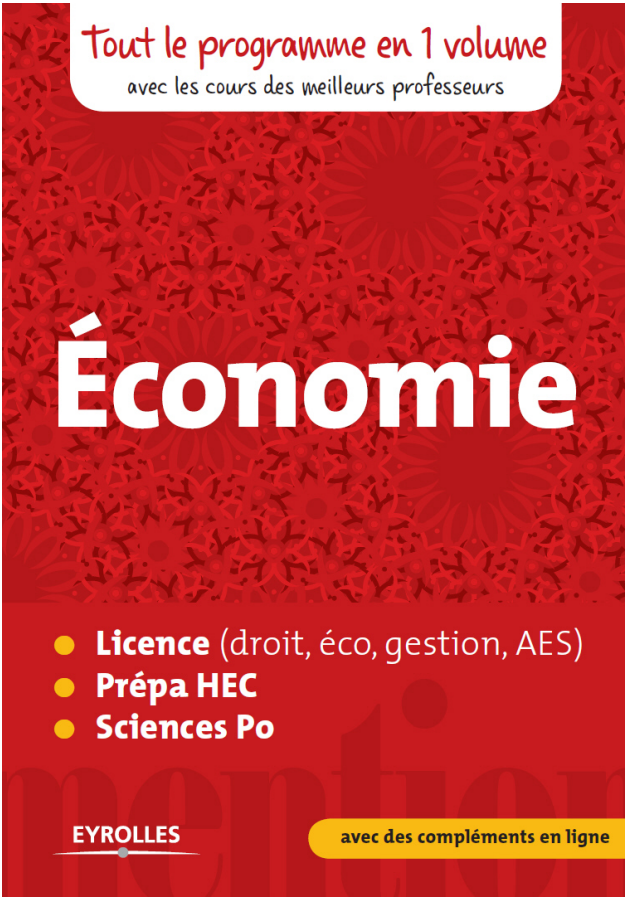
Économie

Tout le programme en un volume
avec les cours des meilleurs professeurs



EYROLLES

2008

Sommaire

Chapitre 1 – Histoire économique 7

I. Microéconomie (les agents et leur coordination)

Chapitre 2 – Comportements individuels 28
Chapitre 3 – Système de prix et équilibre 53
Chapitre 4 – Concurrence et économie industrielle 68
Chapitre 5 – État, impôt et politique publique 91

II. Macroéconomie (les agrégats et les marchés)

Chapitre 6 – Monnaie, banque et marchés financiers 111
Chapitre 7 – Économie du travail 129
Chapitre 8 – Croissance économique et développement 152
Chapitre 9 – Commerce international et mondialisation 169
Chapitre 10 – Crises et cycles 189

III. La science économique

Chapitre 11 – Qu'est-ce que l'économie ? 208
Chapitre 12 – Histoire de la pensée économique : les grands auteurs et leurs doctrines 224

Chapitre 1

Histoire économique

JEAN-YVES GRENIER

Atelier François-Simiand, EHESS

GILLES POSTEL-VINAY

Atelier François-Simiand, PSE (INRA), EHESS

Les interactions entre histoire et économie sont nombreuses et dans l'ensemble très asymétriques, les historiens (souvent d'ailleurs économistes de formation) empruntant fréquemment les méthodes, voire certaines problématiques de la théorie économique. Cette bonne entente signifie-t-elle convergence intellectuelle ? Oui, dans une large mesure, car l'économie aide l'historien à mieux construire ses concepts essentiels (l'institution, le marché…) et à formuler, voire à résoudre, les problèmes qu'il se pose. Oui, avec des réserves, que certains estimeront essentielles. La première est soulevée par les économistes eux-mêmes (David, North) quand ils mettent en question certains présupposés de la théorie néoclassique, souvent incapable d'expliquer la non-optimalité de l'allocation des ressources ou les défaillances du marché. La seconde porte sur l'écart qui existe entre la façon économique, déductive, d'appréhender le monde, et les modalités empiriques et narratives de l'historien. Les passerelles existent et les frontières entre les deux deviennent plus poreuses à mesure que prolifèrent des zones intermédiaires au sein des sciences sociales, que ce soit la sociologie économique, le courant *law and economics* ou la réémer-

gence – après une vraie mutation – de l'économie politique. Leur multiplication rend plus nécessaire aujourd'hui qu'hier une réflexion sur la complémentarité des approches en science(s) sociale(s) (R. Bates *et al.*).

UNE « MACHINE À EXPLORER LE TEMPS »

Ce chapitre entend mettre à l'épreuve ces interrogations à propos du temps et de la durée. C'est un point de rencontre parce que tous les phénomènes économiques prennent du temps et que l'introduction sérieuse d'un tel paramètre conduit à repenser les catégories de l'économiste. Les hypothèses d'un marché parfait, d'une information complète ou de contrats totalement explicites peuvent certes être reprises de l'intérieur, et les théoriciens d'aujourd'hui ne s'en privent pas, mais elles doivent aussi être reconsidérées par ces spécialistes du temps que sont les historiens.

LE TEMPS DU CHANGEMENT

Transition au capitalisme et révolution industrielle

La transition est une forme classique pour penser le changement comme temps de passage d'un ordre à un autre. On connaît la portée mais aussi les limites des analyses construites sur cette notion qui, sans surprise, a souvent irrité les historiens : « Vous connaissez, vous, en histoire, une époque qui ne soit pas de transition ? », grommelait Lucien Febvre. Pour autant, l'idée de transition et ses limites mêmes ont nourri l'investigation.

A. Les transitions démographiques et agricoles entre le 18e et le 19e siècle ont surtout mobilisé les historiens, du fait sans doute de la grande disponibilité de sources quantifiables (registres paroissiaux, dîmes…). Pour la transition démographique, qui voit croître la popula-

tion européenne mais aussi asiatique, il s'agissait de répondre à une double interrogation. Pourquoi observe-t-on un net recul de la mortalité au cours du 18e siècle (dès 1730 en Angleterre), alors qu'aucun paramètre démo-économique ne semble l'expliquer ? Quels en sont les effets économiques en termes d'évolution du niveau de vie et de changements structurels (baisse du coût du travail, modification des prix relatifs, etc.) ?

Du côté de l'agriculture, l'enjeu était de comprendre comment un tel surcroît de population (8 à 30 % selon les pays) avait pu être nourri. La réflexion s'adosse à un petit nombre d'hypothèses empruntées à Ricardo (rendements décroissants) et à Malthus (forte élasticité de la réponse de la population à une augmentation de revenus). Si l'on suppose que la technique disponible ne change pas, l'histoire ne peut que suivre de longues oscillations, les phases d'expansion étant suivies d'une récession à mesure que le stock de population augmente plus vite que celui des subsistances. Ce sont les grands cycles agraires pluriséculaires, engendrés par les « ciseaux » ricardo-malthusiens, qu'ont mis en évidence M. Postan pour l'Angleterre et E. Le Roy Ladurie pour la France. Dans cette logique, la transition qui permet de sortir du « piège malthusien » suppose des changements techniques radicaux qu'une abondante historiographie s'est efforcée de repérer à partir du 18e siècle. Le problème reste cependant ouvert puisqu'un tel modèle peine à intégrer deux constats :

- le fait que les crises les plus graves, censées réguler la relation population/subsistances, ne se produisent pas au moment où se vérifient les plus fortes densités d'habitants, ce qui suppose l'existence de formes d'autorégulation démographique ;
- le fait que beaucoup des changements identifiés comme relevant de cette révolution sont attestés depuis le 12e siècle (voir George Grantham).

