back to Duke. But I got spoiled because UBC was so active in theory, and I developed good working relationships with David Donaldson and Chuck Blackorby during that year. CORE was also active. I had already started working on social choice and interpersonal comparisons before going to CORE. Of course, there is Claude d'Aspremont's paper with Louis Gevers (d'Aspremont and Gevers 1977)—it was a great joy to go to CORE and have them as colleagues.

To answer the other part of your question on what I was working on—well, when I went to Duke, I quickly realized that the research on congestible public goods wasn't going to go very far at that stage, given how the direction of the field had changed. So, at the time I only drew a short note out of it (Weymark 1979a) and sent the paper related to the Diamond–Mirrlees efficiency theorem to JET (Weymark 1978a). The condition introduced in it ended up being used a lot in international trade theory for studying tariff policy. It's actually called the Weymark Condition, although it is picked up from a later paper of mine (Weymark 1979b), not that one.

It was at that point that I started working on inequality measurement. Sen's lectures, *On Economic Inequality* (Sen 1973), came out the year I was taking the course with Bob Pollak and Bruce Ackerman. It was available in England before it was published in North America. One of my classmates was from England, so when he went home for Christmas, he had an order from half a dozen of us to bring the book back. We read it, and it was very influential. This was before I went back to UBC as a visitor, but I was reading the research that Blackorby and Donaldson were doing. There is some nice geometry underlying Gini indices, and they generalized it to look at the normative foundations for different inequality indices (Blackorby and Donaldson 1978). I was very influenced by that. I also started learning a bit about formal measurement theory, particularly the book by David Krantz, Duncan Luce, Patrick Suppes, and Amos Tversky, *Foundations of Measurement*—the first volume (Krantz et al. 1971), which was the only one out at that time. I started seeing that the kind of structures they were looking at were what you needed for an axiomatic analysis of inequality measurement.

CdA: Let me interrupt you. When you came to CORE in '79, Jean Gabszewicz was Research Director, and we worked with Alexis Jacquemin on this cartel stability paper (d'Aspremont et al. 1983). This was an IO paper, although the concept is larger than just for IO applications. You had another paper that was said to be in IO, on concentration indices, with Chuck and David. How do you view this work now in retrospect?

JW: OK, let me answer this question first. That paper actually grew out of a project that I started in those first two years at Duke, when I started thinking about axiomatic approaches to inequality and their welfare foundations. There are many formulas for the Gini index. In the JET article that was published in the late seventies by Blackorby and Donaldson (1978), when they worked out the welfare function underlying it, it turned out that a very convenient representation was the following. We rank order incomes from highest to lowest in a discrete distribution, where the weights are one on the highest income, three on the next, five on the next, and so on. Then to normalize, we divide by the number of people squared because then the weights add up to one—the sum of the first n odd numbers is n^2. And I thought that this nice linear structure would give you some nice properties. I generalized this structure by instead of giving a weight of one to the highest income, three to the next highest income, and so on as you go down the income distribution, by just having the weights decrease monotonically, what I call the Generalized Ginis indices. I then asked, "What would be an axiomatic characterization of the Generalized Ginis?" And that's where I was using the foundations of measurement. I had that pretty much worked out before I went to UBC.

CdA: This was your 1981 paper?

JW: Yes, the one that got published in *Mathematical Social Sciences* (Weymark 1981a).

CdA: So, you started working on that before?

JW: Before going to CORE—it wasn't published until afterwards. The paper that went to JET (Weymark 1978a) out of my thesis got me to start working on a paper on tax reform while I was at Duke, which I finished off at UBC. In this paper (Weymark 1979b), I have a reconciliation of recent results on optimal commodity taxation. There were some results that Diewert (1979), Dixit (1979), and Guesnerie (1977) had, among others, and it wasn't clear how they were connected. They seemed to be inconsistent. I worked out what was involved and why they were getting different results. I was doing that at the same time as my work on inequality, but I put priority on getting the tax stuff written up.

When I went to UBC, of course, I had both Blackorby and Donaldson as colleagues. David was actually on leave that year. He was renting a house on an island between Vancouver and Vancouver Island and coming back once a week. I rented his house. The arrangement was that he kept a room in the house and he could stay there. During his trips back I told him about this work that I hadn't written up yet, the axiomatic characterization of Generalized Ginis. I told him that I didn't have the principle of population as one of my axioms, but that it would be nice to see what is the class of indices that satisfy the principle of population in addition to the axioms I had used to characterize the Generalized Ginis. We didn't have a clue what the answer was. I went to visit him on this island for a weekend. We had no resources, just the problem and how to proceed.

Some time before that, I think before I went to Duke, I started learning some basics of functional equations. Blackorby together with Dan Primont and Bob Russell had written a book on duality, separability, and functional structure (Blackorby et al. 1978), and they used these tools a lot. They looked pretty interesting to me. Dan Primont had directed to me to the work of János Aczél on functional equations (Aczél 1966). It was pretty simple to pick up the basics. It turned out this was the tool we needed.

After doing a bunch of manipulations, we ended up with a functional equation that is an example of what's called a Cauchy Equation. It's the equation on the real line: $f(xy) = f(x)f(y)$. But the twist with our problem was to solve it when the domain is the non-negative integers. We knew that $f(x) = x^r$ is the solution on the non-negative line provided that you have some continuity or monotonicity. Initially, we didn't have a clue what to do with this integer domain.

We realized—David was really the one that had the key insight—that you need to go back to Aczél's classic book (Aczél 1966) and look at the solutions for the various Cauchy equations like $f(x + y) = f(x) + f(y)$ that don't impose a regularity condition like monotonicity or continuity. You can build up the real line from what's called a Hamel basis. It's like spanning the space. You get to specify what the value of the function is on each element of this basis, and everything follows from that. There's a unique way to express the numbers in terms of this Hamel basis, and David realized that something was going to be analogous on the integers; it's the fact that you can factor any integer into a product of primes. And so without any regularity conditions, we could specify what the value of the function is on each prime number.

But our problem actually had some regularity and continuity. Once you put that in, it just becomes the same solution as the case with the non-negative integers. When we undid all the transformations, we obtained the class of what we called the Single-Parameter Ginis. For one parameter value, which in this case is two, we have the regular Gini. Our class provides a way of moving between having no concern for inequality in the utilitarian case up to maxi-min. We had no clue that that was going to be the answer when we started working on this problem.

We obtained all the results for that paper in that weekend. I've never been so productive in my life. We wrote the paper up within the next two or three weeks. It helped that David wasn't teaching. It was one of the best experiences in journal submission I've ever had—the paper (Donaldson and Weymark 1980) was accepted with no revisions within a few weeks. We had another result though, which we added in. That's never happened since then. So, it was a very gratifying experience.

To connect back to Claude's question about IO—there were a lot of indices floating around measuring concentration, and a few people had tried to look at welfare foundations for them. There's something called the "Numbers Equivalent" way of measuring concentration that sounds a lot like the equally distributed income in inequality measurement. That's what led to this paper on concentration indices (Blackorby et al. 1982). At the time, the *European Economic Review*—this was when it was still associated with the European Economic Association—had a special issue on topics related to IO once a year, of which Claude was one of the guest editors. This material was finished up while I was at CORE. Because of having Claude as a colleague and co-author, we thought that was the place to send this paper—that's how it ended up there.

The other paper that Claude mentioned was on cartel stability (d'Aspremont et al. 1983). CORE had this very interesting system for discussion papers that I quite liked. Before a paper could go into the discussion paper series, it had to be refereed internally. One of the ways that CORE tried to integrate newcomers, visitors, and postdocs was to give them one of these papers to referee and comment on soon after they arrived. I'd never worked in industrial organization—never taken a course in it—before being asked to review a cartel stability paper that Claude had written together with Jean Gabszewicz and Alexis Jacquemin. They used what's called the price leadership model: there's a dominant cartel that sets the price and the other firms behave competitively relative to it. They were working with a continuum of firms and formalized some intuition of George Stigler's that cartels wouldn't be stable (Stigler 1964), that there were always incentives for the cartel to break up. It was a fairly short note, so it didn't take me long to review it.

I was intrigued and I thought that a lot is special because of the fact that they were working in a continuum—what would happen if you worked with a finite number of firms? I don't remember how explicit the concept of stability was in the working paper that I had reviewed, but when you went to the discrete case you had to be really precise. And the relevant concept of stability, I thought, was that the cartel was stable if someone leaving the cartel is worse off as a result when they take into account the effect that the departure is going to have on the price and, vice versa, nobody wants to join the cartel if after they do the prices change in a way that is adverse. It didn't take long to figure out that with that concept of stability in the finite case, things don't unravel like they do in the continuum—there always is a stable cartel. In the end, we combined our research, and that became the article. Remarkably, in retrospect, nobody wanted to publish it—we had it turned down two or three times.

After that, I went back to Duke for a year. Then I got hired by UBC on a permanent basis. The article was still floating around then. One of my colleagues was editing the *Canadian Journal of Economics* (CJE), so we sent it there and they accepted it. But they didn't want the appendix. We had a nice example in the appendix. Eventually, it got incorporated in another paper by Claude and Jean (d'Aspremont and Gabszewicz 1986). I'm really proud of the work that Claude, Alexis, Jean, and I did.

That joint paper and the Generalized Gini paper of mine are my two most highly cited papers. Both of them ended up being picked up by other literatures. The cartel stability paper gets cited so much because environmental economists think about it. They study how to make environmental agreements between countries, and they need some concept of stability. They use the concept of stability that was introduced in that article. In retrospect, it would have been better to exposit our paper in terms of a multi-stage game, which we didn't do.

At some point, Peter Wakker came across my Generalized Gini paper and realized that it was formally equivalent to what John Quiggin had done on decision-making under uncertainty (Quiggin 1982). At that time, Quiggin called it "anticipated utility." It later became known as "rank-dependent expected utility." I didn't use these terms, but I had an independence axiom that is analogous to the co-monotonic independence axiom in decision theory, which played a key role in my axiomatization. There are a lot of formal similarities between inequality measurement and decision-making under uncertainty. If you think about the work by Tony Atkinson on measuring inequality and stochastic dominance (Atkinson 1970), it is related to Michael Rothschild and Joseph Stiglitz's paper on the comparative analysis of risk (Rothschild and Stiglitz 1970). They both appeal to the same stochastic dominance results. So largely because of Peter, who prepares annotated lists of papers relevant for decision theory and circulates them, people learned about my Generalized Gini paper. Then it started getting cited and read a lot.

CdA: Later on, you worked on multidimensional inequality, then you came back to the Gini indices. What is fascinating with the Gini index is that the social evaluation function can lead either to relative or to absolute indices.

JW: That's right. Because the iso-welfare curves are linear within a rank-ordered set, you can easily move between the relative and absolute perspectives. One thing that has characterized my research career is that I cycle back to topics. I'll work on something for a while and think, "OK, I don't have any more good ideas about that so I'll put it aside and work on something else." Then a few years later, I might come across something that sparks my interest. For example, when research on multidimensional inequality started taking off, I came back to inequality measurement (Gajdos and Weymark 2005).

CdA: Now let's switch to something else. You insist on the fact that inequality indices should be normatively founded. In some sense, the idea of introducing interpersonal utility comparisons seems to be a natural follow-up of this assumption. So you went in this direction. First, there was your diagrammatic introduction (Blackorby et al. 1984), which was quite successful, I would say. There was also a publication in *Scientific American* that used it (Blair and Pollak 1983b).

JW: Yes.

CdA: Then you went on with your paper on the proof of Arrow's Theorem, (Blackorby et al. 1990) and your survey chapter with Walter Bossert (Bossert and Weymark 2004). Recently, you also had a paper with Michael Morreau on welfarism, introducing different individual scales to measure well-being (Morreau and Weymark 2016). So, this has been a long line of research.

JW: It actually ties in with my work on Harsanyi as well. Trying to sort out what was going on with interpersonal utility comparisons in his work really goes back to my grad school days. But it was when I worked with Blackorby and Donaldson on diagrammatic social choice that I started publishing on this issue. In this work, we were interested in employing a social welfare functional approach to characterize different kinds of social objective functions in order to move away from the nihilism of Arrow's ordinal non-comparable framework. There was a period in the latter part of the seventies when Blackorby and Donaldson were amazingly prolific in generating a lot of great papers. They have a paper in the *Journal of Public Economics* (Blackorby and Donaldson 1977) where they are looking at quasi-orderings in a social context, and that led them to start thinking about this literature on social choice and interpersonal comparisons that was using Sen's framework (Sen 1974). This was the period when this work was done. What year did your paper with Louis Gevers come out?

CdA: It came out in '77. I started working on it in '75, I would say.

JW: In '77 the paper (d'Aspremont and Gevers 1977) came out, so Chuck and David didn't know it before then. But that and work by Peter Hammond (Hammond 1976) and Amartya Sen (Sen 1974) were really some of the most significant work being done in social choice theory, and very congenial to the kinds of interests we all had. David and Chuck had written one paper (Blackorby and Donaldson 1977) which is closely related to the Suppes grading principle (Suppes 1966); it involved ordinal comparisons, but they could be incomplete. David gave a seminar based on some of the work he and Chuck were doing. When he talked about social welfare functionals with ordinal level comparability, he provided a little diagrammatic illustration in the two-person case to provide some intuition. After the seminar, we started talking about it, and I suggested to David that this idea must be applicable to different assumptions about measurability and comparability of utility. So together with Chuck, we sat down and worked it out. This was towards the end of my time as a visitor at UBC, before I went to CORE. We actually finished and wrote our paper (Blackorby et al. 1984) while I was at CORE. Chuck came for a brief visit to CORE. He was at Nuffield College for part of that year.

