

THEORIE DU CHOIX SOCIAL ET ECONOMIE NORMATIVE

La théorie du choix social étudie les règles d'évaluation et de décision collectives en se plaçant au plus haut niveau d'abstraction et de formalisme. Fixée par Arrow dans une monographie célèbre de 1951, sa démarche consiste à formuler, puis à examiner, du point de vue de leur compatibilité logique éventuelle, différents principes qui paraissent s'imposer normativement à la collectivité, quelle qu'en soit la nature concrète (société, groupe, assemblée délibérative,...). Ses deux objets privilégiés sont, d'une part, les procédures de vote, d'autre part, la répartition sociale des ressources économiques. On l'envisagera ici principalement sous le deuxième aspect, qui la rattache étroitement à la problématique de l'économie normative.

La théorie économique contemporaine se propose d'analyser les interactions d'individus idéalement rationnels, qu'elle suppose confrontés à différentes possibilités de coopération, d'antagonisme ou d'indifférence réciproques. Elle comporte simultanément un volet descriptif et un volet normatif, qui s'articulent notamment grâce aux « théorèmes fondamentaux » liant la notion d'équilibre général et celle d'optimum de Pareto. Dès le début du 20ème siècle, une spécialité particulière, l'économie du bien-être, s'était donné pour objet l'évaluation des situations économiques et, singulièrement, des modalités de la répartition. La supériorité du point de vue arrowien explique que, dans l'après-guerre, l'économie du bien-être se soit effacée derrière la théorie du choix social. Mais à son tour, celle-ci a manifesté les limites inhérentes au niveau de généralité qu'elle adopte: largement par contrecoup, il s'est développé une spécialité nouvelle d'économie normative, qui vise à élaborer concrètement les principes de la juste répartition des biens. On présentera ce courant à la suite de ceux du choix social et de l'économie du bien-être.

Le théorème d'Arrow et l'analyse du choix social.

Dans *Social Choice and Individual Values* (1ère édition, 1951; 2ème édition corrigée, 1963), Arrow construit un objet théorique nouveau -- le choix social -- qui lui permet d'abstraire certaines propriétés communes au processus de la décision collective en

démocratie et au processus allocatif réalisé par le marché. La plurivocité de ce concept et, donc, de la notion technique associée -- la fonction de choix social --, en fait tout à la fois l'intérêt et la fragilité. Arrow soumet la fonction de choix social à une liste de contraintes normatives, qu'il présente comme universelles; de cette liste résultera finalement, au terme d'une brillante démonstration combinatoire, le fameux théorème d'impossibilité. Pour autant que les contraintes posées a priori paraîtront justifiées, le théorème atteindra le but que lui assignait Arrow: au-delà de l'ingéniosité mathématique, il désignera une difficulté structurelle, une limitation quasi-physique, de l'existence collective. Si, en revanche, les contraintes déclarées universelles ne l'étaient pas en réalité, si, notamment, il s'avérait qu'elles ne s'appliquent pas toutes ensemble aux deux classes d'interprétations recherchées, la politique et l'économique, le théorème pourrait demeurer un instrument technique appréciable, mais il perdrait l'essentiel de sa portée philosophique.

Par définition, la fonction de choix social relie R_1, \dots, R_n , les relations de préférence des n individus, à $R = f(R_1, \dots, R_n)$, la relation de préférence collective. Les relations de préférence individuelles sont toujours des préordres, c'est à dire des relations totales et transitives. Les propriétés additionnelles de la fonction f constituent autant d'axiomes, que l'ouvrage va examiner en détail, tant sur le plan normatif que sur le plan logique. L'axiome O impose à R d'être elle-même une relation de préordre, quel que soit le « profil » de préférences R_1, \dots, R_n . L'axiome faible de Pareto P requiert que si les n individus préfèrent strictement x à y , la collectivité préfère aussi strictement x à y . L'axiome d'indépendance I pose que, si deux profils de préférences R_1, \dots, R_n et R'_1, \dots, R'_n coïncident relativement à deux options x et y , les préférences collectives qu'ils induisent, $f(R_1, \dots, R_n)$ et $f(R'_1, \dots, R'_n)$, coïncideront à leur tour relativement à x et y . L'axiome de non-restriction U demande que f soit définie sur tous les profils de préférences concevables. Enfin, l'axiome de non-dictature D exclut qu'il y ait un individu i -- le « dictateur » -- tel que, pour tout profil de préférences considéré, et pour tout couple d'options x et y , la préférence stricte de i entre x et y soit automatiquement reproduite par la collectivité. Le théorème d'Arrow affirme que, pour un nombre d'options de choix supérieur ou égal à trois, aucune fonctionnelle ne réunit les propriétés O, P, I, U et D. Par commodité mathématique,

certaines manuels démontrent le résultat sous l'axiome fort de Pareto, P^+ , qui renforce ainsi P : si tous les individus préfèrent au moins faiblement x à y et que l'un d'entre eux, au moins, préfère strictement x à y , la collectivité préfère strictement x à y ; par ailleurs, si tous les individus sont indifférents entre x et y , la collectivité sera indifférente entre x et y .

Les axiomes d'Arrow reflètent certaines exigences du fonctionnement des assemblées démocratiques. Si l'on interprète les relations de préordre R_1, \dots, R_n comme les classements des motions ou des candidats par les votants, et la relation R comme l'effet du vote, les axiomes P ou P^+ deviennent des formalisations élémentaires du principe démocratique. L'axiome D paraît naturel dans cette interprétation politique du système arrovien. I peut se justifier par une commodité de procédure largement reconnue: pour départager un nombre d'options au moins égal à trois, il est banal que les assemblées les mettent aux voix deux par deux. En outre, ainsi que le sous-entend l'appellation anglaise, *independence of irrelevant alternatives*, I peut exprimer une forme de rationalité du choix collectif: lorsque celui-ci consiste à trancher entre deux options x et y , les préférences individuelles relativement à x et y semblent devoir constituer toute l'information nécessaire. De même, U peut recevoir une interprétation à la fois procédurale et normative: il convient que la méthode de décision collective s'applique à toutes les combinaisons possibles de votes. Quant à l'axiome O , il décrit une exigence souhaitable plutôt qu'une contrainte de définition ou qu'une norme obligatoire.