B. La transition vers la révolution industrielle a aussi souvent été pensée avec les concepts économiques usuels. C'est le cas notamment des travaux qui, inspirés par la tradition smithienne, l'ont conçue

comme transition vers un monde de marchés à la fois plus vastes et intégrés entre eux. C'est le cas également des analyses qui ont construit leur chronologie sur un basculement des conditions de production. Au cours des années 1950-1970, de nombreux chercheurs se sont efforcés de prendre la mesure de la croissance européenne en longue période en reconstituant des comptabilités nationales rétrospectives. En dépit d'une sophistication croissante, ces travaux ont conforté la chronologie admise. Ils ont souligné l'avance anglaise au tournant du 18e et du 19e siècle (même si les évaluations initiales ont été revues à la baisse au point de paraître bien pâles comparées aux taux que connaissent aujourd'hui la Chine ou l'Inde), puis le déplacement progressif du centre de gravité des économies occidentales vers l'Allemagne ou les États-Unis. Ces rythmes divergents ont alors été expliqués comme renvoyant d'abord à l'investissement et aux formes d'accumulation de capital physique et humain pratiquées dans chaque pays, donc comme principalement internes.

L'attention récente aux effets de la globalisation sur les économies a modifié cette perspective. Privilégiant des processus externes comme les migrations, les flux de capitaux et l'unification des prix, une nouvelle historiographie a rouvert des débats. Le message en est plus optimiste, notamment dans les travaux consacrés à la période marquée par l'extension des politiques libre-échangistes, de la fin des *Corn Laws* anglaises (1846) à la Première Guerre mondiale. Dans ce cadre, les pays les plus riches, dont on avait jusque-là souligné les trajectoires divergentes, paraissent s'inscrire dans un vaste processus d'équilibre au terme duquel les salaires réels de leurs ouvriers sont rattrapés par ceux des pays du reste de l'Europe. Sans doute le constat ne vaut-il que pour les salaires non qualifiés, de sorte qu'on ignore les effets éventuels des efforts faits à l'époque en matière d'éducation (voir O'Rourke et J.G. Williamson). Sans doute peut-on aussi s'interroger sur le cadre même de ce type d'analyse (voir par exemple J. V. Nye). Retenons du moins qu'une histoire s'est bâtie ainsi sur une analyse économique et le type particulier de durée qu'elle induit, en l'occurrence ici la durée

nécessaire à l'équilibre. L'histoire se scinde entre un passé durant lequel les économies sont des isolats et une transition qui s'opère au rythme de la convergence – rapide au 19e siècle, contrarié pendant une part du 20e siècle.

C. La temporalité de la transition a un lien privilégié mais non exclusif avec les approches macroéconomiques. En témoignent les recherches sur la question des transferts intergénérationnels, inspirées par celles de G. Becker sur les transferts au sein de la famille. Un tel modèle implique une division des tâches et une complémentarité entre le rôle des familles (en matière d'éducation) et de l'État (en matière de retraite). Il a donc, lui aussi, la propriété de suggérer une histoire. *Avant*, régnerait une sorte d'âge du fer puisque les rapports entre générations relèveraient de l'échange, les parents laissant espérer un héritage à leurs enfants « en échange » de telle ou telle contrepartie – essentiellement un soutien pour leurs vieux jours. *Après*, ces flux s'inverseraient dès lors que la famille devenue altruiste rendrait à sa progéniture des services essentiels (dans le domaine de l'éducation) que l'État ou le marché ne pourraient remplacer qu'à un coût prohibitif. Le modèle implique donc un présent, un passé et une transition. En l'occurrence, la césure se situerait au 19e siècle.

Sans doute l'historien se défiera-t-il d'une chronologie qui fait évoluer en sens inverse le rôle de la famille et celui de l'État, en supposant que famille et État soient des invariants historiques identifiables à des catégories évidentes et stables. Il récusera aussi l'idée que le monde occidental ait connu une transition au terme de laquelle l'assistance serait passée à l'État après avoir relevé de la seule logique familiale, comme ce serait aujourd'hui le cas des pays en voie de développement (voir P. Horden et R. Smith). En revanche, il retiendra comme hypothèse le récit suggéré par cette transition. Au sein des familles, les solidarités se réorienteraient, les transferts descendants succédant aux transferts ascendants dès lors que ceux-ci seraient pris en charge par des marchés (comme ceux de l'assurance qui faisaient défaut jusque-là) ou par l'État.

À en juger par ce qui s'est produit aux États-Unis et en France, on observe bien un tel enchaînement (voir A. Bommier *et al.*). Pourtant, force est de reconnaître qu'au cours des derniers siècles, la politique d'éducation et la politique des retraites ont évolué séparément sans interférer l'une avec l'autre. Elles ont connu des phases d'accélération presque simultanées, mais cette concomitance fut fortuite et, si chacune a pu s'en trouver renforcée, il est sûr que leur rencontre ne fut pas conçue comme un objectif politique. Si le quasi-contrat beckerien de coopération intergénérationnelle peut devenir un projet pour le futur, il n'a été – à l'échelle des États du moins – le futur d'aucun passé. En revanche, rien n'interdit de penser que la réalité ait été différente à l'échelle locale, qui est celle à laquelle ces choix ont été effectués depuis l'époque moderne.

D. Les récents succès de certains pays asiatiques, en particulier la Chine, ont démontré que l'Occident n'était pas le seul lieu d'élection du capitalisme. Ce qui soulève une question à laquelle des historiens ont tenté de répondre. L'avance de l'Europe occidentale est en fait récente, et l'histoire de l'économie mondiale jusqu'au début de l'industrialisation est un jeu complexe d'échanges, en particulier de technologies, au sein duquel l'Ouest est souvent en déficit par rapport à l'Est (surtout la Chine).

Un sinologue américain, K. Pomeranz, a entrepris de comparer les cas chinois et européen. Il part du constat qu'en 1750, à la veille de la révolution industrielle européenne, les situations sont très similaires : mêmes caractéristiques démographiques, niveaux proches de consommation et de production, efficacité comparable des institutions de marché, etc. Par contre, une spécificité forte des États européens est leur propension à se projeter vers l'extérieur. Aucune autre région du monde n'a consacré autant d'énergie et d'argent pour conquérir des territoires lointains (Amérique du Nord et du Sud) ou contrôler des régions proches (Afrique). Or cette mainmise territoriale a initié un circuit d'échanges : biens manufacturés et main-d'œuvre *versus* biens agricoles et matières premières. Les importations agricoles en particulier

celles provenant des États-Unis et de la Russie, ont libéré, avant toute révolution agricole, une partie de la population active des tâches de production des biens de subsistance, population qui a pu ainsi se consacrer à des activités manufacturières.

À l'inverse, la Chine n'a pas mené à bien une telle expansion outremer, préférant conquérir des marges (comme le Xinjiang) ou exporter ses populations dans des régions proches (Indonésie, Malaisie…), selon une logique d'empire que ne pouvait pratiquer l'Europe, divisée en pays de taille réduite. La Chine est dès lors restée prisonnière d'une proto-industrialisation bloquée par les rendements décroissants de l'agriculture. Les rizières chinoises entre 1750 et 1850 ont dû mobiliser un volume croissant de population afin de produire du riz à des coûts toujours plus élevés ; et cela au détriment des investissements et de la consommation de biens manufacturés que les entrepreneurs chinois savaient pourtant aussi bien fabriquer et vendre que leurs homologues européens. C'est l'état stationnaire qu'Adam Smith et les économistes classiques envisageaient comme l'avenir de l'Europe !

L'analyse de Pomeranz a beaucoup été discutée. J. Goody lui reproche ainsi de s'en tenir à une approche strictement économique et de négliger d'autres critères concernant, par exemple, la structure sociale, le rôle de l'État ou le comportement des entrepreneurs. Surtout, si Pomeranz a raison de refuser l'opposition entre un Occident dynamique et une Asie bloquée sur des technologies traditionnelles, il sous-estimerait cependant beaucoup le rôle du changement technique dans la révolution économique occidentale.

Le temps des cycles

Pendant longtemps, l'analyse des conjonctures a tenu un rôle central dans la construction de l'histoire économique comme discipline, ce qui s'explique en partie par l'abondance des sources sérielles disponibles, durablement privilégiées comme indicateur à la fois des continuités et du changement historique.