I can't overstate how important the year at CORE was for my career. At CORE, I got to explain my research and get feedback from Claude and Louis Gevers. It was the first time I'd ever been to Europe. Both with this kind of work and the work I was doing on taxation theory, Europe was where all the action was. As a graduate student, I read a lot of the work that Roger Guesnerie was doing. It was very influential, but his work is very tough, very technical for what people were doing in that field. I have very geometric insights, and there is a lot of geometry underneath Roger's work on taxation theory. So, I invested a lot of time in learning it.

CdA: You met Roger Guesnerie during that period?

JW: Yes. I had been out of grad school for three years before going to CORE and I was just starting to get publications. Through the connections of people who were very kind to me in arranging introductions, I received a number of seminar invitations that year that proved to be very important. One was a short trip to England, where I gave a talk at Warwick. At the time Avinash Dixit was one of the faculty there who was working on optimal tax, as were Nick Stern and Jesús Seade. The next stop was Nuffield College and Jim Mirrlees' famous seminar—he invited me. That was when Amartya Sen was at Nuffield. It was the first time I met him. Amartya, then as now, is a very, very busy person, but he is one of those scholars who is very encouraging and supportive of younger people, as is Jim Mirrlees. I had a fifteen or twenty minute appointment to speak with Amartya, and I went to explain to him the diagrammatic social choice paper. He ended up giving me forty-five minutes to an hour and has been supportive ever since. Mirrlees was the same way. He was tremendously supportive.

Then later in that year, I got the opportunity to go to Paris at Roger Guesnerie's invitation to give a talk. That's the first time I got to know him. I presented some work I was doing on taxation. When we went to dinner, it turned out that Nick Stern was in town, so it was the three of us. There's a famous symposium in the *Journal of Public Economics* on optimal income tax with discrete types. Nick has a paper (Stern 1982), and Roger has a paper with Jesús Seade (Guesnerie and Seade 1982) which has been very influential. I was hearing about this work in progress during the dinner. The following year I was back at Duke and Roger was visiting at Penn that year, so I arranged for him to come and give a talk, and he gave that paper. I studied it very carefully, and I've used it a lot.

There's one other seminar I gave that year which turned out to be very influential, and that was Martin Hellwig's seminar in Bonn. Martin was a professor at Bonn at the time. That was the first time I met Martin, whose work in taxation has been very influential to me. Much later, during Martin's time at the Max Planck Institute in Bonn, I visited a number of times.

FB: Had he been working on taxation already at that time?

JW: No.

FB: The earliest piece I know is from '86 (Hellwig 1986), I think.

JW: Initially, it was just the case that he was kind enough to invite me, which is how I got to know him. He's been supportive over the years. I've been fortunate that there were many senior people in the profession who were very kind and supportive when I was starting up, and that has had a big influence on how I deal with people. I try and be helpful and supportive with junior scholars in the profession.

FB: I have maybe a more personal question. When I was preparing this interview, I asked a couple of colleagues what they would ask you. One question I got from François Maniquet was whether you are a welfarist. Maybe, we'll just start by defining what it is—how would you define a welfarist? Then it comes time to confess whether you are one or not.

JW: Well, welfarism is a form of consequentialism. A consequentialist's view is that if you're deciding how to rank different actions or different options, you only pay attention to the consequences. You don't care about the procedure, you don't think about rights, and so on. Welfarism is a particular form of consequentialism in which all you care about are welfare consequences, that is, the utilities that result. If you're ranking different alternatives, the only thing we care about are the utilities that are generated by those alternatives, not whether they respect people's rights or anything else like that.

I'm not a welfarist, but I find it very useful to understand its foundations. I've given you a general definition. In the context of social choice and welfare economics, you want to know what in the nature of social rankings and social welfare functions makes them welfarist. Certainly in the literature we talked about earlier on social choice and interpersonal comparisons, there were characterizations of welfarism in a multi-profile context. When Amartya Sen got an honorary degree from the Université catholique de Louvain, they had a special issue of the *Recherches Économiques de Louvain* in his honor, and Blackorby, Donaldson and I published a paper in it (Blackorby et al. 1990). The main part of the paper built on the diagrammatic paper (Blackorby et al. 1984) that we talked about earlier, which just used two people to show the basic geometry. We used that geometry to build an n-person, geometrically based proof of Arrow's Theorem. But there was actually a bit more to it than that. We showed exactly what was needed if you have just a single profile of preferences, what this kind of welfarism requires. It turns out that it is equivalent to Pareto indifference.

I've realized that I didn't answer something Claude asked about the *Scientific American* piece (Blair and Pollak 1983b). As I mentioned earlier, Bob Pollak taught me social choice as a graduate student. He also worked on social choice theory. Bob was editor of the *International Economic Review* (IER). We sent the diagrammatic paper someplace else and they weren't interested; we didn't know what to do next. Pollak said, "Send it to the IER." He was very sympathetic to our paper. He really liked it, so he studied it carefully.

Doug Blair became a faculty member at Penn just after I left, so I didn't know him until a little while later. Bob and Doug started doing a lot of research on social choice theory in which you relax the collective rationality requirement and don't demand social orderings. They were particularly interested in looking at acyclic choice (Blair and Pollak 1982). When you look at acyclic social choice, the kind of rules that satisfy a set of axioms depends very much on the number of people versus the number of alternatives. It becomes very difficult technically to get complete solutions. They worked on this problem and had to develop some new theorems on colorings and graphs to solve it (Blair and Pollak 1983a). In the process, they realized that our diagrammatic approach could give some insights to more than just the case where you have social orderings. As a result, they decided to write up a piece for a broad audience, which was published in *Scientific American* (Blair and Pollak 1983b). So, our diagrammatic approach went public.

In fact it was very nice that they did this because their article came out when I was a permanent faculty member at UBC. One of the nice features of UBC at that time—it's since disappeared—was a faculty club where people go to lunch. They had a cafeteria and you would see people from other departments. I had taken a course in differential equations as an undergraduate from Z. A. Melzak. He was a wonderful person and a great teacher. And he'd seen it. He came in one day and said, "I saw your work being discussed." He was proud that one of his students had his work discussed in *Scientific American*. That piece was probably the one thing that disseminated my research more than anything else.

CdA: You said that your interest in interpersonal comparisons of utility came out of your reading of Harsanyi. You said also that you have pursued subjects, year after year, abandoning and coming back to them. One area where you have been very active is the famous question of John Harsanyi's defense of utilitarianism (Harsanyi 1955) and Sen's criticism (Sen 1976). What is your opinion on this now? You mentioned that measurement theory was already important also in your seminal paper from

1981 (Weymark 1981a). On Harsanyi, you also have a recent piece which introduces measurement theory (Weymark 2005a). So what is your opinion now on this debate?

JW: Well . . .

CdA: Now, or you can say what was your opinion in the past also. [All laugh.]

JW: It hasn't changed a whole lot. As I said, I tried to make sense of Harsanyi's statements. His aggregation theorem, roughly speaking, says that given a set of lotteries to be socially ranked, if individuals have preferences over them that satisfy the expected utility axioms, the social preferences also satisfy these axioms, and the social preferences are connected to the individual preferences by a Pareto principle like Pareto indifference, then if these preferences are represented by von Neumann–Morgenstern utility functions (von Neumann and Morgenstern 1944), the social utility function is an affine combination of the individual von Neumann–Morgenstern functions. So, basically, social welfare has a weighted utilitarian form.

With Harsanyi's impartial observer theorem—using Rawlsian terminology—you imagine yourself behind a veil of ignorance in which there is an equal chance of you being anybody. This is a problem of choice under uncertainty. Harsanyi (1955) argues that if you think that you have this equal chance, you're going to end up with utilitarianism. But the arguments are all pretty loose.

In Harsanyi's early work, like the'55 paper, he also considers a different argument in support of utilitarianism due to Marcus Fleming (Fleming 1952) that is based on the separability of a social ordering over vectors of utilities. So, Harsanyi has what he thought were three different decision-theoretic arguments for supporting utilitarianism, or at least weighted utilitarianism.

The year I was at UBC as a visitor, I got Blackorby and Donaldson interested in looking at this problem and we wrote a working paper. It first came out at UBC. Then Claude was our referee when I wanted to put it out as a CORE discussion paper (Blackorby et al. 1980) when I was I postdoc there. The point of that paper was to argue that the social rankings sort of look utilitarianan, but they really aren't. Moreover, in each of these three models, you might get a different ordering of the alternatives, so how can they all support utilitarianism?

It wasn't a very well written paper. We used a lot of functional equations—Pexider equations. We asserted something without actually proving it. It turned out that the relevant mathematical theorem from the functional equations literature for the domain we were using wasn't yet known. I think that in terms of the way we presented our results, we made a mistake. Instead of using the lottery model that Harsanyi had used for uncertainty, we used a state contingent alternatives model of uncertainty, but with objective probabilities. It was a model that Arrow used in the early 1950s (Arrow 1964). We did that because Blackorby and Donaldson (Blackorby et al. 1977) were working with that model.

We sent our paper to *Econometrica*. Harsanyi was obviously one of the referees, and he clearly didn't understand what we did. We had it rejected, and we talked about how to proceed. Bob Pollak was the editor of the IER, and we talked to him. He said: "Send it to me, but revise it, take the reviews into account." But we actually never did that. It's something I regret.

CdA: It was never published?

JW: It was never published. There were parts of it that were published (Blackorby et al. 1999). What we were focusing on at that time was that these three different arguments didn't actually produce the same social ordering. How in the world can it be an argument for utilitarianism when you're getting different orders? We had pretty much, but not precisely, identified what the real issue was. The argument that was used to show this was related to the expected utility representation theorem. It didn't really matter if we used state contingent alternatives or lotteries, as von Neumann and Morgenstern had done. If you use lotteries, the axioms are about binary relations over a set of lotteries: there are ordering, continuity and independence axioms. The representation theorem says that we can represent these preferences with a von Neumann–Morgenstern utility function, so it's linear in the probabilities. It doesn't say that you have to—any increasing transform is a perfectly acceptable representation. I think this is something that Harsanyi never recognized—the difference between "can represent" and "must represent."

When you get outside of normative contexts like welfare economics and social choice, it doesn't matter which representation you use. You just pick one for convenience; the von Neumann–Morgentern one is a nice representation to work with. But when, in the normative context, it is to be used somehow as the basis for interpersonal comparisons, the choice of representation matters a lot.

The basic idea is that you can take nonlinear transforms of the utilities and still have a representation of the preferences. Essentially, in the '80 paper (Blackorby et al. 1980), we had specified exogenously that there were welfare relevant measures of utility, and that what each of these three different approaches was doing was to take different nonlinear transforms of them before summing across individuals. That is why you were getting different outcomes: with the impartial observer model, with Harsanyi's aggregation theorem, and with Fleming's separability approach.

In retrospect, we realized that we had the first proof of a version of Harsanyi's aggregation theorem for a different model of uncertainty. Eventually, we cleaned that result up. As I said, there was a little gap. We had asserted what the solution to a functional equation is, but on a non-standard domain. At some point, Walter Bossert read our paper, and he knew of a paper by Radó and Baker (Rádo and Baker 1987) that had come out in the mid-eighties, a few years after we wrote this draft, that actually solved this Pexider equation on the relevant domain. So, we could appeal to it. Eventually in the '90s, we pulled that part out and it became a *Journal of Mathematical Economics* paper (Blackorby et al. 1999) proving Harsanyi's aggregation theorem in that framework.

But, in the '80 paper, we didn't discuss at all the debate between Sen (1976) and Harsanyi (1975) on whether Harsanyi had actually proven anything about utilitarianism. Amartya clearly understood that there were issues related to the choice of representation, and he has a very important paper in *Theory and Decision* (Sen 1976) where he makes this point. Amartya is usually one of the most clear writers that you encounter, but that paper is not clear. He expressed his argument in the form of a Socratic dialogue. Unless you have actually figured it out on your own, you don't understand it. There was a back and forth between him and Harsanyi (Harsanyi 1977; Sen 1977), and it was clear that Harsanyi never saw the point of what Amartya did. I realized that the paper I had done with Chuck and David was what you needed to really address this.

There was a conference in Berlin in the mid- to latter part of the '80s where I made a first stab at sorting this issue out, but I hadn't actually written a paper yet. That eventually became the "Reconsideration of the Harsanyi–Sen Debate" paper which was published in the Elster–Roemer volume (Weymark 1991). John Roemer and Jon Elster held a conference on interpersonal comparisons of utility. At the time, John Roemer was at Davis, where they held their conference. Chuck and David were invited to present a paper, and I was invited to be a discussant. What John and Jon neglected to tell me was that this was the first of a two-part conference. There was going to be a conference at the University of Chicago where Elster was at the time; I think it was one year later. The people who were discussants the first time would give papers the second time. So, I was supposed to produce a paper. I had been making an attempt at working on the Harsanyi–Sen debate after the Berlin conference, so I thought I would write this work up because it was now starting to come together. I asked Chuck and David to work on it, the three of us, because it was really building on what we had been doing. They said they already had a paper for this volume, so I should just do it.

I regret that they are not co-authors on this paper. The one thing thing that I don't think is clear enough in my paper in the Elster–Roemer volume is how much is based on our joint research. I do say that it is heavily influenced by my work with Chuck and David, but that's an understatement. The key idea that you can take different nonlinear transforms of the welfare relevant utilities is in our unpublished paper (Blackorby et al. 1980). And I'm applying that idea, and specifically looking at the Harsanyi–Sen debate in the first part of my paper. But I had a tight deadline to finish the conference paper. It's a long paper, and I really should have split it into two. The first part was to trying to formally explain what Harsanyi actually has shown, and to show that you really shouldn't be interpreting this as theorems about utilitarianism, which is what Sen's point was (Sen 1976). Sen said that Harsanyi just had representation theorems which are not really about utilitarianism because the utilities you should use should be the welfare relevant ones, and there's no reason to believe that they are in Harsanyi's theorems.