La conclusion du théorème, suivant laquelle O est incompatible avec P , I , U et D pris ensemble, vient généraliser des paradoxes depuis longtemps recensés par la « théorie des comités ». On sait notamment que, si une assemblée vote sur trois motions x , y et z , considérées deux à deux, elle peut, dans certaines configurations des votes, aboutir à un classement cyclique de x , y et z . Cette observation, qui remonte à Condorcet, tombe comme un cas particulier sous le théorème d'Arrow: en effet, la règle majoritaire vérifie les axiomes P , I , U et D ; le théorème prédit qu'elle doit violer la condition restante. La force de l'analyse arrovienne est qu'elle permet tout aussi facilement de retrouver certains paradoxes importants de la théorie

économique du bien-être et, ce faisant, d'en révéler la parenté inattendue avec celui de Condorcet.

L'optimum économique, les différentes interprétations de l'utilité et les comparaisons interpersonnelles de préférences.

L'économie du bien-être, et donc l'interprétation économique du théorème d'Arrow, reposent sur certaines notions préalables -- optimalité, préférence, utilité, comparaisons interpersonnelles -- que l'on rappellera brièvement. Depuis le début du siècle, la théorie économique générale, et cette branche particulière qu'on nomme économie du bien-être, *welfare economics*, privilégient un concept normatif particulier, celui d'optimum de Pareto: il désigne un état de l'économie x tel que les agents ne puissent s'accorder à lui préférer un autre état y ; plus précisément, tel que, pour tout autre état y , s'il se trouve un agent qui préfère strictement y à x , il doit s'en trouver un autre qui préfère strictement x à y . L'avantage de cette notion, qui paraît tout d'abord banale, est qu'elle suppose un minimum d'information sur les caractéristiques individuelles. La contrepartie de cette facilité d'utilisation est qu'elle débouche sur un classement seulement partiel des options: il resterait encore à départager les optimums de Pareto, qui, même pour un profil donné des préférences individuelles, peuvent être très nombreux.

A vrai dire, le *Manuel* (1907) de Pareto ne présente pas l'optimalité dans ces termes exactement, mais on peut considérer la différence entre la présentation de l'optimalité par Arrow, adoptée ici, et celle de Pareto lui-même, comme étant d'ordre secondaire. Le point important est que ces deux auteurs, représentatifs, en cela, de la théorie économique du 20^{ème} siècle, voient dans l'utilité un concept auxiliaire, à caractère purement technique. Pour eux, celui-ci désigne un indicateur numérique susceptible de représenter les propriétés qualitatives de la préférence, qu'ils comprennent comme une activité subjective d'évaluation, comme le principe d'explication non apparent des choix individuels. L'indicateur n'a pas d'existence en lui-même; les propriétés qu'il manifeste appartiennent toutes à cette faculté supposée ontologiquement première et suffisante, qu'est la faculté d'évaluer et de choisir. Suivant la conception archétypique du 20^{ème} siècle, il serait erroné de

chercher à rattacher l'utilité à une expérience psychologique originaire, ainsi que le faisaient les auteurs du 19^{ème} siècle. Il est significatif que la notion ne se perpétue le plus souvent, aujourd'hui, que dans l'expression figée de « fonction d'utilité ».

Une telle fonction sera dite ordinale si elle ne représente que l'aspect ordinal de la préférence, c'est-à-dire les propriétés de classement. Dans ce cas, elle ne peut être unique absolument, mais seulement à une transformation croissante près. Une fonction d'utilité sera dite cardinale si elle représente en outre les propriétés relatives à l'intensité de la préférence, celles que l'on nomme, justement, cardinales. Dans ce dernier cas, la fonction d'utilité est unique à une transformation affine croissante près: la restriction d'unicité est donc beaucoup plus forte que celle des fonctions ordinales. Supposons, par exemple, que la donnée de préférence consiste dans le classement décroissant des options: A-B-C, ainsi que dans l'indication suivante: la préférence pour A, par rapport à B, est plus intense que la préférence pour C, par rapport à D. Alors, les nombres 3-2-1, 35-25-15, 2000-1400-800... représentent correctement les données; ils appartiennent à une seule et même fonction d'utilité cardinale. En revanche, les nombres 35-0-(-10) ou 2000-19-5 représentent seulement un aspect de la préférence considérée; ils appartiennent à une seule et même fonction d'utilité ordinale, mais à des fonctions d'utilité cardinales différentes. Le problème qu'illustre cet exemple -- celui de l'existence d'une fonction d'utilité qui représenterait aussi l'aspect cardinal de la préférence individuelle -- ne se pose pas nécessairement en pratique, car la signification des données relatives à l'intensité des préférences est sujette à caution. Selon Pareto et ses disciples, l'observation des choix concrets de l'individu permettrait d'établir s'il préfère A à B, et B à C, mais non pas s'il préfère plus vivement A à B qu'il ne préfère B à C. Au moment où il écrit *Choix social et valeurs individuelles*, Arrow approuve entièrement cette analyse, et récuse donc le concept de cardinalité au profit de celui d'ordinalité.

Il importe de distinguer ces notions, qui s'appliquent aux fonctions d'utilité (ou aux préférences) individuelles, et la propriété qu'ont éventuellement ces fonctions (ou les préférences correspondantes) d'être comparables d'individu à individu. En bonne logique, quatre cas sont possibles: ordinalité et incomparabilité, cardinalité et incomparabilité, ordinalité et comparabilité, cardinalité et comparabilité.

L'axiomatique arrowienne du choix social doit figurer dans la case « ordinalité et incomparabilité ». D'une part, la formalisation de la préférence qu'elle adopte -- par de simples données de classement -- interdit aux propriétés cardinales de jouer un rôle quelconque. D'autre part, la dépendance entre l'individuel et le collectif -- telle que la décrit la définition de la « fonction de choix social », précisée par la condition I -- ne permet pas à la collectivité de tenir compte d'une forme quelconque de comparaison: même la modalité ordinale, qui semble la plus facile à admettre -- « x satisfait mieux la préférence de i que celle de j » -- demeure inexprimable dans le formalisme choisi par Arrow. Après avoir récusé les propriétés cardinales en général, celui-ci exclut donc toute possibilité de mise en rapport des préférences. En cela encore, il se rattache fidèlement à l'école parétienne: l'auteur du *Manuel* et ses disciples avaient déjà conclu que la satisfaction ne peut se comparer d'un individu à un autre. L'intérêt de la notion d'optimum de Pareto vient, très précisément, de ce qu'elle permet de définir le bien-être social sans requérir cette problématique comparaison.