A. Tout commence avec les travaux des économistes qui, de Juglar à Schumpeter en passant par Marx et Kondratiev, ont exploré la dimension cyclique de la régulation capitaliste. Identifiant des cycles plus ou moins longs dans le mouvement des prix, ils les ont mis en relation avec divers facteurs comme la production, le profit, le commerce extérieur ou les investissements.

Dans les années 1920, en particulier, étudiant des séries de prix depuis 1789, Kondratiev a repéré des mouvements de longue durée (25 ans), de croissance puis de récession, qu'il a proposé de relier à l'investissement. Schumpeter, quant à lui, a souligné le lien qui existait entre les phases d'essor économique et les « grappes » d'innovations majeures (comme la révolution ferroviaire), l'épuisement de ces innovations étant à l'origine du retournement de la croissance. L'évolution du capitalisme serait ainsi marquée par une « destruction créatrice » liée à l'apparition de nouvelles techniques.

D'autres, F. Simiand notamment, ont insisté sur l'origine monétaire de ces cycles, l'accroissement de la quantité de monnaie (découverte de mines d'or ou d'argent, émission de papier-monnaie ou octroi de crédits) engendrant une phase de croissance longue, alors que les restrictions monétaires génèrent un ralentissement prolongé.

B. Cette analyse des cycles a beaucoup influencé les historiens, le paradoxe étant que ce sont les modernistes qui s'en sont le plus inspirés malgré la nature profondément différente des cycles préindustriels. Publiés en 1933 et en 1944, les deux livres d'E. Labrousse proposent l'une des analyses économiques les plus achevées du 18e siècle et de la fin de l'Ancien Régime. Profond connaisseur des économistes anciens, il leur emprunte leur explication de la crise d'ancien type qu'il formalise à son tour en la complétant. Il reproche en effet au schéma avancé par Quesnay ou Turgot de mal relier le mouvement de la production et celui des prix, faute notamment de percevoir combien ceux-ci sont affectés par les substitutions de consommation. Surtout, il introduit l'inégalité sociale au cœur du modèle puisque, souligne-t-il,

l'évolution des prix bénéficie à « une minorité féodale » alors qu'elle pénalise « la masse de la nation ». Cette contradiction ne fait que s'aggraver au cours de cette période d'inflation lente qu'est le siècle des Lumières, ce qui conduit à une accentuation des crises qui culmine avec celle de 1789, l'économique et le social engendrant dès lors le politique.

C. Les années 1980 et la disparition apparente de toute régularité conjoncturelle ont considérablement réduit l'intérêt que les historiens accordent à l'étude des formes de la conjoncture, cette dernière n'étant plus dotée de la capacité qui lui était jusqu'alors reconnue de dire la structure.

En revanche, la question de l'inégalité, centrale déjà chez Labrousse, reste une donnée structurelle majeure dans la réflexion des historiens sur les rythmes de l'économie. Elle peut en effet être pensée comme une dynamique cyclique. Il y a là un problème classique des sciences sociales, et « critiques et partisans du capitalisme continuent à s'opposer sur l'existence ou l'inexistence d'une tendance qui tirerait la distribution des revenus et des richesses vers plus d'égalité » (Peter Lindert). Les uns soulignent que le fossé entre riches et pauvres se creuse. Les autres pensent que le développement des rapports marchands porte en lui-même la réduction des inégalités à l'intérieur des pays et entre eux.

Dans un article célèbre de 1955, S. Kuznets a proposé une voie moyenne. L'industrialisation engendrerait un cycle au cours duquel les inégalités commenceraient par s'accroître pour se réduire dans un second temps. Depuis, les recherches sur ces questions se sont multipliées sans que l'accord ne se fasse ni sur l'évolution diagnostiquée ni sur les mécanismes qui l'expliqueraient. P. Lindert, par exemple, sans remettre en cause la « courbe de Kuznets », admet qu'elle apparaît aujourd'hui comme une sorte de « voie lactée » entourée d'un halo d'incertitude. D'autres, au contraire, invitent à sortir de la logique du cycle pour mettre en avant les effets de l'inflation du 20e siècle et des politiques fiscales redistributives (voir Thomas Piketty).

Le temps de l'épuisement

Si l'économie a fourni des outils pour analyser la temporalité propre à la transition ou au retour cyclique, elle a aussi cherché à modéliser la finitude et l'épuisement. De prime abord, ce problème paraît être l'apanage du présent. Pourtant, c'est loin d'être le cas. Les manuels de marchands du 18e siècle évoquent ainsi la diminution des ressources halieutiques (voir R. Allen et I. Keay), et c'est avec la crise des années 1930 que les travaux théoriques font les avancées les plus significatives.

Sans surprise, ce sont les gisements pétroliers qui ont alors suscité le plus d'analyses dans le sillage des travaux de H. Hotelling. Pour celui-ci, extraire le pétrole d'un gisement donne lieu à une rente due à la différence entre le coût marginal de production et le prix du marché qui s'accroît à mesure que la ressource se raréfie. Cette rente de rareté est optimale quand, à l'épuisement de la ressource, son prix est tellement élevé que sa demande est nulle. Au-delà de son domaine spécifique, l'apport d'un tel schéma est d'introduire le caractère intertemporel des décisions des agents. Caractériser le choix économique par des contraintes sans dimension temporelle est une solution « tout à fait insatisfaisante » lorsqu'on l'applique à une industrie « pour laquelle la réalisation d'un état stationnaire est une impossibilité physique » (J. Bourdieu).

LE TEMPS DE LA PERMANENCE OU DE LA TRÈS LONGUE DURÉE

Longtemps le concept de longue durée, inventé par F. Braudel, n'a pas fait problème, appuyé qu'il était sur l'idée de structure alors au cœur des sciences sociales et sur une décomposition fortement déterministe. Les historiens, désormais méfiants à l'égard de situations considérées comme presque immuables, se sont dépris de cette notion. C'est au carrefour de l'histoire et de l'économie, discipline pourtant peu sensible aux durées historiques, que la longue durée a retrouvé de la vita-

lité. *« History matters »*, affirment D. North et P. David, leur objectif commun étant d'ajouter la dimension du temps à l'analyse néoclassique traditionnelle.

Dépendance temporelle et inefficience des marchés

Un processus d'allocation économique est dit *path dependent* lorsque son histoire, en particulier des événements lointains de nature aléatoire, exerce des effets durables sur l'efficience du système économique. Le principe d'un processus *path dependent* est le suivant : d'abord, un blocage *(lock in)* apparaît à un moment donné dans l'évolution d'une institution (ou d'un système technique), puis l'existence de rendements croissants rend ce blocage de plus en plus irréversible.

L'exemple devenu canonique est celui étudié par P. David de l'ordre des lettres sur le clavier des machines à écrire (QWERTY). Le choix du QWERTY, ordre non optimal, s'est imposé lors des débuts de la machine à écrire (inventée en 1868) afin de ralentir la vitesse de frappe et ainsi éviter que les tiges porte caractère ne se coincent. Le QWERTY devient rapidement un standard, imposé par la firme qui l'a inventé, Remington, alors en position dominante sur le marché. Un effet de *lock in* se produit, aucune entreprise n'étant en pouvoir de fabriquer ou d'utiliser un autre clavier même quand, en 1936, Dvorak en invente un plus rapide. Plus le temps passe, plus l'irréversibilité dans l'usage du QWERTY se renforce. Ainsi, un événement sans grand relief historique mais conditionné par un état temporaire de la technique (la mécanique des tiges porte caractère) a provoqué un blocage de longue durée dont les effets durent encore. L'objectif affiché des partisans de la *path dependency* est atteint : prouver à la communauté des économistes qu'une analyse en termes d'équilibre sans durée ne peut pas expliquer l'inefficience de nombre de systèmes économiques et donc les échecs de marché (*markets failure*).