Let me go back to Harsanyi's '55 paper (Harsanyi 1955), which is where he has both the aggregation theorem and an impartial observer theorem. In his '53 paper (Harsanyi 1953a), Harsanyi has the basic idea of the impartial observer theorem, but it is for a very special case. He also talks about Fleming's paper (Fleming 1952). But in the latter part of the'55 paper, he talks about the basis for making interpersonal comparisons. Essentially, what he is saying is that there is some kind of master function for making them. In his terminology, the satisfaction you get from an alternative is a function of all the variables—he calls them causal variables—that affect how much satisfaction you get from the alternatives, with the same function used for everybody. If you know the value of these causal variables, and you apply it for one person, then for another person, you can use this master function to make interpersonal comparisons. It builds on an earlier paper that is not well known, "Welfare Economics of Variable Tastes" (Harsanyi 1953b), in which Harsanyi is trying to do welfare economics when someone's tastes change. He had the idea that because it's the same person, you can put yourself in the place of your other self and make a comparison. Underlying some of this research is also the whole idea of empathetic preferences; it's all mixed in.

In the latter part of Harsanyi's '55 paper, he goes beyond the kind of ordinality you have with the von Neumann–Morgenstern axioms. The latter part of my Elster–Roemer paper was trying to build a case for a utilitarian theorem using the kind of analysis that Harsanyi employed, but without restricting attention just to the preferences. Rather, I used the welfare relevant utility functions, which might not be perfectly measurable. But I was rushed because of the deadline, so that part of my paper is not developed well enough. Most people don't even read that part, I think. It didn't get picked up on for a long time. I think I should have put it aside and done it later.

John Broome has been very critical of Harsanyi's work that uses this kind of basis for making interpersonal comparisons (Broome 1993). He is also critical of the later related work of Serge Kolm (Kolm 1998). I think that Broome's criticism of Harsanyi is off the mark. As far as Kolm is concerned, I think that he's probably right. As I said, Harsanyi was talking about some cardinal measure of well-being—he called it satisfaction. Broome says if we use this causal variable kind of approach to talk about interpersonal comparisons of preferences, it is problematic. I think that his arguments are right. But once you talk about interpersonal comparisons of satisfactions in a cardinal framework, they don't apply. Kolm used an ordinal framework. I've never really followed that up very much, but I think there's something there.

As I said, when I wrote that paper, I was very close to the deadline. The people who were at the second conference didn't have a whole lot of time to prepare their papers, plus mine was long. I sent it to Harsanyi. I had not met him at that time. He wrote back a very nice letter. One of the things that I found very interesting was that he has an argument in his letter justifying interpersonal comparisons that he never put in any of his published work. He builds up the interpersonal comparisons by comparing the gain going from a to b versus c to d—difference comparisons. You can build up a cardinal theory from that, but it's not in his work. I think that the most successful material that he actually wrote on interpersonal comparisons was the last part of his'55 *Journal of Political Economy* (JPE) article (Harsanyi 1955), where he has his causal variable approach.

CdA: That's going in the direction of using measurement theory.

JW: That's right. You asked about my other article on measurement theory. Once I wrote my paper on Harsanyi for the Elster–Roemer volume (Weymark 1991), it was hard to not keep writing about his ideas. I think my next most important paper about Harsanyi is "Measurement Theory and the Foundations of Utilitarianism" (Weymark 2005a), which is where I started looking at the history of expected utility theory and its use in welfare economics. There was a conference in honor of Arrow in Caen that Maurice Salles organized on issues related to the history of social choice and welfare economics. By that time, Harsanyi's early papers were 50 years old. In part because of work that had been critical of my paper in the Elster and Roemer volume by John Broome (Broome 2008) and by Mattias Risse (Risse 2002), I had gone back to looking at von Neumann and Morgenstern's book (1944), seeing how it connected better with normative applications.

One of the arguments of Broome and Risse was that you can go beyond what Harsanyi did with expected utility theory. Broome appeals to "naturalness" when choosing a utility representation. Risse says that there is more to expected utility theory than what I had talked about. If you go back to von Neumann and Morgenstern, they do have more, and it really ties in with the work on measurement theory that I'd done elsewhere. The way we normally present expected utility theory for lotteries, as I've indicated, is that you have binary relations over the set of lotteries, and you also have three axioms. (Things are somewhat more complicated if you allow for compound lotteries.) When von Neumann and Morgenstern develop their expected utility representation, they say that they are also actually considering an operator, and they are making a novel contribution to measurement theory in doing so. The operator is taking convex combinations of lotteries. This is what Risse has in mind.

In what's called the representational theory of measurement—what Krantz, Luce, Suppes, and Tversky's major masterpiece volume (Krantz et al. 1971) is all about—you start with an extensive relational structure. For example, if you're thinking about length, you have an empirical relational structure. Think about having a bunch of rods. How are we going to measure their lengths? You can put two rods side by side, and if one rod goes farther than another, you say that this rod is at least as long as that rod; so we have a binary relation. But we can also have an operator called a concatenation operator, where you put two rods end to end and compare that combined object to the other rod. Then we move into what's called a numerical relational structure, where we want numbers to summarize the empirical relations. It's generally natural, it would seem, that the way you would translate measurements into a numerical structure is that instead of objects you use the numbers that you assign to objects. You translate "at least as long as" into "bigger than or equal" for the numbers, and your concatenation operator is going to be "plus." So, when two concatenated objects are the same length as a third object, in the representation, the length number you assign to the long object is the sum of the other two lengths.

In von Neumann–Morgenstern expected utility theory, you have an infinity of operators, one for each possible weight used to form a convex combination, and you ask that this operation be preserved in the representation. Back in the space of lotteries, the convex combination is, for example: $[\alpha \cdot P] + [(1-\alpha) \cdot Q]$, where P and Q are lotteries. Then in the representation, you want the utility of this combination to be a weighted sum of the utilities assigned to lotteries P and Q, with weights α and $(1-\alpha)$, respectively. This part of von Neumann and Morgenstern's theory got lost. Harsanyi based his work on Jacob Marschak's exposition of von Neumann and Morgenstern theory (Marschak 1950), which is much easier to understand than what was said in von Neumann and Morgenstern's book. Marschak doesn't use convex combination operators; he just works with the binary preference relations. I think that because of this history of thought reason, Harsanyi didn't base his work on von Neumann and Morgenstern, but used Marschak instead.

So, is it possible to resuscitate Harsanyi's conclusion that he really has utilitarianism if account is also taken of these operators? The argument in the measurement theory paper is, no. I have said that we have an empirical relational structure—which here is a set of lotteries, a binary relation on it, and these convexifying operations. We want to translate it into some numerical structure, and we are going to assign utility numbers to lotteries in the order of preference in such a way as to preserve the convexifying operation.

But we do that just because it's simple to work with this kind of number system. You do not have to use this numerical relational structure. Consider the analogous issue when measuring length. Instead of putting two rods end to end, you could put them at right angles and measure the length of the whole thing by the length of the hypotenuse. You have a degree of freedom. So, there is not a unique numerical structure that you can associate with the underlying empirical measurement procedure. This is completely analogous to the fact that you do not have to restrict yourself to the von Neumann–Morgenstern representations when you do not take into account the convexifying operation. There is a degree of freedom. This is recognized in the Krantz, Luce, Suppes, and Tversky book (Krantz et al. 1971). They do not have a big discussion of it, but it is quite early on in the book. I was actually a little bit nervous presenting that paper because Pat Suppes was in the audience, and he is one of the founders of the field and the guru. I thought, "I sure hope I got this right!" And he said, "Yes, it's right!"

CdA: When did you start working on strategy-proof social choice and single-peaked preferences, which is a very important topic in your work with Michel Le Breton?

JW: My work on strategy-proofness started with Michel. Maybe, I should say how our collaboration started. I first met Michel at the World Congress of the Econometric Society at MIT. That was in 1985. Michel's first papers looked at Arrovian social choice and the consistency of the Arrow axioms when you start putting economic structure on the set of alternatives and on the allowable set of preferences. Perhaps, I should just say a few words about that before I go on to strategy-proofness.

CdA: Yes, that's an important topic, too.

JW: There had been some earlier work on that issue, very important work, by Kim Border and Jim Jordan (Border and Jordan 1983) and by Gilbert Laffond. Gilbert Laffond's thesis (Laffond 1980), most of which was never published, related to this. Hervé Moulin's work on single-peaked preferences and strategy-proofness (Moulin 1980) also comes in. There were some important papers looking at structured domains and . . .

CdA: trying to escape the . . .

JW: trying to escape Arrow's impossibility theorem in his framework, not always successfully. Eric Maskin (Maskin 1976) had some early important work on this issue. Norman Schofield was putting together a volume on social choice and political economy. I don't remember if he approached just one of us, or asked the two of us to write a joint paper, or asked us each to prepare a paper. Michel and I had been talking about these issues, and so we decided to write a joint paper (Le Breton and Weymark 1996).

Michel had been primarily working in the Arrow framework where you're ranking alternatives (e.g., Bordes and Le Breton 1989). There is another way to look at Arrow's Theorem from a more choice-theoretic perspective, where you take a profile of preferences and a feasible set, and then choose something from this set. You can reformulate Arrow's Theorem in this framework. David Donaldson and I had been working on this kind of problem (Donaldson and Weymark 1988). I also have an article—without any economic structure—that I wrote with Allan Gibbard and Aanund Hylland (Gibbard et al. 1987). Michel and I had talked a lot about this work.

We wrote an expository piece for the Schofield volume (Le Breton and Weymark 1996) and then we thought we should do something more original together, and we had a couple of ideas. I think the most important one was about strategy-proofness. The proofs of the theorems in the classic papers of Allan Gibbard (Gibbard 1973) and Mark Satterthwaite (Satterthwaite 1975) and the literature that followed on from them were not well suited for starting to include the kind of structure you see in economic models. The reason is that in the proofs you have steps where you take somebody's preference and consider an alternative that is ranked second in the preference. If you are working with economic preferences—for example, if you have two public goods and people have continuous, monotone, convex preferences, or spatial preferences, where there actually is a best alternative—then there's no second best alternative because of the continuity of preferences. So, the proof strategy used for unrestricted domains just doesn't lend itself very well to looking at structured domains.

Then there was a watershed event. It was anticipated by some earlier work of Salvador Barberà (Barberà 1983). The seminal paper using what's called the option set methodology is really the one by Salvador Barberà and Bezalel Peleg in *Social Choice and Welfare* in 1990 (Barberà and Peleg 1990), where they show how to prove the Gibbard–Satterthwaite Theorem on a space of continuous preferences. When you specify the preferences of some subset of the population, an option set is the set of alternatives that are still attainable for some reported preferences of the other people. The way you prove results is by investigating the structure of the option sets. So, if we have a dictatorial rule and I am the dictator, when I report my preference, there's only going to be one alternative in the option set for the rest of the people, my most preferred alternative, because I determine what is chosen. If, on the other hand, I look at any subset of the rest of the population, the option set they generate is the whole set because they have no influence. So, you start to look at what the option sets look like when you consider different structured domains.

After the Barberá and Peleg piece, there was an explosion of activity. At the first meeting of the Society of Social Choice and Welfare in Caen, there were a number of really important papers, some of the early work using their methodology. Michel had arranged for me to come visit him in Marseille, where he was at the time, for a few weeks. I went there to work with him. We had decided that we were going to work on strategy-proofness, looking at a particular restricted domain where you had some separability assumptions. Then in the middle of my visit, we went to Caen for this conference before going back to Marseille to work. We were using the Barberá–Peleg methodology, but it was on a domain where we could not use the kind of constructions that had already been used—we had to come up with some new ideas. That led to my first paper on strategy-proofness (Le Breton and Weymark 1999).

I did quite a bit of reading before I went to work with Michel. It was great because Michel had already started working on strategy-proofness. He had taken some of the thesis research by Gilbert Laffond (Laffond 1980) and worked with him and Georges Bordes to extend it (Bordes et al. 2011). So, Michel knew the literature really well. I was the novice coming into it. We made good progress on what eventually became our main paper on this topic (Le Breton and Weymark 1999).

As a sideline, around that time, Michel was working with Arunava Sen, looking at when, in these separable domains, you can decompose the outcomes into single issue decisions (Le Breton and Sen 1999). I wrote a paper on that as a result of that period of activity (Weymark 1999). It came about as the result of the confluence of two things. First, the fact that we had been working on Arrow-style social choice with restricted domains. Jointly, we had the survey paper in the Schofield volume (Le Breton and Weymark 1996). Separately, Michel had been working on Arrow's Theorem on restricted domains using the social preference approach and I'd been using the social choice theoretic approach. Then, second, all of a sudden, restricted domains became important in the strategy-proofness literature, so we worked on merging the two literatures. Eventually, we were asked to do the *Handbook of Social Choice and Welfare* chapter on Arrovian social choice on restricted domains. We wrote an enormously long chapter (Le Breton and Weymark 2011). It actually came out years after we wrote it. We just missed the first volume and then had to wait seven or eight years for the second volume to come out. It was very fruitful collaboration, and Michel was a great person to work with.

CdA: OK, just a question: recently, you have used tools of social choice to analyze biological problems on measuring group fitness. Do you see this as a promising avenue? Do you see other applications in biology and social choice? I'm curious about that.

JW: On applying social choice ideas in non-traditional areas, more generally, you can think of philosophy of science. With regard to that series of papers, I need to actually step back again to my undergraduate days.

In high school, I'd taken lots of science, but when I went to university, I decided not to pursue science anymore. I did lots of math, but I didn't take science courses. One of the requirements to get my degree as a Bachelor of Arts was to take one or two science courses. I took a computer science programming course. I'd never taken biology as a high school student, and there was a course called something like Genetics and Evolution for Arts students. I thought, "I'll take this, it sounds pretty interesting." I had read some Darwin in the introductory first year course that I'd taken. And the biology course was terrific. Because of my math skills, it wasn't hard for me when we studied Mendelian genetics and needed to figure out how the alleles combine, and their proportions, and all that. Many of my classmates could hardly do anything combinatorial. I thought that the subject was just fascinating. Since then, when I have time, I read popular accounts. The textbook we had was actually a really good book (Lerner 1968). So, I look at some of the biology literature, mostly popular writing.