L'utilitarisme du 19^{ème} siècle procédait de choix philosophiques fort différents. Bentham, John Stuart Mill, Jevons s'accordent sur ce principe général d'évaluation: le meilleur état de la société est celui qui réalise « le plus grand bonheur du plus grand nombre », c'est-à-dire, dans un langage plus technique, celui qui « maximise la somme des utilités individuelles ». Les utilitaristes ne s'accordent sans doute pas sur le contenu à donner à la notion d'utilité. Bentham y voit primordialement une mesure relative du plaisir et du déplaisir, qu'il veut envisager suivant leurs différentes qualités psychologiques élémentaires, principalement l'intensité, la durée, la certitude et la proximité. Cette conception à la fois simpliste -- par son hédonisme avoué --, et raffinée -- puisqu'elle débouche sur une représentation multi-dimensionnelle de l'utilité --, s'édulcore chez Mill ou Jevons. Mais, quelles que soient leurs différences philosophiques, les utilitaristes du 19^{ème} considèrent tous le concept comme découlant d'une psychologie du sentiment; et ils en font le terme primitif de leurs systèmes d'éthique et de de politique. On chercherait en vain chez eux une doctrine élaborée de la préférence. Pour autant qu'ils l'analysent également, l'activité de choix et d'évaluation ne revêt qu'un caractère logiquement second: c'est

parce que l'utilité d'un acte excède celle d'un autre que l'individu choisit d'accomplir le premier. Une telle formule est à comparer avec le point de vue moderne, qui renverse l'ordre hiérarchique des concepts et évite de les relier causalement (les valeurs prises par la fonction d'utilité « représentent » les propriétés de la préférence, elles ne sont pas causées par elle). Enfin, et pour autant qu'on néglige les complications introduites par la conception multi-dimensionnelle de Bentham, les utilitaristes s'accordent nécessairement pour affirmer que les utilités individuelles sont à la fois cardinales et comparables entre elles. En effet, la « maximisation de la somme des utilités individuelles » n'a de sens qu'à la condition suivante: il faut que la collectivité puisse mettre en balance le gain d'utilité que représenterait pour l'individu i le remplacement de l'option x par l'option y , avec la perte d'utilité que celui-ci représenterait éventuellement pour un autre individu j .

Dès le 19^{ème} siècle, on a critiqué la doctrine dans les termes que Rawls devait employer bien plus tard: on a souligné que l'utilitarisme pouvait justifier le sacrifice intégral d'une fraction de la société à une autre, au motif que les gains de la première l'emporteraient largement sur les pertes de la seconde. Cette conséquence extrême, inattendue, d'une doctrine qui revêt tout d'abord l'apparence neutre d'une formule de calcul, figure, énoncée en termes inquiets, dans une lettre prémonitoire de Francis Place à Bentham. La réponse que les benthamiens, plutôt que Bentham lui-même, proposeront bientôt à cette difficulté, consiste à mettre en avant le principe des utilités marginales décroissantes: les transferts de ressources à l'intérieur de la société seront toujours limités, puisque le gain incrémental du groupe avantagé par le transfert diminue constamment, alors que la perte du groupe désavantagé augmente constamment. Ainsi combiné, par Jevons et bien d'autres, à l'un des principes les plus célèbres de l'économie néo-classique, l'utilitarisme perd l'aspect inquiétant que relevait Place. S'il a irrité les doctrinaires de la fin du 19^{ème} siècle, ce n'est pas du tout parce qu'il cautionnerait une morale du sacrifice, mais, à l'opposé en quelque sorte, parce qu'il impliquerait une égalité rigide de la répartition. Cette nouvelle conséquence, peut-être surprenante, de la formule de maximisation benthamienne, résulte de la démonstration suivante: supposons que les fonctions d'utilité des individus, variant avec leur revenu, soient toutes identiques et

présentent la propriété des utilités marginales décroissantes; alors la somme de ces fonctions atteint sa valeur maximale au point d'égle répartition du revenu.

On voit donc que l'utilitarisme est, politiquement et éthiquement, une doctrine ambivalente: elle est relativement peu déterminée tant que l'on ne précise pas la nature du problème traité, et les hypothèses qui en découlent sur les fonctions d'utilité individuelles. Du point de vue de la précision des conséquences, l'utilitarisme n'en reste pas moins très supérieur au calcul parétien. Toute solution utilitariste est aussi un optimum de Pareto, mais non l'inverse; la règle de Bentham permet donc de restreindre l'ensemble des optimums. La précision supérieure de la règle provient de ce qu'elle incorpore le principe « cardinalité et comparabilité » que récuse, justement, les parétiens, parmi lesquels il faut compter l'auteur de *Choix Social et Valeurs Individuelles*.

Forces et faiblesses de l'économie du bien-être; retour au théorème d'impossibilité.

Avant le théorème d'Arrow, les recherches d'économie du bien-être devaient culminer dans les deux résultats que les manuels de microéconomie appellent « théorèmes fondamentaux de l'économie du bien-être ». Le premier de ces énoncés affirme que tout équilibre général de l'économie est un optimum de Pareto. (Un équilibre général est défini comme un état des prix et des demandes en biens et services tel que l'offre et la demande s'égalisent sur chaque marché.) Plus technique et surprenant, le second énoncé affirme que tout optimum de Pareto peut s'obtenir comme équilibre général d'une économie dont les ressources initiales -- c'est-à-dire les dotations des agents antérieurement à la production et à l'échange -- seraient réparties de façon appropriée. Ces deux théorèmes ont des conséquences importantes en philosophie politique. Ils reprennent et clarifient l'idée smithienne de la main invisible: alors que la *Richesse des Nations* entremêle constamment les considérations d'équilibre et d'optimalité, la théorie moderne distingue soigneusement la définition de ces deux concepts avant d'en expliquer les relations logiques, au travers, notamment, des deux théorèmes.

Si importants que soient ces résultats pour l'analyse d'institutions idéales, il est clair qu'ils n'épuisent pas le contenu d'une théorie économique de l'intérêt général. Car il faut aussi évaluer les conséquences de changements particuliers affectant l'économie, et, tout particulièrement, comparer entre elles différentes mesures de politique économique: la suppression des barrières douanières n'a-t-elle que des effets bénéfiques à long terme, comme le prétendent Ricardo et l'école libérale? Comment se mesurent les effets redistributifs, et finalement les conséquences en matière de bien-être, d'un impôt indirect proportionnel et d'une taxe forfaitaire? L'examen de ces deux questions concrètes suppose que soient précisées les normes générales de l'évaluation collective; elles tombent donc dans le champ de l'économie du bien-être, tout aussi bien que les considérations, d'essence théorique, auxquelles obéissent les deux « théorèmes fondamentaux ». Les travaux de Marshall, puis ceux de Pigou dans son *Economics of Welfare* (1920), s'efforçaient justement de répondre aux interrogations naturelles du public et des hommes politiques en matière d'évaluation collective.