Les économistes néoclassiques ont réagi. D'abord, sur le plan factuel, en soulignant que l'inefficacité du QWERTY par rapport à d'autres claviers n'est pas prouvée : en somme, il ne s'est pas imposé par hasard. Ensuite, on peut considérer que le choix d'une norme n'a pas d'effets macroéconomiques majeurs. En effet, l'écart de productivité avec la meilleure technique disponible est nécessairement limité, car tout écart important crée des opportunités incitant les entreprises à investir. Enfin, les néoclassiques considèrent que de deux choses l'une : soit l'inefficience de la technique sélectionnée par la *path dependency* n'était pas prévisible, alors elle ne pouvait être prise en compte par la rationalité économique ; soit elle l'était, et l'hypothèse de rationalité des comportements fait qu'elle aurait été choisie. En bref, il n'y a pas de *market failure*.

Institution et coût de transaction : l'école néo-institutionnaliste

North estime nécessaire de combler l'une des grandes faiblesses de la théorie néoclassique, en l'occurrence le fait de considérer comme données des institutions de marché dont l'inachèvement ou l'imperfection sont au contraire un problème majeur. C'est parce qu'il reproche à la New Economic History de reposer exclusivement sur l'analyse marginaliste que North s'ouvre à la nouvelle économie institutionnelle. S'inspirant des travaux fondateurs de R. Coase, ce courant s'intéresse aux déterminants des coûts de transaction, en particulier au sein de l'entreprise, et à la définition des droits de propriété. North reprend cette problématique en l'élargissant à l'ensemble des institutions d'échange et à la très longue durée historique. North et Wallis procèdent ainsi à une estimation de l'ensemble des coûts afférents aux transactions dans l'économie américaine entre 1870 et 1970 : résultat convaincant puisqu'ils se monteraient au quart du PIB en 1870 et à la moitié un siècle plus tard.

C'est cependant la question des droits de propriété qui a donné lieu aux analyses les plus probantes car elle pèse directement sur l'efficience

du système économique, que ce soit leur absence qui conduit à un gaspillage de ressources rares (Alchian et Demsetz), ou leur insuffisante définition qui rend plus difficile l'application des contrats.

D. McCloskey discute ainsi la pérennité des systèmes d'*open field*, caractérisé par des droits collectifs importants et donc censés peu efficients, mais qui traduisait aussi un comportement rationnel face aux risques de mauvaises récoltes. Nombre d'études ont également souligné que la politique des enclosures conduisait dans le long terme à une réduction des coûts de transaction, position contestée par d'autres qui considèrent au contraire que l'efficience globale de l'agriculture anglaise n'a pas augmenté mais que les enclosures ont surtout conduit à une redistribution de la propriété foncière en faveur des plus riches (voir Robert Allen).

À l'inverse, J.-L. Rosenthal a souligné combien l'identification plus précise des biens fonciers permise par ce grand bouleversement des propriétés que fut la Révolution française a conduit à une nette amélioration de la productivité agricole, incitant par exemple à des opérations d'irrigation.

Tous les domaines de l'économie sont *a priori* concernés par l'application d'un tel paradigme. D. North et B. Weingast l'ont ainsi étendu aux questions financières dans un article célèbre consacré aux effets de la *Glorious Revolution* anglaise (1688). L'instauration d'un contrôle parlementaire sur les finances publiques aurait augmenté la confiance des prêteurs, diminué les coûts de transaction, ce qui renforça le marché des capitaux, atout maître de l'économie anglaise jusqu'à nos jours.

La force du paradigme des coûts de transaction est sa généralité. La conversion récente au capitalisme de nombreux pays renforce encore sa dimension téléologique. Là réside aussi, paradoxalement, l'une de ses faiblesses aux yeux des historiens, car cette temporalité n'est pas sans rappeler les philosophies historicistes qui faisaient de l'extension temporelle d'un concept la clef de la dynamique historique. Une autre

interrogation réside dans la dimension fonctionnaliste d'un paradigme pour lequel la rationalité des transformations historiques est à comprendre *in fine* à travers l'unique prisme de l'utilitarisme, même si, plus que d'autres, North a insisté sur l'impact du politique sur les incitations économiques qui régissent les institutions.

Théorie des jeux et équilibres multiples

D'autres analyses inspirées par le courant néo-institutionnaliste renvoient à des équilibres plus durables encore, notamment ceux que rendent possibles certaines institutions informelles de marché. Comme celles-ci intègrent des facteurs culturels, sociaux et politiques qui affectent les formes de coordination, les anticipations, l'information disponible et jusqu'à la possibilité de recourir à la contrainte légale, elles manifestent une forte dépendance à l'égard du passé.

Tel est le cas des institutions marchandes à l'émergence desquelles s'intéresse A. Greif, qui mobilise à cet effet conjointement l'histoire et la théorie des jeux. Cela le conduit à mettre l'accent sur le rôle d'institutions qui ont permis aux marchands d'établir des relations d'agence ne dépendant ni de contraintes familiales, ni d'une éthique partagée ou de sanctions juridiques. Greif examine ainsi le cas des marchands maghribis, descendants de commerçants juifs venus d'Irak au 10e siècle.

Les relations d'agence entre eux étaient régies par une institution reposant sur un mécanisme de réputation multilatérale. Ces marchands choisissaient leurs agents en leur sein, mais il était d'usage constant que si un agent se montrait déloyal envers l'un d'eux, il ne devait plus jamais être employé par aucun membre du groupe. D'un côté, leur réseau social et commercial fournissait les canaux d'information requis pour détecter la fraude et en répandre la nouvelle si elle était avérée. De l'autre, puisque la punition était multilatérale, elle rendait le contrat auto-exécutoire : l'agent devait rester honnête à l'égard du marchand avec qui il était en affaire sauf à prendre le risque de ne plus jamais travailler avec aucun Maghribi. Ce type de punition multilatérale s'est

révélé particulièrement efficace. Le gain attendu de la coopération et la certitude que les règles en matière d'embauche seraient bien respectées assuraient la « fermeture » du système.

Le travail de Greif est admirable d'inventivité et de précision, les paramètres du jeu finalement retenus résultant non d'une supposition initiale mais d'un ajustement patient entre prédictions du modèle et constats empiriques à partir des archives. Son entreprise n'en soulève pas moins deux questions qui illustrent bien la difficulté du dialogue entre histoire et économie. La première est que l'institution de la confiance au sein des Maghribis repose sur une croyance qui est supposée être connaissance commune, tous les acteurs sachant la nature des interactions qui les lient. On reconnaît l'hypothèse d'anticipations rationnelles : c'est parce que les acteurs savent d'avance que l'institution leur est utile que cette dernière est créée. Ce fonctionnalisme a sans doute sa part de vérité mais il est assez pauvre pour faire comprendre dans sa généralité l'émergence des institutions.

En outre, l'analyse utilisée en termes d'équilibre de Nash suppose des interactions entre deux joueurs seulement (ici un marchand et le reste de la communauté) ce qui permet de spécifier les équilibres mais élude la difficulté majeure qui réside dans la coordination à distance d'une multiplicité d'agents. En ce sens, la durée générée par l'institution de Greif est plus un temps du projet qu'un temps historique.

LE TEMPS DE L'APPRENTISSAGE

L'invention de l'*Homo œconomicus* (et du marché concurrentiel) conditionne l'effectivité des mécanismes analysés par le théoricien comme par l'historien. Or, à la différence de l'analyse néoclassique qui envisage cet être abstrait de façon purement déductive (ce qui a sa propre logique), l'historien de l'économie doit s'efforcer de façon plus inductive d'en repérer l'émergence progressive, jamais achevée.

L'historicité des formes du marché

L'invention du marché et des principes modernes de l'échange est l'enjeu d'un vieux débat dans les sciences sociales. Dès la fin du 19e siècle, les anthropologues se divisent entre modernistes et substantivistes. Pour les premiers, les sociétés primitives sont régies par les mêmes principes de calcul et de choix rationnel que les sociétés contemporaines, faisant ainsi de l'*Homo œconomicus* un être intemporel. Les seconds, au contraire, insistaient sur les différences radicales dans les comportements matériels entre les primitifs et les modernes.