There's a journal called *Biology and Philosophy*, and there was a social choice article there by Samir Okasha (Okasha 2009), who ended up being a co-author on one of these papers. Samir is a very distinguished philosopher of science who is best known for his work on philosophy and biology.

There's a lot of controversy over the level at which natural selection occurs because not only is it at the gene level, it is at the individual level—there is a biological hierarchy. Organisms are composed of cells. Can there be selection at both levels? So, it's a group selection problem. Samir wrote a book on this subject that won the Lakatos Prize (Okasha 2006), which is a big deal for a philosopher of science. At some point, he started learning individual and social choice theory. He's got a brilliant paper in *Economics and Philosophy* where he looks at cells mixing during meiosis as a "behind the veil of ignorance" argument, tying together Harsanyi with Mendelian genetics (Okasha 2012). In any event, he learned about the social choice theory work of Arrow and Sen.

When you are measuring group fitness, the standard way that it's done by the people who look at this multi-level problem is to think of fitness as having two components. There's what is called viability. Roughly speaking, you might say that this is the probability that whatever organism you are looking at is going to survive to reproductive age. And then if it does, the fecundity part is how many offspring they're going to produce. The standard way of measuring fitness in this formal way is to multiply together a measure of viability and a measure of fecundity.

So, you do this at the cell level. What do you do at the group level? The standard way is just to take an average. And one of the things that people who do this are interested in is what are called evolutionary transitions, understanding what happens when a new level of the biological hierarchy arises—for example, the origins of multi-cellular organisms. If you are going to use this understanding to explain that an evolutionary transition is fitness enhancing, you have to have some more precise notion of fitness than just loose talk.

Samir realized that this was an aggregation problem of the kind studied in social choice theory using Sen's social welfare functionals (Sen 1974). He took the same framework and reinterpreted it in terms of fitness functions. Say we are looking at cells and an organism, where the organism is composed of cells. You're going to have measures of fitness for each of the cells that are actually functions that depend upon the state of the world. You're reinterpreting a utility function as a fitness function. And you are going to aggregate them into the group fitness function. Then, Samir applied his approach to some of the issues that had been considered in the literature. By the way, there is an alternative way other than the average one mentioned earlier of constructing a fitness index.

One of the problems with the averaging approach used to construct a fitness measure is that if you think about the origins of multi-cellular organisms, you get specialization. You get some cells specialized in metabolic activities, which is related to viability, and other cells specialized in reproduction. If you measure the fitness of the cell as a product of some measures of viability and fecundity, the metabolic cells are going to have no fitness because they're not helping in reproduction, and vice versa for the reproductive cells, the gametes. So this is not going to help explain the fitness benefits of specialization that you get in the organisms that are multi-cellular. The point of my first paper in this series was to argue that this is the wrong framework (Bossert et al. 2013a). But Samir was onto the right idea—social choice has the tools to deal with this problem.

Now there's an extension of Sen's framework, largely pioneered by Kevin Roberts, who has done some of the most important work in social choice. He called it the double aggregation problem (Roberts 1995). If you think about social welfare functionals, you are taking a profile of the utility functions for each person and aggregating them into a social order of the alternatives. You capture interpersonal comparisons by saying that if we take certain kinds of transforms of these profiles, things are invariant. In the ordinal non-comparable case, if I take independent monotone transforms of the utility functions, that new profile has to lead to the same social ranking because we haven't changed any of the allowable information. Well, you can think about that as if some planner has constructed this social welfare functional, and it's his or her interpersonal comparisons that are underlying it. But what if we had somebody else making the comparisons? They might make them differently. So, you have a multiple planner problem—there is double aggregation. You have to aggregate for a given planner, and you have to aggregate across them. This is now usually called the extensive social choice problem.

I did this with Chloe Qi—who is a former student—and Walter Bossert in our first paper, which was published in *Biology and Philosophy* (Bossert et al. 2013a). It was really a comment on what Samir did, saying that you have to disaggregate more. You have to think of the viability and the fecundity functions separately so that you can capture the gains from specialization. That's like having two different outside observers. One is looking at it from the viability perspective, and one is looking at it from the fecundity perspective. This is the formal structure you have in extensive social choice.

I had the basic idea after I had read Samir Okasha's paper, but I had a very sharp undergraduate who wanted to write an honors paper with me, Chloe Qi. So, I suggested to her that she work on this problem under my direction. She made some good progress. Part of what she did was related to what I have just described, and part was about another index by Richard Michod and his co-authors (Michod et al. 2006). Our later paper involved characterizing this index. That research became the second article of the series (Bossert et al. 2013b). When it came time to actually clean this work up and publish it, Walter agreed to help out, and we ended up with two articles.

The second index I mentioned is in an article by a large group of authors (Michod et al. 2006). One of them is Yannick Viossat, who is an evolutionary game theorist at Paris-Dauphine. When I was writing this work up—I can't remember if I actually had a draft of it—there was a one-day conference that Philippe Mongin put together at Paris-Dauphine where I presented this research. I had not met Viossat prior to that, but we had a chance to talk. I was going from there to LSE. Their philosophy department is primarily a philosophy of science department, and they have something called the Choice Group that has a regular seminar in which I was to give a talk. I presented this material there. Samir Okasha was at Bristol and I had not met him yet, but I had written to him and corresponded about this research. He said that he would come to the seminar, which he did. We talked for quite awhile afterwards and he suggested that we work together.

Samir proposed that we work on something called inclusive fitness. Say you have two siblings and there is a choice between saving yourself and saving them, each sibling sharing half your genes. In terms of evolutionary survival, the same number of your genes survive by saving two of your siblings or by saving yourself. This idea had been raised by J. B. S. Haldane (Haldane 1955), but not in any formal way. Bill Hamilton, who was one of the great evolutionary biologists of the last century, formalized this idea into what is called Hamilton's rule (or measure) of inclusive fitness (Hamilton 1964). Samir suggested we work on axiomatizing it. Together with Walter Bossert, we axiomatized it, and that became the third paper in the series (Okasha et al. 2014).

Samir's social choice research was connected to Thomas Kuhn's theory about the structure of scientific revolutions (Kuhn 1962), and his subsequent work looking at selecting between theories (Kuhn 1977). You have these various criteria: simplicity, fit, and so on. And it's an aggregation problem. Samir published a very influential article on this topic in a philosophy journal called *Mind* (Okasha 2011). I refereed it. I had actually read it before learning of his other work. At *Mind*, they use double blind refereeing, so I initially didn't know who had written this piece. But once I found out, I didn't know Samir's work, so I looked it up, and that is how I found the *Biology and Philosophy* piece.

One of the things about Samir's paper on theory choice is that I think that he's dead right that the social choice framework is the right framework to look at that problem. The one weakness of his article is that it is not so clear that unrestricted domain—which he uses—is a natural assumption. Michael Morreau, a philosopher who also became another co-author of mine, wrote a comment on that point that also appeared in *Mind* (Morreau 2015). This has now become a hot topic in philosophy of science. Just a week ago, there was a meeting of the European Philosophy of Science Association in Germany where they had a whole session on it.

CdA: OK, I think we should now go on to optimal taxation. Maybe, Felix could ask some questions on that?

FB: Yes, sure. We'll start from social choice and then move to the core of optimal taxation. I have a recollection of a paper by Samuelson in the'60s about Arrow's Theorem, which is called "Arrow's Mathematical Politics" (Samuelson 1967). What he's suggesting there is that this is maybe about mathematics and maybe about politics, but certainly not about economics. According to Samuelson, the typical public policy problem the economist would study is one where you have enough data available to know what the set of Pareto efficient allocations is. And the set of Pareto efficient allocations is something that only depends on ordinal information. Then you are left with the problem of how to select among the different Pareto optima.

That's the problem the economist faces, and then social welfare functions are just one convenient way of formalizing a preference over the set of Pareto optima. How do you like this defense of the use of social welfare functions?

JW: The exposition in Samuelson's paper leaves a lot to be desired. I don't think we can know for sure exactly what he was claiming, There were a couple of points that I think he's making. One is that he's got a different framework from Arrow. Arrow has a multi-profile framework. Samuelson, I think, is saying that we've got the preferences, and we only need to worry about that one profile. Not everybody agrees with that, but that paper has been very influential in leading to what's called single-profile social choice. When people do single-profile versions of Arrow's Theorem, they have to assume a neutrality principle directly, which is basically saying that if the pattern of preferences for a over b across people is the same as c over d, then the social ranking of a and b is the same as the social ranking of c and d. You get that in the multi-profile version by having a rich domain assumption together with Pareto indifference and independence of irrelevant alternatives. But if you only have one profile, you've got no independence condition to give the neutrality.

An alternative interpretation of what Samuelson was talking about is that preferences may change and you have to allow for that, but you don't require independence. So, you're not forcing neutrality. He's very critical of neutrality. He has an example in terms of chocolates. You've got a hundred chocolates to share between Felix and Claude. They only care about how many chocolates they get. Look at transferring chocolates between people. If we take one distribution and move a chocolate from Claude to Felix, you say, "Well, that is better," maybe because Felix didn't have any. Any time we transfer another unit from Claude, neutrality then implies that it's better. But then Felix will end up with everything. So, neutrality is a terrible principle.

That's Samuelson's fundamental objection. I think he really is in a single-profile framework, which is fine for some applications. I believe that people differ on how they feel about independence in a multi-profile case. This issue played a role in the paper in Sen's honor where we proved the full n-dimensional version of Arrow's Theorem based on geometric arguments (Blackorby et al. 1990). The first part of that paper has a result on single-profile welfarism, where we show that in a single-profile case that welfarism is equivalent to Pareto indifference. That partly grew out of thinking about Samuelson's paper.

FB: Would the single profile interpretation be appropriate for a Mirrleesian model of income taxation (Mirrlees 1971), where we just write down a utility function and based on that utility, we characterize the whole set of Pareto efficient tax schedules?

JW: When you look at the Mirrlees model in detail, you've got some cardinal measure of utility, which you don't have with Samuelson—at least in that paper of Samuelson's. How you interpret whether he allows for cardinality or not depends on which paper you're reading. I think it's an issue that you can't pin down based on textual analysis. Some papers sound like Samuelson only considers ordinal cases, others like there's more. The Mirrlees model is single profile, but it's a profile of utility functions that have cardinal significance, so that you can at least compare utility gains and losses when you talk about utilitarianism.

FB: Right. Given the complications of all these assumptions, do you think it's a strength or a weakness of the Mirrleesian approach to work on the single-profile assumption? Is it a strength, in the sense that it makes us talk about something we want to talk about without having to talk at the same time about too many other complicated factors? Or is it a weakness, in the sense that you simplify the value judgments which are inherent in the analysis so much that it's inappropriate?

JW: No, because we can take a multi-profile approach, the Sen welfare functional approach. I should say that although Sen developed this approach and did a lot of the important work using it, there's actually a precursor to it. There's a brilliant chapter on social choice in Luce and Raiffa's *Games and Decisions* (Luce and Raiffa 1957) book from the mid '50s. They don't use the Sen terminology, but social welfare functionals are there. They have matrices, with rows (or columns—I forget which) being the people, the columns being alternatives, and the entries the utilities. You want to aggregate these matrices into social rankings. They don't do much with it, but you can use that framework to axiomatize social objectives. In the paper I've mentioned before that Claude wrote with Louis Gevers (d'Aspremont and Gevers 1977), there's an axiomatization of utilitarianism using Sen's version of this framework.

So, you've got a foundation for the utilitarian objective function using this framework. You could take the Mirrlees problem and say, here's one profile and we're going to use this criterion. and if the preferences change, we'll redo the analysis. If you go back to my comparative statics of optimal income tax paper (Weymark 1987), when I change a parameter in the preferences, we're moving to a new profile and seeing what happens to the solution. You can justify the objective function in the way that Claude and Louis did. Mirrlees was only looking at a single profile, but when you do the comparative statics, you could be changing the profile and seeing what happens to the solution.

FB: To make sure I understand this correctly, you're saying that there is no need to have the single profile in the Mirrleesian analysis? To allow for multiple ones, you would just have to pick a welfare function for every possible profile. Then you would basically do the same analysis that you did before?

JW: Yes, and in his case it's always utilitarianism.

FB: OK, good. You have the work with Craig Brett about the political economy approach to distributive income taxation (Brett and Weymark 2016, 2017). Is this a way of trying to figure out what the relevant social welfare function is, at least for a problem of income taxation?

JW: Well, that wasn't the way we thought about it. There had been a few early papers on voting over tax schedules showing how you would get cycles. Kevin Roberts had a very important paper (Roberts 1977) where you get around this. The impetus for our work was a very important unpublished paper of Ailsa Röell (Röell 2012). When voting over nonlinear income tax schedules, you will generally have Condorcet cycles. Ailsa's interesting observation was to allow for nonlinear tax systems, but to restrict the set of tax schedules we vote over so as to avoid cycles.

In the Mirrlees income tax problem (Mirrlees 1971), you've got two sorts of constraints. You pick a nonlinear tax schedule that's the same for everybody, so how much tax you pay only depends upon how much income you have. There's a taxation principle which says that having people optimize off the common budget set implied by that common tax schedule is equivalent to picking the allocation directly subject to the standard self-selection constraints. So, one way of writing the incentive constraint in that problem is to pick the allocation subject to self-selection constraints. The allocation also has to satisfy the standard materials balance constraints—you can't have more consumption than you've produced. So those are the constraints in the Mirrlees problem. As an objective, he has welfare maximization, in his case, utilitarian maximization.