Mais les auteurs des années 1930 devaient contester les techniques introduites par les marshalliens, au motif qu'elles excèdent généralement l'ordinalisme parétien. La « nouvelle économie du bien-être », comme on en vint à la nommer alors, prétendait ramener toute la partie normative de la science économique à la seule considération des optimums de Pareto. Pour surmonter l'inconvénient majeur du calcul parétien -- la partialité du classement --, la *new welfare economics* comptait sur le « principe de compensation »: on dira qu'un optimum de Pareto x est socialement préférable à un autre y si les individus qui gagneraient au remplacement de y par x peuvent, en x , indemniser les autres de toute leur perte et néanmoins conserver une partie de leurs gains. Cependant, malgré les efforts poursuivis par Hicks, Kaldor et Scitovsky, le principe a résisté à toute élaboration rigoureuse. On doit à Arrow d'avoir fait comprendre pourquoi ces tentatives étaient logiquement condamnées à l'échec. Lorsqu'on reformule la recherche des optimums dans le langage arrowien, on obtient une fonction de choix social qui vérifie les conditions U, I, D, ainsi, naturellement, que P et même P+. Cette fonction satisfait à l'une des conditions contenues dans O -- la transitivité -- mais, par nature, viole la condition restante --- celle de totalité. Le

théorème implique alors cette conséquence: tout renforcement du critère qui le rendrait total viole nécessairement la transitivité. De fait, toutes les propositions que l'on a faites, dans les années 1930, pour comparer les optimums de Pareto à l'aide du principe de compensation, ont débouché sur des exemples de cycles, exactement comme en théorie du vote. En économie comme en politique, la force du théorème d'Arrow vient de ce qu'il révèle une impossibilité de principe derrière des difficultés que l'on avait tout d'abord appréhendées cas par cas.

L'importance de cette démonstration négative dépasse la réfutation des tentatives particulières de la *new welfare economics*. L'économiste d'aujourd'hui aura retenu de *Choix Social et Valeurs Individuelles* une leçon durable: il sait que le parétianisme ne constitue pas une base logiquement suffisante pour élaborer la théorie économique du bien-être. Le vrai paradoxe du théorème d'impossibilité tient peut-être finalement à ceci: il implique une réduction à l'absurde de la doctrine à laquelle, en 1951, Arrow lui-même disait adhérer.

Les développements de la théorie du choix social.

La théorie est née du théorème: ses premiers travaux ont consisté à parcourir les différentes solutions concevables au problème, à la fois logique et normatif, soulevé par celui-ci. Un premier groupe d'auteurs a contesté la définition de la fonction de choix social et l'axiome I, qui font étroitement système. (En effet, Arrow a défini sa fonction f d'une manière qui permette à la collectivité de considérer simultanément plusieurs profils de préférences; mais cette liaison des profils ne devient effective que grâce à l'axiome I.) Dans les années 1950, A. Bergson, Little et Samuelson ont nié que les conditions « interprofils » s'appliquent valablement aux problèmes traités par l'économie du bien-être: celle-ci raisonne sur une population fixée d'individus, dont chacun s'identifie à une relation de préférence donnée une fois pour toutes. Les mêmes auteurs reconnaissent que les conditions litigieuses peuvent avoir un sens pour l'analyse politique, s'il est vrai que le problème formalisé par la fonction f concerne la mise en place d'une *constitution* (dont on attend évidemment qu'elle s'impose à des populations multiples). La critique revient donc à contester

l'identification originale, chez Arrow, des procédures d'allocation économique et de vote.

Elle explique le développement d'une théorie du choix social à « profils simples »; mais celle-ci, d'une certaine façon, a répondu aux objections de Samuelson et de Little dans un sens favorable à Arrow. Pollack, Rubinstein et d'autres ont examiné les équivalences éventuelles entre l'analyse à « profils multiples » et à « profils simples »: il est apparu que, si l'on introduit un nouvel axiome qui joue le rôle de I, un profil unique comportant des possibilités combinatoires suffisamment nombreuses suffit à l'obtention des résultats souhaités, en particulier de la conclusion d'impossibilité. De telles recherches interagissent avec celles qui visent en priorité à affaiblir l'axiome U et qui privilégient l'hypothèse d'une sous-classe donnée de profils (« domaine économique »). La préoccupation qui guide ce groupe de travaux, auxquels participent notamment Border, Le Breton ou Weymark, n'est pas très différente de celle qu'exprimaient les premiers adversaires de la fonction de choix social. Une application spécifiquement économique du théorème d'impossibilité supposerait en effet qu'on le démontre pour un ensemble d'options bien particulier, typiquement, les « paniers de consommation » de la théorie microéconomique, et pour des préférences elles-mêmes particulières, parce que déterminées également par cette théorie. On peut vouloir se limiter à des préférences égoïstes (c'est-à-dire affectées par la seule consommation de l'individu considéré), monotones (c'est-à-dire croissantes avec la consommation), continues, convexes (cette dernière propriété correspond techniquement à celle des utilités marginales décroissantes). La question est alors de savoir si l'on peut reproduire la démonstration d'impossibilité en restreignant f au sous-domaine défini à l'aide de telle ou telle de ces considérations microéconomiques. Comme la réponse est souvent affirmative, le point de vue initial d'Arrow paraît, ici encore, avoir été conforté.

Plus important conceptuellement, et beaucoup plus problématique, le parétianisme sous-jacent au théorème d'impossibilité prêtait le flanc à la critique. Réhabilitée brillamment par Sen, l'analyse des comparaisons interpersonnelles d'utilité constitue sans doute le développement théorique principal des années 1970 et du début des

années 1980. Les auteurs qui y ont participé -- Sen lui-même dans *Collective Choice and Social Welfare* (1970), d'Aspremont, Gevers, Hammond, Roberts, Blackorby et bien d'autres encore -- ont sauté le pas que Arrow n'avait pas voulu franchir en 1951. Ils ont exploité, par contraposition, un passage dans lequel celui-ci paraissait désigner l'hypothèse responsable de la conclusion d'impossibilité: « si nous excluons la possibilité de comparaisons interpersonnelles d'utilité, alors la seule manière de passer des goûts individuels aux préférences sociales qui soit satisfaisante, et qui s'applique à un large ensemble de n-uples de préférences individuelles, est ou bien imposée ou bien dictatoriale » (V,4). Il ne s'agissait pas pour autant de cautionner l'utilitarisme, mais de replacer les différentes méthodes d'évaluation collectives dans un même cadre rationnel, afin d'en préparer l'appréciation normative.