Les historiens s'insèrent dans ce débat, en particulier autour de l'œuvre de M. Rostovtzeff, qui analysait l'économie de l'Empire romain sous des traits qui rappelaient le monde capitaliste de son temps. La problématique rebondit après la Seconde Guerre mondiale avec K. Polanyi, qui considérait que le paradigme de l'économie de marché concurrentiel et autorégulé n'était valable que pour une courte période, allant du milieu du 19e siècle jusqu'à la Seconde Guerre mondiale, époque après laquelle s'impose le capitalisme administré.

C'est un historien qui lui répond dans les années 1970, F. Braudel préférant une distinction non pas diachronique mais synchronique entre économie de marché et capitalisme, le second par sa puissance ayant toujours su manipuler la première.

Ce débat se poursuit aujourd'hui avec d'autres enjeux, en particulier sous la plume de juristes réfléchissant à la croisée du droit et de l'économie. R. Steinfeld insiste ainsi sur l'historicité d'une composante du marché souvent considérée comme acquise depuis longtemps en Europe occidentale : la liberté du travail. En effet, si l'esclavage a vite disparu du continent européen, la frontière entre travail libre et travail forcé ou contraint est des plus floues durant tout le 19e siècle. Les *Master and Servant Acts* anglais prouvent que l'application des contrats de travail relevait moins de la justice que de la police, la coercition prévalant sur la libre négociation au sein du marché du travail. Ces pratiques ne cessent pas en Angleterre avant les années 1875, et aux

États-Unis avant les années 1930. La fiction, entretenue par une lecture sans contrôle d'Adam Smith, de l'émergence définitive du travail libre dès la fin du 18e siècle est donc fausse.

L'invention de l'*Homo œconomicus*

D'autres exemples historiques confortent cette analyse selon laquelle les comportements économiques rationnels sont le produit d'un apprentissage. Déjà, le sociologue allemand G. Simmel, auteur de la *Philosophie de l'argent* (1908), a montré que l'usage de plus en plus généralisé de l'argent comme équivalent général a progressivement conduit à une désincarnation de la relation de l'individu aux objets et aux autres individus. De là, un rapport au monde abstrait et calculateur qui s'est progressivement généralisé en Occident entre la Renaissance et la fin du 19e siècle. Beaucoup d'études empiriques illustrent cette réflexion, à propos, par exemple, du difficile apprentissage des règles du marché financier par l'épargnant américain à la fin du 19e siècle (voir R. Cull et L.E. Davis), ou bien, à la même époque, de l'initiation progressive des immigrants américains aux mécanismes de la consommation et de la rationalité monétaire (voir V. Zelizer).

Ces transformations sont toujours à l'œuvre, les comportements se modifiant à mesure que le capitalisme invente de nouveaux rapports aux objets économiques. Un sujet de réflexion pour les historiens de demain, par exemple, sera peut-être les effets, sur les comportements des intermédiaires financiers, de l'invention en 1974 de la formule de Black et Scholes pour calculer le prix des options.

Mise au travail et *consumer revolution*

Une question majeure de l'histoire économique est de comprendre comment des populations très étrangères à la vie industrielle, ses contraintes de lieu (l'usine) et de rythmes, ont pu accepter de se soumettre à cette nouvelle discipline. L'explication traditionnelle, de Marx à

Foucault, insiste sur la dimension coercitive d'une telle évolution, subie plus que voulue par les prolétaires de l'industrialisation. E. Thompson a insisté en particulier sur l'apprentissage difficile de la contrainte temporelle par les ouvriers anglais entre le 18e et le 19e siècle. Des pistes de réflexion assez différentes ont été récemment proposées.

La première, dans une tradition libérale, explique ce phénomène non pas en termes de contrainte mais d'adhésion. Ainsi, G. Clark estime que les ouvriers eux-mêmes sont demandeurs d'une telle discipline pour être forcés au travail, afin d'accroître leur productivité, dont ils attendent en retour une augmentation salariale.

Une seconde piste, plus argumentée et mieux balisée, insiste sur l'importance de ce que les historiens britanniques ont appelé la *consumer revolution*, c'est-à-dire le désir d'acquérir des biens (de plus en plus nombreux et diversifiés) qui se répand en Angleterre après le milieu du 18e siècle. Ce nouveau rapport à la consommation, perçue plus dans une optique de jouissance que de stricte utilité, gagne tous les milieux sociaux. Il pourrait contribuer à expliquer l'émergence d'une nouvelle relation au travail. À l'ancienne relation « mercantiliste » (à partir d'un certain seuil, l'offre de travail répondrait négativement à une hausse de salaire) se substituerait ainsi une relation conforme au modèle économique standard. Comme illustration de ce processus, on retiendra les analyses de F. Cooper, qui explique la conversion au 19e siècle en Afrique d'une partie de la main-d'œuvre au travail salarié régulier, jusque-là honni, par la nouvelle disponibilité de biens de consommation qui stimule le désir d'acquérir de l'argent.

AUTRES TEMPS, LE CONTREFACTUEL ET LE VIRTUEL

Pour identifier les causes qui ont rendu nécessaire telle action plutôt que telle autre, l'historien peut être tenté par une reconstruction *a posteriori*. Elle peut prendre deux formes.

Téléologie et histoire contrefactuelle

L'une des réponses les plus convaincantes apportée à cette difficulté est la *conterfactual history*. Pour l'historien, il s'agit de rien moins que de se mettre dans la position de faire « comme si » il ignorait comment se sont achevés les événements qu'il observe. L'effet de connaissance naît alors de la confrontation de données historiques empiriques avec les futurs que suggère tel ou tel modèle formel.

Le cas classique est l'étude de R. Fogel qui se demande ce qu'aurait été la croissance américaine du 19e siècle si les canaux avaient remplacé les chemins de fer. La méthode est celle de la théorie marginaliste. Le résultat est sans appel puisque Fogel en conclut que le « gain social » du chemin de fer par rapport aux canaux n'aurait pas excédé 2 % du PNB.

Les travaux de la New Economic History ne furent pas sans effet, ils conduisirent à mettre en doute les visions industrialistes traditionnelles du développement par étapes à la Rostow, qui considérait que les chemins de fer relevaient de l'« axiome d'indispensabilité ».

Les critiques furent nombreuses. Certaines d'ordre technique, comme la difficulté de spécifier les fonctions de coût, la non-prise en compte du trafic des voyageurs (pour lequel l'avantage comparatif du train est considérable), ou encore la non-prise en compte d'externalités non mesurables comme l'unification des marchés. D'autres critiques furent en fait une défense de la rationalité du marché, dans une perspective assez proche des réactions qu'auront quelques années plus tard les économistes néoclassiques contre la *path dependency*. C'est par exemple ce que propose Jon Elster, dont McCloskey résume ainsi l'argument : *« If Fogel had developed a theory of invention to draw a less vague picture of road transport without railroads, he would have faced the problem that the very theory would predict the existence of railroads. After all, railroads were actually invented and therefore should be predicted by a sound theory of innovation. »*

Temps virtuel et anticipations

Le souci d'éviter toute forme de téléologie historique a récemment orienté la recherche dans de multiples directions. Observons que cette réflexion sur la multiplicité des futurs possibles du passé s'inscrit dans l'intérêt récent, en particulier de la part des économistes, porté aux modélisations non déterministes, où les effets de non-linéarité et de bifurcation sont pris au sérieux (J. Lesourne). Les développements ces dernières années de la nouvelle théorie du commerce international (P. Krugman) ou de l'économie géographique (A. Burns), pour laquelle la dimension historique est importante, en sont des illustrations.

Comment l'historien peut-il prendre en compte ces scénarios ? Une application directe est envisageable, comme le montrent les travaux sur la *path dependency* dont l'objectif est de montrer que la trajectoire dessinée par le choix, à tout moment du temps, de la meilleure technique n'est qu'un des futurs possibles du passé. Observons également que les réflexions comparatistes évoquées plus haut ressortent de cette préoccupation puisqu'il s'agit de mettre en question l'unicité de la trajectoire européenne en imaginant que d'autres configurations auraient pu conduire à d'autres inventions du capitalisme ou de la révolution industrielle.