Ailsa said, "Let's keep those constraints." (She actually has another constraint which I'll just leave aside because she doesn't use it for her voting result in any serious way.) "We've got different types of people, and we ask each one, if you were a dictator, what tax schedule would you pick?" You have to satisfy the same constraints as the Mirrlees problem, the incentive constraints and the materials balance constraint. Then we're only going to vote over these selfishly optimal schedules. In the case where the utility functions are quasilinear, linear in consumption, she has a remarkable result: because people only differ in the skill level in the Mirrlees model, it's a one-dimensional asymmetric information problem. We can index the selfishly optimal schedules by the type that proposes each one. She shows that the preferences are single peaked over the schedules when you index them in that way. Then you can use Black's Median Voter Theorem (Black 1948). She does that by first identifying some of the qualitative properties of the selfishly optimal schedules. Her proof that there is single-peakedness is quite clever—it's not straightforward.

There's a version of the paper that was circulated in '96. It actually isn't quite complete; the conclusion isn't all there. It's something she worked on as a graduate student in the '80s, but she never did anything with it. I'd gotten a copy from Ailsa when I saw her in the late '90s. I'd read it and thought that this was a really remarkable paper, and that I should work on something relating to it at some point. One time Craig Brett was visiting me to work on something else and it went faster than we thought, so we had a couple of days before he went home. I said, "Let's delve into this," because I thought there were techniques we had used that would simplify the analysis.

In my first *Journal of Public Economics* paper on optimal income taxation (Weymark 1986b), I was working in a quasilinear environment. I investigated it using sub-optimization, eliminating one of the variables in a first stage and then optimizing over the second variable. The quasilinearity played a big role in doing that. When utility is quasilinear in consumption, in the first sub-optimization problem, optimal consumptions are chosen for given incomes. When the functions that show how the consumptions optimally depend on incomes are substituted into the objective function, you are left with something that no longer has the consumption variables. You can't just pick arbitrary incomes because the incentive constraints require that the incomes be non-decreasing, which is the only constraint other than the non-negativity constraint on consumption.

I thought that if we used these insights to reduce the dimension of the problem by sub-optimization, we should be able to do more than Ailsa had done. We first started to do this with a finite type model, which is what she was working with. We got stuck at some point, so we moved to a continuum. I think that we can now move back to the finite type case using the same basic ideas (Brett and Weymark 2020). Using sub-optimization, we obtain a problem that only involves solving for the incomes as a function of type.

Mirrlees (1976) showed that with the incentive constraints, there's a first-order condition—it's essentially an envelope condition—and then there is a second-order condition that the incomes be non-decreasing. Most of the optimal income tax literature just ignores the second-order conditions. In trying to understand the problem, we thought we would ignore that first, and then add it back in.

We quickly discovered that the solution for what a particular type would propose had a very simple form if you ignore the second-order condition. When we're trying to find incomes as a function of type, the maxi-max schedule lies above the maxi-min schedule. If we look at some type in the middle of the distribution, it wants to use maxi-max below its type so as to pull resources up towards itself because it is the biggest type of the people below it. For people above it, it wants to pull resources down towards itself, so it's like it has a maxi-min objective with respect to higher types. But if you do that, because the maxi-max schedule lies above the maxi-min schedule, you're going to have a downward discontinuity, and that violates the second-order condition. It turns out what you have to do is iron. You've got to put in what we call a level bridge. So the full characterization with all the constraints is that as you increase the type, you first follow maxi-max, but at some point you bunch a whole group of people together—they all get the same income and the same consumption, and then for higher types, you follow maxi-min schedule. So you have to prove that it's the right thing to do to link the maxi-max and maxi-min schedules in this way and determine where the end points are, which is what we do.

Once you've got that, it turns out to be relatively straightforward to prove the single-peakedness result. But that is because we've got a complete characterization of what the selfishly optimal schedules look like. So, that paper arose from trying to understand what was really driving Ailsa Röell's results, and thinking that there were techniques we had that might allow us to simplify the problem. In the process, we actually learned a lot more about the structure of the problem.

FB: So in some sense, you are saying that there is a really close relation between optimal taxation and mechanism design?

JW: That's right. But when you're working with these optimal tax problems, you're not allowing right from the start all possible mechanisms. We make the assumption that we're picking the same tax schedules for everybody, so how much tax you pay only depends on your own income, not on everybody else's income. In work that Felix has done, he calls them simple mechanisms (Bierbrauer 2011). You could have a more general mechanism. He's investigated when, in continuum models, these simple mechanisms are optimal. We're following the optimal income tax tradition; you just start with that as a restriction on a class of mechanisms you want to look at.

FB: There was one aspect of my question which you have not yet answered. You can take the political economy perspective on the Mirrleesian problem, and you can take the perspective of welfare maximization. How do these two things talk to each other? One way, and that was what I had expected as an answer is to say that we have the tradition of asking questions in the spirit of the second welfare theorem. So, we ask, "What among these welfare optimal or Pareto efficient outcomes can actually be decentralized in one way or another?" And political competition would be one way of decentralizing outcomes. There could also be a more simple-minded motivation, to say we want to have a prediction about what tax schedule the political process will produce. None of this has been present in your answer.

JW: No, but the second theorem of welfare economics is fundamental for understanding the whole point of optimal tax theory. The second theorem of welfare economics says that under appropriate convexity assumptions, if we pick your favorite Pareto optimal outcome, then we can obtain it if we suitably lump sum redistribute resources initially and let the markets work.

The thing about the second theorem is that in order to determine what those personalized lump sum taxes and transfers are, you need a lot of information. You need to know the preferences of the people, and you need to know the endowments. And the point of departure for optimal tax theory is that we don't know that, and so we're going to look at more limited kinds of information. In the income tax literature, you do that by saying that we can't have personalized tax schedules. Everybody has to face the same schedule, but we allow for nonlinear ones.

It also motivates, in the commodity tax literature, the fact that everybody is going to face the same commodity tax rates. In terms of trades on the market, everybody has the same budget set. If they have different endowments, they're going to have different budget sets in terms of final consumptions, but what you actually observe are market transactions, everybody has the same budget set. So, the motivation behind the whole literature is that we're limited because of the asymmetric information about what kind of instruments we have.

Voting is one mechanism for making decisions. Welfare maximization is another. I think they're both worth exploring and trying to see if there are nice connections between the two. I haven't done much on what the connections between the two are. In a sense, this paper we've just been talking about (Brett and Weymark 2017) has some because the lowest skilled uses a maxi-min objective and the high-skilled proposes the maxi-max one. So, there are connections.

FB: What I found interesting in your answer is that you say both welfare maximization and voting are ways of making a decision.

JW: Yes.

FB: That's interesting. You wouldn't say that welfare maximization is just a nice thought experiment for researchers, and voting is what happens in the real world? You would say both are ways to come to a decision, even if it's just interesting to compare them?

JW: Yes.

FB: It's not that one is empirically more plausible than the other?

JW: Well, actually, probably not. I think if you wanted to get something more empirically plausible, you'd have to start getting into bargaining solutions.

FB: Right.

CdA: In representative democracy, this would be quite reasonable, no?

JW: The paper I did with Felix and Craig Brett (Bierbrauer et al. 2013) was not looking at voting. You have two countries and people could move between them, and that's going to limit the amount of redistribution. In each country, you have average utilitarian governments. They are welfare maximizing within each country, and they want to redistribute from rich to poor based upon whoever finally ends up in its country. But people can move, so if you try and tax the rich too much, they're going to move to another country. If you offer really great redistribution at the bottom end, the poor from another country are going to move in, and it really limits the amount of redistribution.

There's really nice work on these kind of tax competition models, like Massimo Morelli, Huanxing Yang and Lixin Ye's piece in *AEJ: Micro* with countervailing incentives (Morelli et al. 2012). The middle group wants to draw resources from the top and the bottom, the top want to move resources up towards them, and the bottom want to move resources down towards them. How that gets traded off in the voting process could depend upon the relative size of this middle group.

There are nice things going on in this literature, but there's not a lot of it yet. There's some combining of welfare maximization and political economy. I think they're compatible. I know some people don't think that is possible. Morelli told me they initially had some concerns raised about their work because people didn't like the fact that there was a mix of political economy and welfare maximization in the same problem.

FB: I would like to get back to the second welfare theorem and also to the relation between the theory of taxation and the theory of mechanism design. When you were discussing the second welfare theorem, you made the observation that the moment we have private information on productive abilities, we can no longer apply it. We need something else. But there are two alternatives. One is, as you said, we can go for optimal mechanisms, or we could go for tax systems. And there are some situations where the two are the same thing, but there are other environments where they aren't. For instance, if you have a finite population with a known cross-sectional distribution of productive abilities, that's a particular environment that we know from work of Thomas Piketty (Piketty 1993) that gives rise to correlations among types. If you have just two people and you know one is rich and one is poor, types are perfectly negatively correlated and, in principle, you could come up with a mechanism that exploits this correlation to have a first best alternative.

JW: Yes. Before you go on, the big difference between these finite type models and the continuum models is that if one person deviates by not telling the truth, you can't detect it in the continuum model, but you can in the finite model. That's what

Piketty is exploiting—you know the distribution, it's common knowledge. And so if what you see isn't the right distribution, you know somebody has deviated. But the mechanisms you need to deal with this, what he calls generalized tax schedules, are complex. How much tax I have to pay in one of his mechanisms depends not only on my own income, but on everybody else's income, in principle.

FB: Yes, yes.

JW: So, I think the justification that somebody like Jim Mirrlees would offer is: because we've got this asymmetric information, we need to use anonymous tax schedules. It's a little too quick. As you were saying, in these finite type models, while we could use something more complex and still respect the asymmetric information, we need to justify what in your own work you call simple mechanisms—where the tax schedule only depends upon your own income. At some point, you say that generalized tax schedules are unrealistic. People aren't going to put up with having their tax depend on what somebody else's income is. So, it's a natural way to start restricting the class of mechanisms you look at; you're not looking at optimal ones.

FB: OK, you answered the question that I was after. I wanted to understand the modeling choice that you made. It's an appeal to realism. I have a related question. In public finance, we had this important paper by Atkinson and Stiglitz in '76 (Atkinson and Stiglitz 1976), where they're saying that in many situations, commodity taxation is superfluous if you have the income tax at hand. In what sense is studying Ramsey models of taxation after the Atkinson and Stiglitz critique different from studying Mirrleesian taxation after Piketty's critique?

JW: That's a tough one! [All laugh.]

FB: Because the basic argument is the same. You say there is something better, it's even respecting all the constraints that we have. Why don't you go for the thing that's better?

JW: OK, I've given you an answer in terms of the Piketty approach.

FB: So, your answer to that critique was to say that the Mirrleesian income tax problem looks so much more reasonable. But still you would never study a Ramsey problem.

JW: No, I've never written a paper on Ramsey problems—I teach them.

FB: So, what is the difference? Put differently, what is the difference in the balance between simplicity and theoretical appeal? Why does it make you go for Mirrlees rather than Ramsey in one kind of situation, and make you go for Mirrlees over Piketty in another one?

JW: I really don't have a good answer to that off the top of my head. The Atkinson–Stiglitz kind of problem is something I've never worked on. There's huge literature on it, and I've taught it from time to time. I haven't taught tax theory all that often, and it's not something I've really thought through. But it's a good question; it's something to think about.

FB: Let's move on to your *Econometrica* paper on the comparative statics properties (Weymark 1987), which you mentioned already. You did this based on the assumption that preferences are quasilinear in leisure. How do you think of this modeling choice as of today? Today, many people assume preferences that are quasilinear in consumption. Jonathan Gruber and Emmanuel Saez would defend that on empirical grounds. What are your thoughts on this? Did you have a deep reason to do it with preferences that are quasilinear in leisure, or could it just have been the other way around, and you just picked one?

JW: It could have been the other way around. There is a reason why it was done the way I did it. But let me address your question about people that claim that if you're going to use quasilinearity, you should use quasilinear in consumption, rather than quasilinear in income. Notice that the voting work we've talked about has quasilinearty in consumption, but the earlier work I did on comparative statics is in terms of quasilinearity in income. Laurent Simula (Simula 2010) worked out what it looked like in the other case.

I don't think the empirical evidence is all that relevant. We're looking at optimal income tax systems. What they look like could be very far from what we're observing in the non-optimal tax systems that are out there in practice. So, the fact that empirically it looks like we're closer to quasilinearity in consumption, that's only when you're looking at it in terms of these non-optimal tax systems. What things look like if we're really looking at an optimal tax system could be very, very different.

Much like how the work I did with Craig Brett on voting grew out of trying to deeply understand what, in that case, Ailsa Röell did, the choice of quasilinearity in income in the original paper was motivated by a nice short paper by Jean-Charles Rochet and Stefan Lollivier in JET in the early '80s (Lollivier and Rochet 1983) where they were investigating a Mirrlees model with quasilinear preferences. They assumed quasilinearity in income. They had this very interesting result that in this quasilinear framework with a continuum of types, any bunching is all down at the bottom of the skill distribution if the distribution is uniform.

I haven't written very many papers using a continuum approach because I find that I have much better intuition working with discrete types and spaces. You can understand the geometry underlying the problems and the incentives much easier. In trying to understand the Rochet and Lollivier paper, I asked, "What would happen with a finite number of types?" The reduction of the dimension of the problem that I mentioned before that happens in their paper is kind of implicit. You can do this reduction in a finite type problem as long as you have some mild assumptions. In a Mirrlees-style problem, I used a weighted utilitarian objective. You want the adjacent downward self-selection constraints to all bind and you also want the economy-wide production constraint, or the government budget constraint, to bind.

The way you do this dimensionality reduction—when you're assuming quasilinearity in income—is to first fix the distribution of consumption across types, and find what the optimal income vector is conditional on these incomes. As a first step, just fix the lowest income. We know that you can find the income of the second-lowest person by going up to the consumption of type two that we pre-specified along the type two indifference curve through the type one bundle. When you've got up to that income, it tells you what the type two income is. And then just iterate up. But when all is said and done, you may not have budget balance. Then you just move the whole schedule parallel so you don't upset these binding constraints, and that pins down where you should have set the lowest income (see Weymark 1986b).