Plus précisément, Sen et les auteurs de ce courant introduisent une « fonctionnelle de choix social », qui est définie non plus sur les profils de préférences, mais sur les profils de fonctions d'utilité, puis ils proposent de partitionner le nouveau domaine de définition de différentes manières, en posant que la fonctionnelle sera invariante relativement à la partition choisie. Les hypothèses sur les partitions concernent les quatre cas de figure déjà présentés « ordinalité et incomparabilité », « cardinalité et incomparabilité », etc., ainsi que d'autres que l'on n'a pas encore envisagés ici. Ce formalisme permet donc de retrouver l'hypothèse d'Arrow comme cas particulier d'une analyse plus générale, qui est compatible avec toutes les modalités concevables de comparabilité et d'incomparabilité des préférences. La partition arrowienne, « ordinalité et incomparabilité », est la moins fine de toutes: à un profil d'utilités donné, elle identifie tout autre profil dont les utilités se déduisent des précédentes par des transformations croissantes quelconques. La partition la plus fine de toutes est celle qui distingue tous les profils d'utilités entre eux: c'est la partition dont les cellules sont des singletons; elle s'interprète en disant que, du point de vue collectif, toutes les propriétés des indicateurs numériques sont, d'une part, pertinentes du point de vue de la préférence individuelle, d'autre part, comparables d'individu à individu. Entre ces deux extrêmes se déploie une combinatoire que les théoriciens des années 1970-1980 ont examinée à la lumière d'axiomes souvent inspirés de ceux d'Arrow, mais qu'ils reformulaient dans le

langage de la fonctionnelle de choix social. Par exemple, Sen a fait voir que, sous ces axiomes additionnels, le choix de la partition « cardinalité et incomparabilité » ramène immédiatement à la conclusion du théorème d'impossibilité. D'autres auteurs ont montré comment la configuration « cardinalité et comparabilité », qui, on l'a vu, est impliquée par l'utilitarisme, permet, inversement, de le caractériser, compte tenu des autres axiomes. De même, on s'est penché sur le cas « ordinalité et comparabilité »: il est vérifié par une règle égalitariste, celle du *maximin*., et s'avère même équivalent à elle, si l'on admet certains autres axiomes. Cette règle consiste à choisir entre deux options x et y suivant que x ou y apporte plus d'utilité à l'individu qui, lorsque chacune de ces options est retenue, apparaît comme le plus désavantagé de tous. Elle s'apparente à celle que Rawls défend dans *A Theory of Justice* (1971), à cette réserve près -- philosophiquement importante -- que Rawls apprécie l'avantage relatif d'un individu par rapport à un autre non pas en termes d'utilité, mais suivant la possession de « biens fondamentaux » pris dans une liste limitative et dotés d'un caractère objectif. La règle du maximin a donné lieu à une généralisation itérative appelée *leximin* : en cas d'égalité des utilités les plus basses, on considère celles qui, pour chaque option, leur sont immédiatement supérieures, et ainsi de suite.

Dans le jargon du choix social, on appelle « base informationnelle » la partition qui est sous-jacente à une méthode d'évaluation donnée. On a contesté cette terminologie: elle met en relief la capacité de discernement de la collectivité, comme dans la théorie microéconomique de l'information, au risque de dissimuler malencontreusement le rôle que jouent les choix éthiques dans l'adhésion à tel ou tel type de règle. On a adressé des objections plus radicales à la théorie des années 1970-1980: l'axiomatisation des « bases informationnelles » aura-t-elle fait réellement progresser la compréhension des doctrines les plus célèbres, comme l'utilitarisme et l'égalitarisme? Sans doute a-t-on trop attendu de la démarche purement formalisante de ces travaux. Mais certains d'entre eux ont livré des conclusions qui paraissent à la fois importantes et surprenantes. Un des résultats de la théorie s'énonce ainsi: supposons que la « fonctionnelle de choix social » satisfasse l'axiome d'indépendance à l'égard des options non-pertinentes, celui de Pareto, un axiome de symétrie entre les individus, appelé « anonymat », un axiome de séparabilité -- dont l'effet est

simplement d'exclure de la décision collective les individus qui y seraient indifférents --; alors, supposant encore un axiome général (c'est-à-dire non seulement ordinal, mais aussi cardinal) de comparaison interpersonnelle des utilités, il s'avère que les deux règles du leximin et de l'utilitarisme constituent une alternative *pratiquement exclusive* de toute autre possibilité. Un tel résultat contribue à clarifier les présupposés sur lesquels s'affrontent, dans la philosophie anglo-américaine contemporaine, d'une part, les défenseurs d'une tradition utilitariste toujours vivante, d'autre part, les néo-égalitaristes qui se reconnaissent dans l'oeuvre de Rawls.

Théories de l'état originaire et modèles de négociation.

Les résultats qu'on vient de restituer n'ont pas de caractère constructif: les équivalences logiques qu'ils exposent ne correspondent aucunement à une genèse, même idéale, des règles de choix social. On peut concevoir que l'on passe des propriétés normatives aux règles suivant une autre méthode, éventuellement formalisée et même axiomatisable, comme l'était la précédente, mais qui serait constructive: au lieu d'agréger de l'extérieur, en quelque sorte, les caractéristiques pertinentes des individus (typiquement, leurs préférences ou leurs utilités), on ferait découler l'évaluation collective d'un processus qui implique activement ces individus, qui transforment les caractéristiques retenues en facteur causal. Les théories dites de l'état originaire (*original position*) correspondent à cette description générale. Elles ressortissent à la philosophie politique tout autant, si ce n'est plus, qu'à l'économie normative; elles font d'ailleurs écho aux théories de l'état de nature élaborées par les juristes et les philosophes des 17^{ème} et 18^{ème} siècles. Dans une certaine mesure aussi, les modèles de négociation (*bargaining*) répondent à une préoccupation de dérivation génétique, plutôt qu'exclusivement formelle, des règles collectives.