Une dernière voie enfin consiste à tenter de reconstituer les anticipations des acteurs en sachant qu'ils ignoraient le futur de l'époque dans laquelle ils vivaient. Il s'agit par exemple d'examiner le comportement d'un agent aux prises avec une hyperinflation, que ce soit l'Allemagne de 1923 (P. Cagan) ou la Révolution française (T. Sargent et F. Velde). Un tel modèle a des mérites mais il prend le parti discutable d'ignorer la variété des comportements réels face à l'inflation et de réduire la diversité des actions individuelles au comportement unique d'un « agent représentatif » supposé accepter les initiatives du pouvoir comme une sorte de fatalité. L'hypothèse est inappropriée dans la mesure où elle fait abstraction des efforts de chacun pour prévoir la politique économique à un moment, précisément, où la population se

constitue en groupes d'intérêts d'autant plus attentifs à anticiper les choix des gouvernements successifs. Ainsi, tant bien que mal, ces groupes réussissent à anticiper la nouvelle configuration politique et, par là, à influer sur les décisions les concernant et sur la dynamique économique (à propos de la Révolution, voir P. Hoffman *et al.*).

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 2

Comportements individuels

BERTRAND LEMENNICIER

Professeur à l'université Paris II Panthéon-Assas

LA THÉORIE TRADITIONNELLE DES CHOIX INDIVIDUELS

La théorie traditionnelle des choix individuels a été développée par Vilfredo Pareto (1906), John Hicks (1939) et Paul Samuelson (1947). Elle repose sur l'idée que l'individu (ou le ménage) cherche à maximiser une fonction de satisfaction U, dont les arguments sont les biens x_i achetés sur le marché :

$$(1) \qquad U = u(x_1, x_2, \ldots, x_n)$$

sous la contrainte de revenu :

$$(2) \qquad R = \sum_{i=1}^{n} p_i x_i$$

où R est le revenu monétaire non salarial et p_i sont les prix monétaires des biens. L'utilité est alors maximale lorsque l'individu ou le ménage, par arbitrage, choisit une combinaison de $x_1^*, x_2^*, \ldots, x_n^*$ telle que l'euro placé dans un bien rapporte autant de satisfaction supplémentaire que le même euro placé dans un autre bien, en faisant l'hypothèse que le supplément d'utilité $Um(x_i)$ obtenu en consommant une unité de plus du bien x_i, diminue au fur et à mesure où l'on consomme davantage de ce bien x_i :

$$(3) \qquad = \frac{Um(x_1)}{p_1} = \frac{Um(x_2)}{p_2} = \ldots = \frac{Um(x_n)}{p_n}$$

C'est une règle d'arbitrage simple : dès que l'euro dépensé dans un bien rapporte plus d'utilité supplémentaire que le même euro dépensé dans un autre bien, on augmente la consommation de ce bien dont l'utilité supplémentaire va diminuer au fur et à mesure que l'on en consomme davantage. On s'arrêtera quand l'euro dépensé dans un bien ou un autre rapporte la même utilité supplémentaire.

De cette optimisation, on déduit les fonctions de demande des biens $x_1^*, x_2^*, \ldots, x_n^*$ dont les deux principaux arguments sont le revenu non salarial exprimé en termes d'un panier de biens dont l'indice de prix est p, et les prix relatifs des biens en termes de ce même indice de prix. Toutes les variations qui ne sont pas la conséquence d'un changement de prix relatif ou de revenu non salarial réel, sont attribuées à un changement de goûts, G.

$$(4) \qquad x_1^* = d_i^*\left(\frac{R}{p}, \frac{p_1}{p}, \frac{p_2}{p}, \ldots, \frac{p_n}{p}, G\right) \text{ pour tout } i = 1 \text{ à } n.$$

Ces trois facteurs – prix relatif, revenu non salarial réel et goûts – expliquent alors le comportement du consommateur. De la variation des prix relatifs et du revenu réel non salarial, on déduit la loi la plus importante en économie : celle de la demande. On peut l'exprimer de la façon suivante : si un bien est normal, c'est-à-dire si sa consommation augmente quand le revenu réel, non salarial, augmente, alors une hausse (respectivement une baisse) du prix relatif de ce bien diminue (respectivement augmente) la quantité consommée.

Son application est d'une extrême généralité. Quand le prix relatif du divorce baisse, le revenu réel non salarial et les goûts étant maintenus constants, les divorces augmentent ; quand le prix relatif du transport aérien baisse, toutes choses égales par ailleurs, les aéroports deviennent encombrés ; quand le prix relatif du crime augmente, toutes choses égales par ailleurs, le nombre de crimes diminue.

Ce modèle traditionnel a été, très tôt, étendu à trois types d'arbitrage :

- entre consommation et loisir ou travail (L. Robbins et J. Hicks) ;

- entre consommation présente et future (I. Fisher) ;
- entre biens associés à des événements aléatoires (K. Arrow).

Ce sont ces extensions que l'on retrouve systématiquement dans les manuels standards de microéconomie. On peut en effet réduire les choix entre travail et revenu, les choix intertemporels et dans l'incertain, à un arbitrage habituel en changeant de manière appropriée la structure des biens (G. Debreu). On différencie les services rendus par les produits ou les personnes non seulement par leur localisation dans l'espace (une nuit dans un hôtel de la Best Western Union à Paris ou à Lille n'est pas le même bien) ou par la date à laquelle on les consomme (ce week-end ou le week-end dans les quinze jours qui suivent), mais aussi par leur « localisation » dans l'espace des « états de nature ». Manger une glace quand il fait beau est un service différent que manger une glace quand il pleut. Commençons par les extensions aux choix intertemporels et dans l'incertain.

Première extension : les choix intertemporels

Les biens x_1 à x_N sont repérés par leur « localisation » dans l'espace temporel grâce à la date à laquelle on les consomme : $t_1, t_2, \ldots, t_T$. Représentons par le tableau suivant l'ensemble des biens :

	Biens dans l'espace intertemporel			
Dates	**1**	**2**	**…**	***N***
t_1	$x_1(t_1)$	$x_2(t_1)$	…	$x_N(t_1)$
t_2	$x_1(t_2)$	$x_2(t_2)$	…	$x_N(t_2)$
…	…	…	…	…
t_T	$x_1(t_T)$	$x_2(t_T)$	…	$x_N(t_T)$

En ligne, on mesure à chaque période, la consommation de tous les biens de 1 à N, et en colonne, la consommation de chaque bien à chaque date, c'est-à-dire le profil temporel de la consommation du bien 1 à N de $t_1, t_2, \ldots, t_T$.

On peut construire un tableau semblable pour les prix :

	Valeur présente des prix dans l'espace intertemporel			
Dates	**1**	**2**	**...**	**N**
t_1	$P_1(t_1)$	$P_2(t_1)$		$P_N(t_1)$
t_2	$\frac{P_1(t_2)}{1+i}$	$\frac{P_2(t_2)}{1+i}$	...	$\frac{P_N(t_2)}{1+i}$
...	...			
t_T	$\frac{P_1(t_T)}{(1+i)^{T_t-1}}$	$\frac{P_2(t_T)}{(1+i)^{T_t-1}}$	...	$\frac{P_N(t_T)}{(1+i)^{T_t-1}}$

On peut alors définir à la date t_i un panier de biens :

$$X_{1N}(t_i) = [x_1(t_i), x_2(t_i), \ldots, x_N(t_i)]$$

et définir un vecteur de prix des biens correspondant :

$$P_{1N}(t_i) = [\frac{P_1(t_i)}{(1+i)^{t_i-1}}, \frac{P_1(t_i)}{(1+i)^{t_i-1}}, \ldots, \frac{P_N(t_i)}{(1+i)^{t_i-1}}]$$

De manière identique, regardons une colonne. La colonne donne le montant du bien i que l'on consommera à chaque période.