Again, the choice of quasilinearity in income was only made only because that's what Rochet and Lollivier did. I could have done it the other way around. Once I saw this in this finite type case, I realized that you could start doing comparative statics, which didn't exist for nonlinear taxes. In fact, there was this paper by Guesnerie and Seade (1982) that I've mentioned where they're looking at nonlinear pricing and nonlinear taxation with finite types. They actually say something misleading about doing comparative statics. If you think about the pattern of binding constraints I described, they say that if I move the bundle for some type, I'm going to get discontinuous jumps in what happens. But that's not the right experiment to look at because what you wanted to do is re-optimize. That's what I did.

If you look at the case where you have an interior solution and there's no bunching, you can solve for each variable separately. The solutions are a function of parameters, and you can start moving the parameters around to get the comparative statics. Doing this also enabled me to understand why you have bunching at the bottom with Lollivier and Rochet—it's not true in general in the finite type model (Weymark 1986a). But if the distance between adjacent types is a constant—so starting with type one, you go up to type two and, say, add two to its skill level, go up to type three and add another two, then the types evenly spaced and there is only bunching at the bottom. Implicitly that's what's going on with Loliiver and Rochet because they worked with the uniform distribution. If you have a uniform distribution in a continuum, it's like having the types evenly spaced.

CdA: So, you seem to say that this specific type of quasilinearity comes out of the problem in some sense.

JW: The choice I used in that paper was simply because my starting point was Rochet and Lollivier and that's what they used, but I could have done it the other way around. And, as I've said, it has been done. For the comparative statics, I only looked at changes in certain parameters. I changed the marginal disutility of income term that appears because things are quasilinear, the slope of the budget constraint, and some of the welfare weights in a weighted utilitarian objective. There are other parameters that could have been changed related to the skill distribution. What happens if we change the value of some skill? What happens if we change the distribution of skills for a given type? At the time, I didn't know how to solve these questions, and so I teamed up with Craig Brett. We tackled them and made progress (Brett and Weymark 2008, 2011).

Craig had been a graduate student at UBC. At the time he was coming through, I never taught any public economics. Chuck Blackorby had decided to learn tax theory and he used Guesnerie's monograph (1995) as the textbook. Craig was in the class and he could understand it. So, Chuck was the supervisor, but I was on the committee. Craig wrote a great thesis, and then afterwards we started collaborating.

FB: So, maybe one last, very broad question on the theory of optimal taxation. I had a disappointing experience when I was a young Ph.D. student. I met Mirrlees, and I had worked on theories of taxation, and it was a great opportunity to talk to him for me. And I asked him what his basic attitude to this whole research program was—whether he thought of this as being a tool for deriving policy recommendations, or an abstract theory of equity-efficiency trade-offs. And he refused an answer. He said I should decide that for myself. How would you answer, and are you sympathetic to this agnostic attitude of Mirrlees?

JW: Well, I think there's aspects of both phenomena. At the level of abstraction that optimal tax theory operates at, it can't pin down specific programs, specific policies. But I think it highlights the kind of considerations that should be taken into account when you design policies. From the other perspective, I think it shows a lot about how equity and efficiency considerations interact, and you need to take them into account when you're designing policies.

Early on when I was working on commodity taxation, I was initially looking at optimal commodity taxes. There was a period around the time that I was a young Assistant Professor that people started looking at tax reforms, I think largely because of work by Feldstein (1976). We're not involved in designing a whole tax system from scratch, we're making marginal changes from where we are. A number of people, like Dixit (1979), Diewert (1979), and Guesnerie (Guesnerie 1977)—in fact, one of Jean Tirole's first publications was one with Guesnerie on tax reform (Tirole and Guesnerie 1981)—started looking at what directions we could move outcomes in a welfare- or Pareto-improving way. In the context of a commodity tax problem, what we need is some knowledge about excess demands for goods on the part of individuals, but we also need to know the elasticities at the aggregate level. We don't need that data at the individual level. There's been some very nice empirical work where you put people into groups and work with aggregate goods, Nick Stern and some of his co-authors (Ahmad and Stern 1984), also Serge Wibaut (Wibaut 1989), Was he a CORE person? He did some nice work where you go out and take the theory to the data, to try and see if there's welfare-improving or Pareto-improving changes in policies.

Actually, I can remember at the time I was at CORE, another person who worked on taxation was Knud Munk. He spent part time at CORE that year while he was working at the European Commission. They were looking at reforming different kinds of policies, like tariffs and agricultural policies. I think what this perspective does, which we talked about at the time, was that rather than looking at one kind of change in isolation, like changing a single tariff, you look for a simultaneous change in the values of a number of policies to see if you can make them, for example, in a Pareto-improving way—trying to make things politically acceptable that way. You can't do this at the level of disaggregation that our models have, but you can do it in terms of more aggregated models.

FB: Thank you. I think you once told me that the only non-academic piece that you have written was on the problem of economic growth in Brazil, if I remember that correctly. And that brings me to the question of how the fields in which you would be ready to communicate with the broader public are related to your fields of expertise as a researcher? Is it that the more you know as a researcher, the less competent you feel about giving advice?

JW: I like academic, theoretical work. I like making precise statements from precise models. And I think when you're giving policy advice, you're going to have to make implicit, kind of fuzzy statements. You're not going to be able to back them up completely by a model. I find that hard to do.

The piece was in a Brazilian journal, *Exame* (Weymark 2010). My Brazilian friends told me that it's the Brazilian version of *The Economist* and they have, on occasion, special issues that matter. They wanted me to write a short piece of a couple of thousand words that they would translate into Brazilian Portuguese on pro-poor growth, if possible related to Brazil. I'd never done anything like this. I knew a little bit about the relevant subject and I knew a little about the Brazilian experience because I'd been reading about experiments with conditional cash payments for doing something like keeping your kids in school or for young mothers going to training sessions where they learn about good nutrition to help raise their kids. I had also read about pension reforms. I knew a bit about these issues and so I thought, well I knew enough about the theory and it wouldn't take me long to put in some illustrations. It ended up taking longer than I thought. I didn't have a lot of time to do it.

This is the first and only work for hire that I did. But one thing I should have known—I knew, but I wasn't really thinking about it—was that the editorial staff would edit the piece quite a bit. They didn't really distort my message, but that's what happens with journalism—you don't get the final say of what things look like. I had somebody who's Brazilian read the piece back to me.

The title in English would be "Dawn over the horizon, how beautiful." I did not pick that title. It deals with the punch line. Brazil in the decade before that piece was written had made unbelievable progress in reducing poverty at the bottom end of the distribution, in part from trade liberalization policies, but also from pension reforms, particularly for the rural poor, and through these conditional cash payment programs. The bottom line of the paper was that the pension reforms had an immediate impact, but these other policies are not going to have an impact for a decade or so because you won't see gains from improving human capital among children until they're on the job market. And so that was where the funny title came from.

Along the same lines, I had another opportunity to do something that had policy relevance, but in the end I didn't do it. In the early '90s, I had a visiting chair at Johns Hopkins. The chair rotated among a few departments. You had to be from the British Commonwealth to hold the chair, and there had been a requirement, which was not being enforced anymore, to give a public lecture on some constitutional issue. This was not long after there had been a lot of debate in Canada about Quebec separating, and in particular among economists about some of the economic issues and consequences of separation. Robin Boadway, who is one of the great public economists and who is also from the same hometown as me, is very good at writing good theory and writing these kind of policy pieces, and he'd written a nice economic analysis of the relevant constitutional issues (Boadway 1992). So, initially I was tempted to do the talk. And then I thought, it's just not me. And Baltimore is so close to Washington DC that I'm sure people from the Canadian embassy would show up and I'll just embarrass myself, so I'll stick with things that I know. So, I just gave talks on my research.

CdA: OK, so maybe moving to some very important tasks that you have done—the many editorial boards on which you have served. You have gone through a huge amount of papers, referee reports, and so on. Can you tell us what for you are the qualities of a good editor? And what challenges you faced, and what is generally your experience in that aspect? It's a very important aspect, and you did a lot of work on that.

JW: Yes. I'm easing out of that because it's very time consuming. Something I said earlier was that when I was starting off, there were some senior people who were very, very good and who were supportive of me. They showed me through example, and so I want to help young scholars. When I do refereeing, one of the things I like about not having double blind refereeing is that if I know it's a young person, I'm willing to spend more time to try and help him or her develop the paper. A young scholar may have a good idea, but it may not be expressed very well—and I'll put a lot of time into providing advice. It's not that I don't give serious comments on other people's papers too, but I try and particularly help young scholars.

With rare exceptions, I want to know that any paper that I recommend as a referee to be resubmitted or accepted is right. Lots of referees don't check for correctness. I became sensitive to this very early in my career. There's a piece by Roger Guesnerie and Jean-Jacques Laffont in JET on taxing price makers (Guesnerie and Laffont 1978). It's an optimal tax paper where they have non-competitive markets. It's a fabulous paper. I was a referee of it and I couldn't understand one result—I thought that it's not stated correctly. I thought I had a counterexample. I puzzled over it, put it aside, came back to it in a few days and thought, "You know, it's close to being right, but I think you need an extra assumption." However, I couldn't find a flaw in their proof. After a couple of days, when I went back to the paper, I found the flaw in the proof, and you did need this extra assumption. I thought, "These guys are brilliant and if they can make a mistake like that, think about mere mortals." A large number of papers I referee have errors that are typically easily fixed. I think it's important when we have papers published that you can count on them. And so as a referee, that's one of my goals.

Initially, my editorial roles were not as a main editor, but often as an associate editor that handles the manuscripts. I would send them out to referees and then I would also prepare a report. In that role, I would also use the same rule. But when I became a Managing Editor of *Social Choice and Welfare*, that was just impossible. And the same is true with *Economics and Philosophy*. I guess a guiding principle is to treat people fairly.

Sometimes there are different ways of approaching a problem, and it can get ideological. I think that on the whole, people working in social choice and welfare economics are pretty open-minded, but it's not true in all areas of economics. On occasion, I'll get something as an editor where I think that the referee is taking too narrow a view and that what the author is doing is reasonable. I think that if reasonable people can disagree about how something is modeled, just because it's not your way, you don't turn it down. Let us get this work out into the public domain, and let people investigate it. So, that's been a major consideration in my editorial policy.

Another thing that I've put a lot of priority on are publication lags. They can get very, very long, which can be a disaster for junior people who are on a tenure clock. I now do research in philosophy too and have edited *Economics and Philosophy*. For young philosophers, it's difficult to even get a permanent job. It's just so much harder in their field. Waiting a year, I just think it's morally wrong to keep people waiting that long. So, when I became a Managing Editor of *Social Choice and Welfare*, I really pushed to do everything we could to make sure that nobody waited more than six months for a decision. And I did the same thing at *Economics and Philosophy*. Many journals ask you to produce a report in three or four weeks. I think that is the wrong margin to push on. You might get a long, complex paper, and if you want a good, serious report, you have to give the referees more time. People don't mind waiting a couple of months, or three months, for a report, to get a serious review. They do mind if they have to wait eleven months, twelve months. And I think in both the journals that I've been a main editor of, there's been good progress made on getting that long tail cut down.

CdA: And did you see an evolution along the years, in terms of the publication process, did it change?

JW: Yes. Now with web-based electronic submissions, for every kind of option, there's a template. A lot of editors—for time saving reasons—tend to have standard kinds of letters and just send them out. They tend to be very impersonal, but you can modify them. *Economics and Philosophy*, when I was of the main editors, wasn't doing web-based submissions yet; they were in the process of transitioning to it. *Social Choice and Welfare* moved to a web-based platform when I was a Managing Editor. You have to put a lot of thought into what your basic letters are going to say because there are variations on the decision letters needed.

So, have a good template, but then don't just use it as a form letter. It's not that hard, once you've got good templates, to make a letter specific to the person it's being directed at. I think it's an obligation as an editor to give a good reason if you're going to turn a paper down. You can't just send the author some standard terminology—regrets, we're turning you down, the referees didn't like it, or whatever. It takes more time, but the actual writing of the letter doesn't necessarily take a whole lot more time if you've got the basic structure in a template that you can build off of.

FB: Do you think that in your fields of expertise, it has become more difficult or easier to publish your work also in general interest journals?

JW: Well, I've never actually published much in general interest journals, so I can't say. I haven't actually sent many papers to general interest journals, so I'm not quite sure what to say about that.

CdA: We know one paper was published fast; it was the paper with Donaldson on Ginis and the principle of population (Donaldson and Weymark 1980).

JW: Actually, my very first publication was extremely quick. It's in a philosophy journal called *Philosophical Studies*, which is actually a pretty good journal. It had a technical result. I was criticizing some work by Nicholas Rescher (Rescher 1975), who had written on altruism. He started with a simple prisoner's dilemma and said, "We have this non-cooperative behavior that leads to something that neither person would like. What if we made them altruistic, so each of them gives a little weight to the other person's utility? By doing that, in some circumstances, we can transform a prisoner's dilemma structure into one where you get outcomes that they all prefer. So, this a good argument for altruism."

The point of my paper (Weymark 1978b) was that you can start with situations that don't have a prisoner's dilemma and by doing what Rescher did, create situations that have the prisoner's dilemma structure. So I wrote this up, and I actually worked out conditions for which the altruism would result in the prisoner's dilemma form and for which it does not, which I included in the paper. *Philosophical Studies* allows a little bit of technical material. This was back when you were corresponding by mail. I had a strong revise and resubmit within ten days of submitting my paper. All the editor wanted was this little result moved into an appendix. That was the fastest acceptance I have had.

CdA: [laughs] And what was the hardest one?

JW: The hardest?

CdA: Maybe it's one which is not published . . .

FB: Always the one you're currently working on!