Ces constructions ont encore ceci de commun qu'elles s'inscrivent dans la problématique contemporaine de la justice distributive. En effet, si l'on amplifie une idée commune à Nozick et à Rawls, la justice doit trouver son sens plein dans l'élaboration de procédures convenables: le caractère juste ou injuste d'un état de

choses reflète primordialement les propriétés éthiques de la méthode qui conduit à cet état. Une telle affirmation ne vaut pas nécessairement pour la forme de justice qu'Aristote appelait « corrective », par opposition à la forme « distributive ». Mais si l'on admet cette limitation, on disposera, grâce à la conception procédurale, d'une manière d'expliquer la justice: suivant les auteurs, l'explicans se nommera « impartialité » ou « avantage mutuel », ainsi qu'on va le voir.

Le « voile d'ignorance » constitue l'une des genèses les plus fréquemment invoquées, aujourd'hui, pour les règles d'évaluation collective; il intervient notamment dans la formulation rawlsienne de l'égalitarisme (le célèbre « principe de différence »). On suppose que les membres de la société sont doués de raison et dotés d'un savoir général, mais ignorent certains traits fondamentaux de leur identité personnelle. Dans ces conditions idéales, l'évaluation d'un état de la répartition ne peut être entachée de favoritisme ou d'égoïsme. Toute décision prise sous le voile d'ignorance doit être déclarée juste, en ce sens particulier, tout au moins, où la justice désigne l'impartialité de traitement des individus. Les décisions seront adoptées à l'unanimité, parce que -- ce sont là deux hypothèses cruciales -- tous appliquent le même critère de choix dans l'incertitude et tous disposent exactement des mêmes informations partielles. Avant que Rawls ne lui donne corps, l'idée d'une décision en situation d'ignorance relative appartient au théoricien des jeux et philosophe Harsanyi, qui en esquisse l'application dès 1953. Par la suite, Harsanyi la précisera en s'appuyant sur la conception de la « préférence fondamentale » (que Sen et Kolm ont également invoquée, mais pour d'autres fins): les individus placés sous le voile évaluent les états de la répartition à l'aide d'une seule et même relation de préférence, qui permet de comparer deux états x et y en se plaçant successivement au point de vue de n'importe quel membre de la société i, j, k, \dots . C'est à cette préférence fondamentale que Harsanyi va appliquer la théorie générale du choix rationnel dans l'incertitude. Ce faisant, il obtient une formulation discutable, mais analytiquement claire, de la notion d'impartialité qui peut fonder celle de justice distributive.

Les conclusions de Harsanyi s'opposent radicalement à celles de Rawls: alors que celui-ci aboutissait à la règle du maximin, celui-là obtient une variante particulière

de l'utilitarisme, qui recommande de maximiser la moyenne des utilités individuelles. Au-delà de différences relatives à l'information maintenue ou supprimée par le voile, la raison de ce désaccord tient au choix du critère de décision: Rawls suppose que des individus ignorants de leur identité feront preuve d'une extrême prudence; ils seront ainsi conduits à préserver les intérêts du plus mal loti qu'ils peuvent incarner sans le savoir. En revanche, Harsanyi suppose qu'ils se trouvent face à une loterie donnant une chance égale d'avoir l'identité de n'importe quel individu de la population: si l'on applique à cette loterie particulière le critère de décision proposé par von Neumann et Morgenstern, celui de l'espérance d'utilité, on obtient immédiatement la règle utilitariste de la moyenne.

Le débat entre Rawls et Harsanyi aura été peu fructueux, en partie peut-être, parce qu'il recouvrait une équivoque: l'égalité dont il est question dans le « principe de différence » porte sur les consommations de certains biens, et n'est donc pas à mettre logiquement sur le même plan que la règle de la moyenne ou de la somme des utilités. En rappelant l'historique de la doctrine, on a d'ailleurs signalé que, sous certaines hypothèses additionnelles, l'utilitarisme cautionnait des conclusions égalitaristes. Une polémique d'un autre ordre, conceptuellement plus intéressante, a opposé Harsanyi et Sen. Pour celui-ci, un raisonnement fondé sur le critère de von Neumann-Morgenstern ne parviendrait pas à dériver l'utilitarisme proprement dit, mais seulement une règle de décision qui se représente, mathématiquement, par une fonction d'utilité sociale additivement séparable, sans que cette fonction ait d'interprétation claire, utilitariste ou autre, pour la théorie de l'évaluation collective. Sen et d'autres auteurs ont étendu cette objection technique à un travail important de Harsanyi, datant de 1955, qui se proposait de dériver à nouveau l'utilitarisme à partir du critère de von Neumann-Morgenstern, mais en remplaçant le procédé du voile d'ignorance par une condition d'unanimité de type parétien.

L'analyse axiomatique de la négociation se rapproche de celle de la « fonctionnelle de choix social » exposée précédemment. Elle comporte à la fois plus et moins d'informations que cette dernière approche: d'un côté, elle précise la situation de choix en y faisant figurer un « point de statu quo »; d'un autre côté, elle en appauvrit la description en n'explicitant pas l'ensemble des options sur lequel porte la

préférence collective. C'est ainsi qu'on définira un « problème de négociation » comme un ensemble de niveaux d'utilités que les individus -- au nombre de deux dans le modèle de base -- peuvent atteindre selon les accords qu'ils passent entre eux. On s'intéresse alors aux règles qui, dans un problème donné, permettent de sélectionner des vecteurs particuliers d'utilité: comme ces vecteurs modélisent l'issue de la négociation, les règles s'appellent « solutions ». Parmi les plus marquantes, celle de Nash consiste à maximiser le produit des gains d'utilité des agents (le gain d'utilité étant la différence d'utilité entre la situation finale et le statu quo ante, qui, par hypothèse, est maintenu en cas d'échec de la négociation). Celle de Kalai-Smorodinsky consiste à égaliser les sacrifices relatifs des agents (le sacrifice relatif est la fraction d'utilité perdue par rapport au gain potentiel maximal que l'agent obtiendrait si les autres réalisaient un gain nul). La théorie s'efforce de redécrire axiomatiquement les différentes solutions avancées. Exactement comme en théorie du choix social, les caractérisations ainsi obtenues permettront de décomposer les jugements de valeur sous-jacents à l'adoption d'une règle donnée.