On peut donc définir un vecteur quantité du bien 1 :

$$X_{1T} = [x_1(t_1), x_1(t_2), \ldots, x_1(t_T)]$$

et un vecteur prix :

$$P_{1T} = [P_1(t_1), \frac{P_1(t_1)}{(1+i)}, \ldots, \frac{P_1(t_i)}{(1+i)^{t_i-1}}, \ldots, \frac{P_1(t_T)}{(1+i)^{t_T-1}}]$$

qui donne la valeur actuelle du prix du bien 1 aux différentes périodes.

Définissons un panier de biens qui est l'ensemble de consommation :

$$X=[x_1(t_1),\ldots,x_1(t_T)\,;\,x_2(t_1),\ldots,x_2(t_T)\,;\,\ldots\,;\,x_N(t_1),\ldots,x_N(t_T)]$$

avec un vecteur prix :

$$P=[P_1(t_1),\ldots\frac{P_1(t_T)}{(1+i)^{t_T-1}};P_2(t_1),\ldots,\frac{P_2(t_T)}{(1+i)^{t_T-1}};P_n(t_1),\ldots,\frac{P_N(t_T)}{(1+i)^{t_T-1}}]$$

Nous avons ramené l'arbitrage dans l'espace des choix intertemporels à un arbitrage ordinaire. Il nous suffit d'ajouter la contrainte de ressource à chaque période et une fonction d'utilité qui prend en compte toutes les consommations à chaque période.

Ce qui nous intéresse dans l'arbitrage intertemporel, c'est l'allocation d'un panier de biens quelconque x_1 au fil du temps. C'est le profil temporel optimal de la consommation de ce panier de biens. Nous avons alors une contrainte de budget pour le panier de bien x_1 qui s'écrit :

$$\text{(a)}\qquad \sum_{t=1}^{T} P_1(t).X_1(t)=W$$

où W mesure la richesse de l'individu qui est la somme des revenus présents et futurs, actualisée au taux d'intérêt du marché :

$$R_{t1}+\frac{R_{t2}}{(1+i)}+\ldots+\frac{R_{tT}}{(1+i)^{t_T-1}}$$

Posons une fonction de satisfaction :

$$\text{(b)}\qquad U=u[(x_1(t_1),x_1(t_2),\ldots,x_1(t_T)]$$

La maximisation de (b) sous la contrainte de (a) donne la condition d'optimum suivante :

$$\text{(c)}\qquad \frac{Um[x_1(t_1)]}{p_1(t_1)}=\frac{Um[x_1(t_2)]}{\frac{p_1(t_2)}{(1+i)}}=\ldots=\frac{Um[x_1(t_T)]}{\frac{p_1(t_T)}{(1+i)^{t_T-1}}}$$

L'euro dépensé en valeur présente dans le bien *i* doit rapporter la même utilité marginale à toutes les périodes. On peut procéder de manière similaire avec les choix dans l'incertain.

Seconde extension : les choix dans l'incertain

Les biens, cette fois, sont repérés par leur « localisation » dans l'espace des « états de nature ». Nous l'avons dit, manger une glace quand il fait beau est un service différent que de manger une glace quand il pleut. L'espace S des biens contingents est établi sur la base que ces états de nature sont mutuellement exclusifs. Si l'on a N biens et S états de nature, nous aurons NS biens contingents avec NS prix, s'il existe NS marchés où l'on peut échanger une situation incertaine contre une situation moins incertaine, ou au contraire plus risquée, sur chacun de ces marchés. Cette approche en termes d'espace des biens contingents a été introduite par K. Arrow en (1953). Définissons $X_{i,s}$ le montant de service rendu par le bien *i* dans l'état de nature *s*. On peut alors représenter dans le tableau suivant l'ensemble des biens contingents :

	Biens contingents			
État de nature	**1**	**2**	**...**	**N**
1	x_{11}	x_{12}	…	x_{N1}
2	x_{12}	x_{22}	…	x_{N2}
...	…	…	…	…
S	x_{1S}	x_{2S}	…	x_{NS}

Posons P_{is} le prix d'un bien contingent, on aura un tableau similaire pour les prix.

Regardons une ligne. Dans l'état de nature *s*, la ligne donne tous les biens que l'on obtient dans l'état *s*. On peut alors définir dans l'état de nature *s* un panier de biens contingents :

$$x_s = [x_{1S}, x_{2S}, \ldots, x_{NS}]$$

et définir :

$$P_s = [P_{1S}, P_{2S}, \ldots, P_{NS}]$$

un vecteur de prix des biens contingents correspondant. De manière identique, regardons une colonne. La colonne donne le montant du bien *i* que l'on recevra selon chaque état de nature. On peut donc définir :

$$x_i = [x_{i1}, x_{i2}, \ldots, x_{iS}]$$

et un vecteur de prix :

$$P_i = [P_{i1}, P_{i2}, \ldots, P_{iS}]$$

qui donne le prix du bien dans les différents états de nature.

Définissons le panier de biens $x = [x_1, x_2, \ldots, x_S]$. C'est alors l'ensemble de consommation : $x = [x_{11}, \ldots, x_{N1}, x_{12}, \ldots, x_{N2}, x_{1S}, \ldots, x_{NS}]$ avec un vecteur prix $P = [P_{11}, \ldots, P_{N1}, P_{12}, \ldots, P_{N2}, P_{1S}, \ldots, P_{NS}]$. Nous avons ramené l'arbitrage dans l'espace des biens contingents à un arbitrage ordinaire. Un panier de biens *X* acheté par un consommateur est composé d'une variété de produits par exemple, un parapluie quand il pleut, un chapeau quand il y a du soleil, un vélo s'il y a une grève du métro, etc. Le coût de ce panier de biens est simplement les quantités multipliées par les prix.

Nous avons alors une contrainte de budget qui s'écrit :

$$\text{(A)} \quad \sum_S P_s x_s \leq \sum_S P_s e_s$$

où e_s mesure le revenu ou la dotation initiale de l'individu dans l'état de nature *s*. Maintenant, représentons la fonction de satisfaction. Elle s'écrit :

$$\text{(B)} \quad U = \sum_S \pi_s u(x_s)$$

où π_s est une probabilité subjective qui donne l'estimation de l'apparition de l'état de nature par l'individu lui-même. Par définition, la fonction d'utilité est indépendante des états de nature.

La maximisation de (B) sous la contrainte (A) donne la condition d'optimum suivante pour le bien *i*, qui est le théorème fondamental des choix risqués selon Arrow :

$$\frac{\pi_1 Um(x_{i1})}{P_{i1}} = \frac{\pi_2 Um(x_{i2})}{P_{i2}} = \dots = \frac{\pi_s Um(x_{iS})}{P_{iS}}$$

L'euro dépensé dans le bien *i* doit rapporter la même utilité marginale espérée dans tous les états de nature.

Revenons sur cette formalisation des choix individuels. Pour ramener les choix dans l'incertain ou les choix intertemporels à des choix traditionnels, il faut supposer l'existence de marchés qui permettent de transférer des biens d'une période à l'autre.

Dans le cas de l'arbitrage intertemporel, ce marché doit permettre de transférer ses revenus présents vers ses revenus futurs, et *vice versa*. Cela nécessite l'existence d'un marché des fonds prêtables où l'on peut passer des contrats tels que l'on prête son revenu courant ou on emprunte sur ses revenus futurs. Or, c'est ce marché qui détermine les prix actualisés. Si un tel marché des fonds prêtables n'existe pas, le profil intertemporel de la consommation est donné par celui de la répartition des revenus dans le temps puisque chacun est contraint de consommer ses revenus courants !

Il en va de même avec les choix dans l'incertain. Il faut pouvoir échanger une situation certaine contre une incertaine ou *vice versa*, une situation incertaine contre une certaine pour chaque état de nature. Si pour chaque état de nature existe un marché, les économistes évoquent la notion de *complétude des marchés*. Mais il n'existe pas nécessairement autant de marchés qu'il y a d'états de nature. Il y a alors *incomplétude des marchés*.