JW: Well, hard ones to publish were my paper with Claude here on cartel stability (d'Aspremont et al. 1983), which ended up being very influential. My Generalized Gini paper (Weymark 1981a) was hard to publish—I only think I sent it to a couple of places before *Mathematical Social Sciences*. *Mathematical Social Sciences* was in its first year of publication. I was really naïve as a junior scholar in not realizing that most places would count a publication there as almost nothing because nobody's heard of the journal. It was started by two mathematicians, and one of them obviously was the editor in charge. The editors were really good, and one of them suggested a way of simplifying the proof. I had a very good experience with them, but not the actual printing by the publisher—my paper looked like a drunk had typeset it. There were ten million corrections on the page proofs and the publisher didn't correct any of them. The whole issue was like that. Arrow had a paper in that issue, too (Arrow 1981). It was unreadable, so the publisher ended up republishing the issue by doing a second set of proofs. But the editorial process was good.

FB: There's one last topic we'd like to touch on: you have also been an academic teacher over the years. So how would you describe yourself as an academic teacher?

JW: You know, at the graduate level, you always have to make a decision about how you trade off breadth versus depth. I don't think all instructors need to make the same kind of trade-off—it helps to have the depth in some courses, and in other courses, the breadth. I tend to lean towards the depth. I want students to really understand the guts of a model, how to carefully derive results. We may not cover a huge range of topics when I teach graduate micro theory or taxation or social choice. All the arguments are there, the proofs are there, and the students do lots of problem solving.

With undergraduates, I lean more towards breadth, but again I want them to understand the models, and I give them lots of problems so that they can practice. One thing that's been true in the USA over the time since I started is that the courses have just been dumbed down. The levels at which the courses are taught are much less than when I started. I refuse to do this. I want to challenge the students; I want them to learn. Many of them will rise to the challenge, but some of them resist.

FB: Imagine the following situation. You have somebody who has passed the stage of coursework, has been proven to be a smart person, and would like to write a job market paper on optimal taxation, for example. The person is ambitious but just feels unable to come up with a good, specific research question to study in the job market paper. Are you more likely to tell the student that she should work harder, read more papers, invest in her mathematical skills? Or are you more likely to say, sometimes inspiration comes, sometimes it doesn't. Have a nice walk around the university!

JW: You know, I haven't been the main supervisor of many students. At Duke, I was on committees, but never supervised; I was too junior to supervise. The UBC Ph.D. program was quite small, but it had a big masters program where they did a thesis. I had some terrific students, and they all came up with their own topics. At Vanderbilt, you do not get very many students that want to do theory, so I'm not doing much supervising. I really think it's important for the students to come up with their own topics. I think that's a big part of what doing research is about. If they don't have great ideas, maybe their research won't go to the really good journals, but it's something they've done on their own.

I've supervised or co-supervised a few undergraduate honors papers, like the research with Chloe Qi on biology. I gave her the topic. It's rare that undergraduates even do something this technical in the USA; they don't know enough about the subject typically to come up with a topic. Another one that I jointly supervised a year ago was an honors paper in philosophy. We gave the topic to the student and let him see what he did with it—we didn't actually pinpoint the questions, just gave some advice about what to explore.

Another one that I co-supervised two years ago with Alan Wiseman from Political Science was an honors thesis by a young guy named Thomas Choate, who's very good (Choate et al. 2019). Thomas worked on an extension of Baron–Ferejohn legislative bargaining (Baron and Ferejohn 1989). He'd learned in a game theory course the year before about Rubinstein alternating offers bargaining (Rubinstein 1982), but learned about the legislative bargaining literature on his own. He went and read it, and came up with a topic. His model was not tractable—a big part of what Alan and I did was show him how to simplify the model to make it tractable. It was amazing that an undergraduate could come up with that. And he was young; he graduated at age 20. He's now at the Stanford Graduate School of Business in the Political Economy program.

So, I'll help undergraduates with the topic because they just haven't had a lot of exposure. With graduate students, I will not help them with a topic. I haven't had a lot of graduate students, but most of the ones I've had have been pretty good. My first student was Bentley MacLeod, who's at Columbia and a Fellow of the Econometric Society, so I was very fortunate there.

FB: What are teaching formats that prepare graduate students for the challenge to find a research topic that can be fruitfully analyzed?

JW: In the North American system, you do intensive coursework for a couple of years and then let the students try and find a topic. It's changing in Europe, but it used to be more that you'd work with somebody. You'd have a narrower focus. You'd start seeing the kinds of problems that your advisor would work on, and maybe come up with a topic related to them. It's almost like an apprenticeship. Many students in the US system have a horrible time making the transition from coursework to a thesis topic.

It helps them sometimes in the more specialized courses where you don't emphasize breadth. You just read a series of articles that are on the cutting edge; then you start getting some ideas. I don't do much of this in my courses because I want the students to have the foundations. Sometimes, I do a little bit of this. I don't teach graduate social choice much—you know, it's not marketable in the USA. But when I do, I might spend two-thirds of the course covering a set of topics so that the students get the basics of the subject; then we read a few papers. Or I will have them write a term paper or make a presentation where they take a couple of papers that are related and relatively new so that they see what kind of issues are getting attention.

On occasion, when I have time and I'm not overloaded with journal editing or administrative work, I organize reading groups. Right now, I've got a reading group. Paul Klemperer and Elizabeth Baldwin are using tropical geometry to study auctions and other things (Baldwin and Klemperer 2019), and I think this has potential for other mechanism design problems. So, we've got a reading group reading papers on mechanism design and also learning tropical geometry. I've got one Vanderbilt student in this group, but at some other university, I might have had more. I've had reading groups where we've had more students than that. They're not part of a formal course structure, but it's doing math and seeing some current literature, and hoping to come up with new research ideas by doing it. You need both breadth and depth, but it doesn't have to be, as I said, in the same course; it can be one person doing the depth, one person focusing on a few papers.

FB: So, ready for your final question?

CdA: Yes, I think we should really stop here, but the last question is to ask you is if you have something to add or question to answer that we didn't ask? Maybe, we have missed something?

JW: One thing I said is that I tend to cycle through topics. I mentioned the special program I was in as a first year student, a lot of it was very philosophical. I never actually took that many philosophy courses. I took a few. More so at the graduate level, like a course in political philosophy. But I've always had this strong interest in philosophy, and it's partly why I'm drawn to subjects like social choice and welfare economics. I've used the past few years to seriously learn more philosophy and to publish in the area. This is something I'd like to continue to do.

CdA: Thank you very much!

FB: Thank you.

Acknowledgements We are grateful to Erika Berthold, Chris Evans, and Risa Pavia for their assistance in preparing a transcript of this interview.

References

Aczél, J. (1966). *Lectures on functional equations and their applications*. New York: Academic Press.

Ahmad, E., & Stern, N. (1984). The theory of reform and Indian indirect taxes. *Journal of Public Economics*, *25*, 259–298.

Arrow, K. J. (1951). *Social choice and individual values*. New York: Wiley.

Arrow, K. J. (1964). The role of securities in the optimal allocation of risk-bearing. *Review of Economic Studies*, *31*, 91–96. Originally published in French in 1954.

Arrow, K. J. (1981). Jacob Marschak's contributions to the economics of decision and information. *Mathematical Social Sciences*, *1*, 335–338.

Atkinson, A. B. (1970). On the measurement of inequality. *Journal of Economic Theory*, *2*, 244–263.

Atkinson, A. B., & Stiglitz, J. E. (1976). The design of tax structure: Direct versus indirect taxation. *Journal of Public Economics*, *6*, 55–75.

Baldwin, E., & Klemperer, P. (2019). Understanding preferences: "Demand types", and the existence of equilibrium with indivisibilities. *Econometrica*, *87*, 867–932.

Barberà, S. (1983). Strategy-proofness and pivotal voters: A direct proof of the Gibbard-Satterthwaite Theorem. *International Economic Review*, *24*, 413–417.

Barberà, S., & Peleg, B. (1990). Strategy-proof voting schemes with continuous preferences. *Social Choice and Welfare*, *7*, 31–38.

Baron, D. P., & Ferejohn, J. A. (1989). Bargaining in legislatures. *American Political Science Review*, *83*, 1181–1206.

Bierbrauer, F. J. (2011). On the optimality of optimal income taxation. *Journal of Economic Theory*, *146*, 2105–2116.

Bierbrauer, F. J., Brett, C., & Weymark, J. A. (2013). Strategic nonlinear income tax competition with perfect labor mobility. *Games and Economic Behavior*, *82*, 292–311.

Black, D. (1948). On the rationale of group decision making. *Journal of Political Economy*, *56*, 23–34.

Blackorby, C., & Donaldson, D. (1977). Utility vs equity: Some plausible quasi-orderings. *Journal of Public Economics*, *7*, 365–381.

Blackorby, C., & Donaldson, D. (1978). Measures of relative inequality and their meaning in terms of social welfare. *Journal of Economic Theory*, *18*, 59–80.

Blackorby, C., Davidson, R., & Donaldson, D. (1977). A homiletic exposition of the expected utility hypothesis. *Economica*, *44*, 351–358.

Blackorby, C., Primont, D., & Russell, R. R. (1978). *Duality, separability, and functional structure: Theory and economic applications*. New York: North-Holland.

Blackorby, C., Donaldson, D., & Weymark, J. A. (1980). *On John Harsanyi's defences of utilitarianism*. Discussion Paper No. 8013, Center for Operations Research and Econometrics, Université catholique de Louvain.
Blackorby, C., Donaldson, D., & Weymark, J. A. (1982). A normative approach to industrial-performance evaluation and concentration indices. *European Economic Review*, *19*, 89–121.
Blackorby, C., Donaldson, D., & Weymark, J. A. (1984). Social choice with interpersonal utility comparisons: A diagrammatic introduction. *International Economic Review*, *25*, 327–356.
Blackorby, C., Donaldson, D., & Weymark, J. A. (1990). A welfarist proof of Arrow's Theorem. *Recherches Économiques de Louvain*, *56*, 259–286.
Blackorby, C., Donaldson, D., & Weymark, J. A. (1999). Harsanyi's social aggregation theorem for state-contingent alternatives. *Journal of Mathematical Economics*, *32*, 365–387.
Blair, D. H., & Pollak, R. A. (1982). Acyclic collective choice rules. *Econometrica*, *50*, 931–943.
Blair, D. H., & Pollak, R. A. (1983a). Polychromatic acyclic tours in colored multigraphs. *Mathematics of Operations Research*, *8*, 471–476.
Blair, D. H., & Pollak, R. A. (1983b). Rational collective choice. *Scientific American*, *249*(2), 88–95.
Boadway, R. W. (1992). *The constitutional division of powers: An economic perspective*. Minister of Supply and Services Canada, Ottawa: A study prepared for the Economic Council of Canada.
Border, K. C., & Jordan, J. S. (1983). Straightforward elections, unanimity and phantom voters. *Review of Economic Studies*, *50*, 153–170.
Bordes, G., & Le Breton, M. (1989). Arrovian theorems with private alternatives domains and selfish individuals. *Journal of Economic Theory*, *47*, 257–281.
Bordes, G., Laffond, G., & Le Breton, M. (2011). Euclidean preferences, option sets and strategyproofness. *SERIEs*, *2*, 469–483.
Bossert, W., & Weymark, J. A. (2004). Utility in social choice. In S. Barberà, P. J. Hammond, & C. Seidl (Eds.), *Handbook of utility theory. Volume 2: Extensions* (pp. 1099–1177). Boston: Kluwer Academic Publishers.
Bossert, W., Qi, C. X., & Weymark, J. A. (2013a). Extensive social choice and the measurement of group fitness in biological hierarchies. *Biology and Philosophy*, *28*, 75–98.
Bossert, W., Qi, C. X., & Weymark, J. A. (2013b). Measuring group fitness in a biological hierarchy: An axiomatic social choice approach. *Economics and Philosophy*, *29*, 301–323.
Brett, C., & Weymark, J. A. (2008). The impact of changing skill levels on optimal nonlinear income taxes. *Journal of Public Economics*, *92*, 1765–1771.
Brett, C., & Weymark, J. A. (2011). How optimal nonlinear income taxes change when the distribution of the population changes. *Journal of Public Economics*, *95*, 1239–1247.
Brett, C., & Weymark, J. A. (2016). Voting over selfishly optimal nonlinear income tax schedules with a minimum-utility constraint. *Journal of Mathematical Economics*, *67*, 18–31.
Brett, C., & Weymark, J. A. (2017). Voting over selfishly optimal nonlinear income tax schedules. *Games and Economic Behavior*, *101*, 172–188.
Brett, C., & Weymark, J. A. (2020). Majority rule and selfishly optimal nonlinear income tax schedules with discrete skill levels. *Social Choice and Welfare*, *54*, 337–362.
Broome, J. (1993). A cause of preference is not an object of preference. *Social Choice and Welfare*, *10*, 57–68.
Broome, J. (2008). Can there be a preference-based utilitarianism? In M. Fleurbaey, M. Salles, & J. A. Weymark (Eds.), *Justice, political liberalism, and utilitarianism: Themes From Harsanyi and Rawls* (pp. 221–238). Cambridge: Cambridge University Press.
Choate, T., Weymark, J. A., & Wiseman, A. (2019). Partisan strength and legislative bargaining. *Journal of Theoretical Politics*, *31*, 6–45.
d'Aspremont, C., & Gabszewicz, J. J. (1986). On the stability of collusion. In J. E. Stiglitz & G. F. Mathewson (Eds.), *New developments in analysis of market structure* (pp. 243–261). Basingstoke, UK: Macmillan.
d'Aspremont, C., & Gevers, L. (1977). Equity and the informational basis of collective social choice. *Review of Economic Studies*, *44*, 199–209.