La solution de Kalai-Smorodinski inspire directement l'interprétation de la justice défendue par Gauthier dans *Morals by Agreement* (1986). De même que la négociation peut poser des questions éthiques, certains auteurs voient en retour, dans la négociation, le modèle procédural de toute justice distributive. Une telle conception identifie la recherche d'une allocation équitable et celle d'un arrangement à l'amiable. La « justice comme avantage mutuel », pour reprendre l'expression de Barry, consacre les avantages et les handicaps naturels des individus; sous cet angle, elle est absolument antagonique de la « justice comme impartialité », dont le procédé caractéristique, le voile d'ignorance, permet de corriger le donné naturel, ou, plus exactement, certains de ses effets. Les deux conceptions présentent toutefois un caractère commun: elles illustrent, chacune à sa manière, le thème de la « justice comme exercice du choix rationnel ». Naturellement, les critères de rationalité, aussi bien que les situations hypothétiques auxquelles on les applique, varient considérablement de l'une à l'autre: Rawls et Harsanyi mobilisent la théorie des choix individuels face à la nature, Gauthier celle des jeux coopératifs. Sans doute le point de vue procédural est-il plus manifeste chez celui-ci que chez ceux-là: les

solutions d'un problème de négociation dépendent toujours des situations empiriques de départ (ici décrites par le point de statu quo et les possibilités de gains des différents agents).

Travaux récents d'économie normative.

Les constructions qui viennent d'être examinées conservent un caractère abstrait. Bien qu'elles interviennent dans une théorie de la justice distributive plutôt que du choix social pris généralement, elles reproduisent un trait caractéristique de la manière d'Arrow: leurs auteurs ne précisent pas toujours sur quels objets porte l'évaluation collective -- états de la répartition ou formes générales des institutions qui l'influencent, allocations de tous les biens ou de certains d'entre eux seulement, allocations destinées à la consommation ou destinées à la production; les précisions que l'on trouve chez certains -- par exemple Rawls -- demeurent extérieures à la modélisation retenue. En cela, les applications du voile d'ignorance ou des schémas de négociation s'opposent aux travaux les plus récents, qui s'efforcent de définir la justice dans des contextes économiques bien spécifiés. Ces travaux ont formulé des critères inédits de justice distributive (ou d'« équité », suivant le terme qui tend à s'imposer aujourd'hui), parmi lesquels on mentionnera notamment: le principe de « non-envie » (Foley, Kolm, Varian), les bornes inférieures ou supérieures de bien-être (Moulin), la solidarité des membres de la société face aux modifications des paramètres économiques ou de la population (Moulin, Thomson), la compensation des handicaps et des inégalités de talents (Dworkin, Roemer, Fleurbaey). Une fois axiomatisés, les critères figurent soit dans des théorèmes d'impossibilité d'un genre nouveau, soit dans la caractérisation de règles allocatives particulières. Celles-ci se regroupent dans deux familles principales: d'une part, les traditionnelles règles de marché, suivant lesquelles la collectivité alloue un budget à chaque individu, et laisse la répartition ultime des biens dépendre des prix obtenus à l'équilibre des offres et des demandes; d'autre part, les règles du type « équivalent-égalitaire », introduites par Pazner et Schmeidler, suivant lesquelles une allocation physique doit être un optimum de Pareto et, simultanément, convenir à chaque individu au moins autant que s'il pouvait choisir lui-même son panier de biens dans un certain

ensemble de référence universel. Plutôt que de détailler une liste, inévitablement incomplète, d'apports récents, on signalera quelques-uns de leurs traits communs, afin d'aider à les situer par rapport aux courants antérieurs.

Les restrictions de domaine que privilégie l'économie normative contemporaine évoquent celles que l'on rencontrait déjà en théorie du choix social: suivant une hypothèse courante, par exemple, les fonctions d'utilité varieront exclusivement, et de façon monotone, avec la consommation de l'individu considéré. Mais au lieu de chercher à classer l'ensemble des allocations envisagées, comme on l'aurait fait dans la théorie des « domaines économiques », on se contentera le plus souvent de définir un sous-ensemble qui satisfasse au critère désiré. Par ailleurs, l'axiome le plus représentatif de l'analyse arrowienne, l'indépendance à l'égard des options non pertinentes, est violé: il apparaît le plus souvent impossible de séparer une allocation équitable d'une autre qui ne l'est pas sur la seule base des préférences relatives à ces deux allocations; les préférences relatives à d'autres allocations réalisables -- voire simplement virtuelles, comme dans la règle de l'équivalent-égalitaire -- sont susceptibles d'influencer le jugement d'équité. En revanche, l'économie normative rejoint Arrow en ceci qu'elle n'envisage guère que des préférences ordinales. Enfin, les règles d'allocation vaudront non pas pour différents profils de préférences ou d'utilités individuelles, comme il en allait dans la théorie de la fonction ou de la fonctionnelle de choix social, mais pour différentes « économies »: ce concept inclut sans doute la donnée des utilités ou des préférences, mais aussi d'autres indications, à caractère objectif, comme la taille de la population, le nombre des biens, la quantité de ressources et la nature des techniques de production, etc.

Les auteurs anglo-saxons appellent *welfarism* la thèse suivant laquelle la satisfaction subjective ou bien-être -- que l'on mesure à l'aide de la relation de préférence individuelle ou de la fonction d'utilité correspondante -- représente la donnée principale, voire exclusive, pour une appréciation normative des règles collectives. On voit, par le dernier trait mentionné, que l'économie normative contemporaine est globalement dégagée du « welfarisme », voire hostile à ce postulat. Par là, elle témoigne d'une attitude caractéristique des années 1980-1990: la dénonciation du « welfarisme » rassemble aujourd'hui des auteurs aussi différents que Rawls,

Dworkin, Sen, Cohen ou Roemer. La polémique vise directement la conception du choix social des années 1970, celle qui correspondait à la première manière de Sen. Elle se développe notamment -- mais non exclusivement -- à l'occasion du débat lancé par Sen en 1980, « Egalité de quoi? ». Si l'on excepte celle du philosophe Arneson, les réponses données à cette question ne font jouer qu'un rôle subalterne aux préférences ou aux fonctions d'utilité, quand elles n'éliminent pas complètement ces concepts.

Parmi les arguments dirigés contre le « welfarisme », certains concernent les propriétés, normativement pertinentes, dont la fonction d'utilité individuelle ne parvient pas à rendre compte. On mentionne au premier chef les droits fondamentaux de la personne, ce qui ne surprendra pas l'historien des idées. Suivant une tradition qui, dès les Révolutions française et américaine, s'est affrontée au « principe d'utilité » -- Halévy expose admirablement ce conflit dans *La Formation du Radicalisme Philosophique* --, les droits fondamentaux constituent une réalité première du monde juridique et moral: ils existent indépendamment de l'utilité (au sens ordinaire aussi bien que technique) qu'ils représentent pour leurs titulaires. Avec Sen, on mentionne également les réalisations (*functionings*) ou les capacités objectives (*capabilities*), parmi les traits individuels qui échapperaient à une réduction conceptuellement satisfaisante aux fonctions d'utilité.

A l'opposé, en quelque sorte, on peut vouloir déclarer éthiquement non pertinents certains facteurs qui influencent la satisfaction des préférences ou les valeurs des fonctions utilités; ces critiques atteindront souvent le principe de Pareto en même temps que les règles « welfaristes ». Les préférences perverses ou malveillantes posent un problème délicat; plus généralement, ainsi que certains utilitaristes l'avaient compris dès le 19ème siècle, il semble que toutes les considérations « extra-personnelles » qui affectent les valeurs d'utilité puissent créer des difficultés au « welfarisme », y compris lorsqu'elles sont inspirées par la bienveillance. Dans un autre ordre, les préférences tendent naturellement à s'adapter aux potentialités; si on ne les corrige pas pour tenir compte de ce facteur d'adaptation, les règles exclusivement fondées sur la considération des utilités ou des préférences pourront légitimer abusivement certaines inégalités de la répartition. Enfin, on a souligné que

les préférences des individus répercutent des évaluations en même temps que des croyances, des jugements de fait aussi bien que des jugements de valeurs. Or les travaux de Hammond, Broome, Mongin, ont bien mis en évidence la fragilité du principe de Pareto -- et par là des règles « welfaristes » qui le mettent en oeuvre, comme l'utilitarisme -- lorsque les jugements de préférence individuels incorporent des probabilités subjectives différentes.

Ni les distinctions internes à l'économie normative contemporaine, ni les arguments qui la rassemblent polémiquement contre les théories précédentes du choix social ou du bien-être collectif, ne doivent faire oublier que toutes les conceptions exposées dans cet article participent de certaines postulations philosophiques communes, d'autant plus massives qu'elles sont rarement explicitées. En premier lieu, toutes s'accordent à voir dans l'individu sinon la source unique des jugements de valeurs portés au nom de la collectivité, du moins leur source primordiale et le seul objet dont ils traitent en réalité. Dans le calcul parétien du bien-être, en théorie du choix social, aussi bien que dans les théories de la négociation ou de la position originaire, les individus sont à l'origine des jugements de valeurs substantiels, qu'il s'agit alors seulement d'« agréger », en s'aidant d'autres jugements de valeur, subordonnés parce que purement formels (par exemple, le principe de Pareto, l'axiome arrovien d'indépendance). S'il est vrai que le schéma se brouille dans les travaux les plus récents d'économie normative, il demeure que les jugements de valeur substantiels ne concernent finalement que l'existence individuelle, dans telle ou telle de ses modalités -- consommation des ressources, « réalisations », etc. En second lieu, les constructions présentées ici entretiennent une relation étroite, mais mal définie et insuffisamment problématisée, avec la philosophie morale traditionnelle. Les normes explicitées par la théorie arrovienne du choix social concernent aussi bien les fondements de la vie politique, que l'appréciation morale de la répartition des biens. Lorsque le projet d'une théorie économique de l'« équité » s'autonomise, l'interprétation éthique devient sans doute prédominante; mais on peut avoir alors l'impression que la moralité, aux yeux de certains auteurs qui ne sont pas des moindres, se réduirait aux seuls aspects de justice distributive. Ainsi, les conceptions parcourues dans cet article tantôt

excèdent, tantôt limitent, le champ d'investigation ordinaire de l'éthique. Enfin, parce qu'ils emploient les méthodes de l'économiste théoricien, et héritent de concepts techniques qu'ils n'analysent pas toujours complètement, les auteurs contemporains en répercutent certaines ambiguïtés caractéristiques. En particulier, même si la critique du « welfarisme » a permis d'en mieux délimiter le rôle, la notion de préférence apparaît à la fois omniprésente et mystérieuse: elle intervient tantôt dans la prise de décision et le choix, tantôt dans ces actes purement idéels que sont l'évaluation et le classement; elle décrit tantôt le bien de l'individu de façon quasi-objective, tantôt son bien-être ou son bonheur, tantôt, seulement, ses goûts. Les économistes du 20ème siècle ont sans doute affermi le fondement de leurs théories en subordonnant l'utilité à la préférence; mais ce remaniement n'a pas dissipé entièrement l'obscurité de leurs concepts premiers.

BIBLIOGRAPHIE

Arrow (Kenneth), Social Choice and Individual Values, New Haven, Yale University Press, 1951 (2ème éd. révisée, 1963).

d'Aspremont (Claude), « Economie du bien-être et utilitarisme », in L.A. Gérard-Varet et J.C. Passeron, Le modèle et l'enquête. Les usages du principe de rationalité dans les sciences sociales, Paris, Editions de la Maison des Sciences de l'Homme, 1995.

Fleurbaey (Marc), Théories économiques de la justice, Paris, Economica, à paraître.

Gauthier (David), Morals by Agreement, Oxford, Clarendon Press, 1986.

Halévy (Elie), La formation du radicalisme philosophique, 3 tomes, Paris, Félix Alcan, 1901-1904, à paraître dans une édition révisée, Paris, P.U.F., avec des postfaces de P.Bouretz, J.P.Dupuy et P.Mongin.

Harsanyi (John), Essays in Ethics, Social Behavior, and Scientific Explanation, Dordrecht, Reidel, 1976.

Kolm (Serge), Modern Theories of Justice, Cambridge (Mass.), The M.I.T. Press, 1995.

Moulin (Hervé), Axioms of Cooperative Decision-Making, Cambridge, C.U.P., 1988.

Pareto (Vilfredo), Manuel d'économie politique, 1907, rééd. in Oeuvres complètes, t.VII, Genève, Droz, 1966.

Pigou (Alfred), The Economics of Welfare, Londres, MacMillan, 1920.

Rawls (John), A Theory of Justice, Cambridge (U.S.A.), Belknap, 1971 (trad.fr. La théorie de la justice, Paris, Le Seuil, 1987).

Roemer (John), Theories of Distributive Justice, Cambridge (Mass.), 1995.

Sen (Amartya), Collective Choice and Social Welfare, San Francisco, Holden Day, 1970.

Sen (Amartya), Inequality Reexamined, Oxford, Clarendon Press, 1992.

van Parijs (Philippe), Qu'est-ce qu'une société juste?, Paris , Le Seuil, 1991.