d'Aspremont, C., Jacquemin, A., Gabszewicz, J. J., & Weymark, J. A. (1983). On the stability of collusive price leadership. *Canadian Journal of Economics*, *16*, 17–25.
Diamond, P. A., & Mirrlees, J. A. (1971). Optimal taxation and public production. I and II. *American Economic Review*, *61*, 8–27, 261–278.
Diamond, P. A., & Mirrlees, J. A. (1976). Private constant returns and public shadow prices. *Review of Economic Studies*, *43*, 41–47.
Diewert, W. E. (1979). Optimum tax perturbations. *Journal of Public Economics*, *10*, 139–177.
Dixit, A. K. (1979). Price changes and optimum taxation in a many-person economy. *Journal of Public Economics*, *11*, 143–157.
Domotor, Z. (1979). Ordered sum and tensor product of linear utility structures. *Theory and Decision*, *11*, 375–399.
Donaldson, D., & Weymark, J. A. (1980). A single-parameter generalization of the Gini indices of inequality. *Journal of Economic Theory*, *22*, 67–86.
Donaldson, D., & Weymark, J. A. (1988). Social choice in economic environments. *Journal of Economic Theory*, *46*, 291–308.
Feldstein, M. (1976). On the theory of tax reform. *Journal of Public Economics*, *6*, 77–104.
Fleming, M. (1952). A cardinal concept of welfare. *Quarterly Journal of Economics*, *66*, 366–384.
Gajdos, T., & Weymark, J. A. (2005). Multidimensional Generalized Gini indices. *Economic Theory*, *26*, 471–496.
Gibbard, A. (1973). Manipulation of voting schemes: A general result. *Econometrica*, *41*, 587–601.
Gibbard, A., Hylland, A., & Weymark, J. A. (1987). Arrow's Theorem with a fixed feasible alternative. *Social Choice and Welfare*, *4*, 105–115.
Guesnerie, R. (1977). On the direction of tax reform. *Journal of Public Economics*, *7*, 179–202.
Guesnerie, R. (1995). *A contribution to the pure theory of taxation*. Cambridge: Cambridge University Press.
Guesnerie, R., & Laffont, J.-J. (1978). Taxing price makers. *Journal of Economic Theory*, *19*, 423–455.
Guesnerie, R., & Seade, J. (1982). Nonlinear pricing in a finite economy. *Journal of Public Economics*, *17*, 157–179.
Hahn, F. H. (1973). On optimum taxation. *Journal of Economic Theory*, *6*, 96–106.
Haldane, J. B. S. (1955). Population genetics. In M. L. Johnson, M. Abercrombie, & G. E. Fogg (Eds.), *New biology 18* (pp. 34–51). Harmondsworth, UK: Penguin.
Hamilton, W. D. (1964). The genetical evolution of social behaviour. I and II. *Journal of Theoretical Biology*, *7*, 1–52.
Hammond, P. J. (1976). Equity, Arrow's conditions, and Rawls' difference principle. *Econometrica*, *44*, 793–804.
Harsanyi, J. C. (1953a). Cardinal utility in welfare economics and in the theory of risk-taking. *Journal of Political Economy*, *61*, 434–435.
Harsanyi, J. C. (1953b). Welfare economics of variable tastes. *Review of Economic Studies*, *21*, 204–213.
Harsanyi, J. C. (1955). Cardinal welfare, individualistic ethics, and interpersonal comparisons of utility. *Journal of Political Economy*, *63*, 309–321.
Harsanyi, J. C. (1975). Nonlinear social welfare functions: Do welfare economists have a special exemption from Bayesian rationality? *Theory and Decision*, *6*, 311–332.
Harsanyi, J. C. (1977). Non-linear social welfare functions: A rejoinder to Professor Sen. In R. E. Butts & J. Hintikka (Eds.), *Foundational problems in the special sciences* (pp. 293–296). Dordrecht: D. Reidel.
Heller, W. P., & Shell, K. (1974). On optimal taxation with costly administration. *American Economic Review, Papers and Proceedings*, *64*, 338–345.
Hellwig, M. (1986). The optimal linear income tax revisited. *Journal of Public Economics*, *31*, 163–179.
Inman, R. P. (1971). Towards an econometric model of local budgeting. *Proceedings of the Annual Conference of the National Tax Association*, *64*, 699–719.

Kolm, S.-C. (1998). *Justice and Equity*. Cambridge, MA: MIT Press. Originally published in French in 1972.

Krantz, D., Luce, R. D., Suppes, P., & Tversky, A. (1971). *Foundations of measurement, Volume I: Additive and polynomial representations*. New York: Academic Press.

Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: University of Chicago Press.

Kuhn, T. S. (1977). Objectivity, value judgment, and theory choice. In T. S. Kuhn (Ed.), *The essential tension* (pp. 320–339). Chicago: University of Chicago Press.

Laffond, G. (1980). *Révelation des preferences et utilités unimodales*. Thèse pour le doctorat, Laboratoire d'Econométrie, Conservatoire National des Arts et Métiers.

Le Breton, M., & Sen, A. (1999). Separable preferences, strategyproofness, and decomposability. *Econometrica*, *67*, 605–628.

Le Breton, M., & Weymark, J. A. (1996). An introduction to Arrovian social welfare functions on economic and political domains. In N. Schofield (Ed.), *Collective decision-making: Social choice and political economy* (pp. 25–61). Boston: Kluwer Academic Publishers.

Le Breton, M., & Weymark, J. A. (1999). Strategy-proof social choice with continuous separable preferences. *Journal of Mathematical Economics*, *32*, 47–85.

Le Breton, M., & Weymark, J. A. (2011). Arrovian social choice theory on economic domains. In K. J. Arrow, A. K. Sen, & K. Suzumura (Eds.), *Handbook of social choice and welfare Vol.1*, (pp. 191–291). Amsterdam: North-Holland.

Lerner, I. M. (1968). *Heredity, evolution and society*. San Francisco: W. H. Freeman.

Lollivier, S., & Rochet, J.-C. (1983). Bunching and second-order conditions: A note on optimal tax theory. *Journal of Economic Theory*, *31*, 392–400.

Luce, R. D., & Raiffa, H. (1957). *Games and decisions: Introduction and critical survey*. New York: Wiley.

Marschak, J. (1950). Rational behavior, uncertain prospects, and measurable utility. *Econometrica*, *18*, 111–141.

Maskin, E. S. (1976). *Social welfare functions for economics*. Unpublished manuscript, Darwin College, Cambridge University and Department of Economics, Harvard University.

Michod, R. E., Viossat, Y., Solari, C. A., Hurand, M., & Nedelcu, A. M. (2006). Life-history evolution and the origin of multicellularity. *Journal of Theoretical Biology*, *239*, 257–272.

Mirrlees, J. A. (1971). An exploration in the theory of optimum income taxation. *Review of Economic Studies*, *38*, 175–208.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics*, *6*, 327–358.

Morelli, M., Yang, H., & Ye, L. (2012). Competitive nonlinear taxation and constitutional choice. *American Economic Journal: Microeconomics*, *4*, 142–175.

Morreau, M. (2015). Theory choice and social choice: Kuhn vindicated. *Mind*, *493*, 239–262.

Morreau, M., & Weymark, J. A. (2016). Measurement scales and welfarist social choice. *Journal of Mathematical Psychology*, *75*, 127–136.

Moulin, H. (1980). On strategy-proofness and single peakedness. *Public Choice*, *35*, 437–455.

Okasha, S. (2006). *Evolution and the levels of selection*. Oxford: Oxford University Press.

Okasha, S. (2009). Individuals, groups, fitness and utility: Multi-level selection meets social choice theory. *Biology and Philosophy*, *24*, 561–584.

Okasha, S. (2011). Theory choice and social choice: Kuhn versus Arrow. *Mind*, *477*, 83–115.

Okasha, S. (2012). Social justice, genomic justice and the veil of ignorance: Harsanyi meets Mendel. *Economics and Philosophy*, *28*, 43–71.

Okasha, S., Weymark, J. A., & Bossert, W. (2014). Inclusive fitness maximization: An axiomatic approach. *Journal of Theoretical Biology*, *350*, 24–31.

Piketty, T. (1993). Implementation of first-best allocations via generalized tax schedules. *Journal of Economic Theory*, *61*, 23–41.

Quiggin, J. (1982). A theory of anticipated utility. *Journal of Economic Behavior and Organization*, *3*, 323–343.

Rádo, F., & Baker, J. A. (1987). Pexider's equation and aggregation of allocations. *Aequationes Mathematicae*, *32*, 227–239.

Rawls, J. (1958). Justice as fairness. *Philosophical Review*, *67*, 164–194.
Rawls, J. (1971). *A Theory of Justice*. Cambridge, MA: Harvard University Press.
Rescher, N. (1975). *Unselfishness: The role of the vicarious affects in moral philosophy and social theory*. Pittsburgh: University of Pittsburgh Press.
Risse, M. (2002). Harsanyi's 'utilitarian theorem' and utilitarianism. *Noûs*, *36*, 550–577.
Roberts, K. (1995). Valued opinions or opinionated values: The double aggregation problem. In K. Basu, P. Pattanaik, & K. Suzumura (Eds.), *Choice, welfare, and development: A festschrift in Honour of Amartya K. Sen* (pp. 141–165). Oxford: Oxford University Press.
Roberts, K. W. S. (1977). Voting over income tax schedules. *Journal of Public Economics*, *8*, 329–340.
Röell, A. (2012). *Voting over nonlinear income tax schedules*. Unpublished manuscript, School of International and Public Affairs, Columbia University.
Rothschild, M., & Stiglitz, J. E. (1970). Increasing risk: I. A definition. *Journal of Economic Theory*, *2*, 225–243.
Rubinstein, A. (1982). Perfect equilibria in a bargaining model. *Econometrica*, *50*, 97–109.
Samuelson, P. A. (1967). Arrow's mathematical politics. In S. Hook (Ed.), *Human values and economic policy: A symposium* (pp. 41–51). New York: New York University Press.
Satterthwaite, M. A. (1975). Strategy-proofness and Arrow's conditions: Existence and correspondence theorems for voting procedures and social welfare functions. *Journal of Economic Theory*, *10*, 187–217.
Sen, A. K. (1970). *Collective choice and social welfare*. San Francisco: Holden-Day.
Sen, A. K. (1973). *On economic inequality*. Oxford: Clarendon Press.
Sen, A. K. (1974). Informational bases of alternative welfare approaches: Aggregation and income distribution. *Journal of Public Economics*, *3*, 387–403.
Sen, A. K. (1976). Welfare inequalities and Rawlsian axiomatics. *Theory and Decision*, *7*, 243–262.
Sen, A. K. (1977). Non-linear social welfare functions: A reply to Professor Harsanyi. In R. E. Butts & J. Hintikka (Eds.), *Foundational problems in the special sciences* (pp. 297–302). Dordrecht: D. Reidel.
Simula, L. (2010). Optimal nonlinear income tax and nonlinear pricing: Optimality conditions and comparative static properties. *Social Choice and Welfare*, *35*, 199–220.
Stern, N. (1982). Optimum taxation with errors in administration. *Journal of Public Economics*, *17*, 181–211.
Stigler, G. J. (1964). A theory of oligopoly. *Journal of Political Economy*, *72*, 44–61.
Suppes, P. (1966). Some formal models of grading principles. *Synthese*, *6*, 284–306.
Tirole, J., & Guesnerie, R. (1981). Tax reform from the gradient projection viewpoint. *Journal of Public Economics*, *15*, 275–293.
Tulkens, H., & Zamir, S. (1979). Surplus-sharing local games in dynamic exchange economies. *Review of Economic Studies*, *46*, 305–313.
von Neumann, J., & Morgenstern, O. (1944). *Theory of games and economic behavior*. Princeton: Princeton University Press.
Weintraub, E. R. (1979). *Microfoundations: The compatibility of microeconomics and macroeconomics*. Cambridge: Cambridge University Press.
Weymark, J. A. (1978a). On Pareto-improving price changes. *Journal of Economic Theory*, *19*, 338–346.
Weymark, J. A. (1978b). 'Unselfishnness' and prisoner's dilemmas. *Philosophical Studies*, *34*, 417–425.
Weymark, J. A. (1979a). Optimality conditions for public and private goods. *Public Finance Quarterly*, *7*, 338–351.
Weymark, J. A. (1979b). A reconciliation of recent results in optimal taxation theory. *Journal of Public Economics*, *12*, 171–189.
Weymark, J. A. (1981a). Generalized Gini inequality indices. *Mathematical Social Sciences*, *1*, 409–430.

Weymark, J. A. (1981b). Undominated directions of tax reform. *Journal of Public Economics*, *16*, 343–369.
Weymark, J. A. (1986a). Bunching properties of optimal nonlinear income taxes. *Social Choice and Welfare*, *2*, 213–232.
Weymark, J. A. (1986b). A reduced-form optimal nonlinear income tax problem. *Journal of Public Economics*, *30*, 199–217.
Weymark, J. A. (1987). Comparative static properties of optimal nonlinear income taxes. *Econometrica*, *55*, 1165–1185.
Weymark, J. A. (1991). A reconsideration of the Harsanyi-Sen debate on utilitarianism. In J. Elster & J. E. Roemer (Eds.), *Interpersonal comparisons of well-being* (pp. 255–320). Cambridge: Cambridge University Press.
Weymark, J. A. (1999). Decomposable strategy-proof social choice functions. *Japanese Economic Review*, *50*, 343–355.
Weymark, J. A. (2004). Shared consumption: A technological analysis. *Annales d'Économie et de Statistique*, *75–76*, 175–195.
Weymark, J. A. (2005a). Measurement theory and the foundations of utilitarianism. *Social Choice and Welfare*, *25*, 527–555.
Weymark, J. A. (2005b). Shadow prices for a nonconvex public technology in the presence of private constant returns. In U. Schmidt, & S. Traub (Eds.), *Advances in public economics: Utility, choice, and welfare. A festschrift for Christian Seidl*, (pp. 61–71). Springer, Dordrecht.
Weymark, J. A. (2010). Alvorado lá no morro, que beleza. *Exame CEO*, *5*, 52–55.
Wibaut, S. (1989). *Tax reform in disequilibrium economies*. Cambridge: Cambridge University Press.