S'il est possible de créer un titre élémentaire qui permet de s'assurer partiellement ou totalement contre un risque particulier à un état de nature,

alors l'échange de ce titre sur un marché donne naissance aux marchés financiers. Le marché financier lui-même sera dit complet s'il est possible d'assurer *tous* les états de nature. Les marchés des options et des dérivés sont une façon de « compléter » des marchés financiers incomplets. Supposons que nous ayons trois états de nature 1, 2, 3, mais simplement un actif qui rapporte un rendement différent r_1, r_2, r_3 dans chaque état de nature : les marchés sont incomplets, puisqu'il y a plus d'états de nature que d'actifs financiers. On peut compléter le marché en créant deux options en fonction de l'apparition attendue ou non des différents états de nature. Le rendement de l'option *p* achetée au prix P*p* est :

$$r_p = [max(0, r_1 - P_p),\ max(0, r_2 - P_p),\ max(0, r_3 - P_p)]$$

et celle *q* achetée au prix P_q :

$$r_q = [max(0, r_1 - P_q),\ max(0, r_2 - P_q),\ max(0, r_3 - P_q)]$$

On a alors trois titres et trois états de nature.

Cette formalisation a donc un double inconvénient : elle masque le fait que l'essentiel est dans le non-dit, à savoir dans l'existence de marchés complets à partir desquels l'arbitrage s'opère, et elle ne nous donne pas une compréhension réelle de la signification du système de prix relatif qui permet cette allocation. Pour pallier ces deux inconvénients, les manuels standards ramènent ces arbitrages dans un espace à deux dimensions. C'est ce que nous allons faire en prenant comme seul exemple les choix intertemporels.

L'exemple des choix intertemporels

Nous avons tous le choix d'aller au cinéma ce week-end ou de retarder cette consommation de quelques mois, en profitant du temps libre des vacances. L'idée est très simple. Il est toujours possible de consommer un panier de biens dans la période courante ou de reporter cette consommation à une période ultérieure en s'abstenant de consommer dans le présent.

Le revenu dont on dispose à chaque période limite la consommation courante. Il peut alors être astucieux de s'abstenir de consommer dans l'année pour épargner une fraction de son revenu courant, et ainsi augmenter ses revenus futurs du produit des intérêts. Avec ces revenus supplémentaires, on peut augmenter sa consommation dans l'année suivante. À l'inverse, on peut vouloir réduire la consommation future, et avec le revenu ainsi libéré, emprunter de l'argent pour augmenter sa consommation présente. Cet arbitrage entre présent et futur n'est pas fondamentalement différent du choix entre deux paniers de biens ou services à la même période. Cependant, il est essentiel à la compréhension des phénomènes liés à l'évolution de la consommation, de l'épargne, de l'investissement et des marchés financiers. C'est la raison pour laquelle il est étudié systématiquement dès les premières années d'enseignement en sciences économiques.

Posons x_1 la consommation d'un panier de biens à la période 1, le présent, et x_2 la consommation du même panier de biens à la période 2, le futur. Nous avons une contrainte de budget par période : une pour le présent, et une autre pour le futur. L'individu dispose d'un revenu présent R_1, et espère dans le futur un revenu R_2. Il n'est pas obligé de consommer la totalité de son revenu présent. Le revenu ainsi conservé peut être prêté à quelqu'un sur un marché des fonds prêtables et rapporter un taux d'intérêt *i*. Appelons S (*saving*) ce prêt. Le revenu futur va alors s'accroître du montant épargné auquel on ajoute le produit des intérêts. L'individu ou le ménage peut aussi emprunter sur ces revenus futurs pour accroître sa consommation présente, il émet un titre qu'il vend sur le marché des fonds prêtables contre une promesse de remboursement dans le futur du capital emprunté avec intérêts. Appelons B (*bond*) cet emprunt. Par définition, le montant maximal que l'on puisse emprunter sur ses revenus futurs est égal à $R_2 - iB = B$, soit $\frac{R_2}{1+i} = B$; incidemment, nous venons de définir la valeur présente d'un revenu futur qui n'est pas autre chose que la quantité maximale

que l'on peut emprunter pour disposer de ce revenu futur dans le présent. Prenons deux périodes, et écrivons les contraintes de budgets de chaque période :

$$(5) \quad P_1x_1 + S = R_1 + \frac{B}{1+i}$$

$$(6) \quad P_2x_2 + B = R_2 + S(1+i)$$

Posons $S = 0$ et tirons B de (6), $B = R_2 + S(1+i)$. Substituons dans (5), on obtient : $P_1x_1 = R_1 + \frac{R_2 - P_2x_2}{1+i}$. Mettons les dépenses d'un côté et les ressources de l'autre, on obtient la contrainte de budget intertemporelle :

$$(7) \quad P_1x_1 + \frac{P_2x_2}{1+i} = R_1 + \frac{R_2}{1+i} = W$$

Du côté gauche de cette équation, on mesure la valeur présente des consommations présente et future, et du côté droit, la valeur présente des revenus présents et futurs, toutes deux actualisées au taux d'intérêt nominal du marché des fonds prêtables. La somme des revenus présent et futur actualisée au taux d'intérêt nominal du marché des fonds prêtables est appelée, par convention, la richesse *W*.

Les ménages ou les individus sont supposés maximiser la satisfaction qu'ils tirent, à chaque période, de la consommation du panier de biens auquel on s'intéresse (panier de biens qui peut recouvrir l'ensemble de la consommation du ménage). Cette satisfaction est représentée par la fonction d'utilité suivante :

$$(8) \quad U = u(x_1, x_2)$$

Le profil optimal de consommation intertemporelle, $x_1^\star$, $x_2^\star$, est obtenu en maximisant (8) sous la contrainte (7) ; les montants désirés de prêt $S^\star$ et d'emprunt $B^\star$ sont obtenus en maximisant (8) sous les contraintes (5) et (6).

Revenons à la consommation et à l'optimisation intertemporelle.

$$(9) \quad \text{Max de } u(x_1,x_2) \text{ tel que } P_1x_1 + \frac{P_2x_2}{1+i} = R_1 + \frac{R_2}{1+i} = W$$

Représentons, dans l'espace consommation présente et consommation future, cet arbitrage. Le maximum que l'on puisse consommer dans le présent est obtenu lorsque l'on emprunte la totalité de son revenu futur :

$$(10) \quad x_1\max = \frac{R_1}{P_1} + \frac{R_2}{P_1(1+i)}$$

Le maximum que l'on puisse consommer dans le futur est obtenu lorsque l'on prête la totalité de son revenu présent :

$$(11) \quad x_2\max = \frac{R_2}{P_2} + \frac{R_1(1+i)}{P_2}$$

Représentons dans la figure suivante ces consommations maximales :

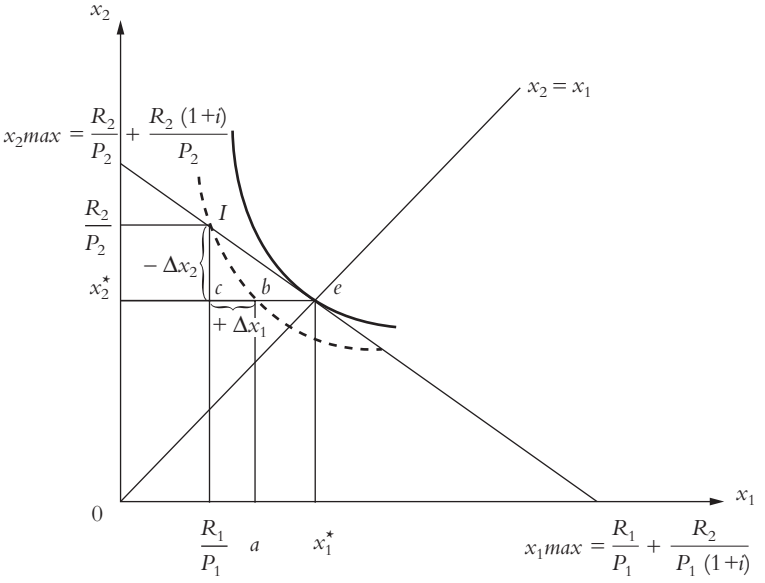


Figure 1 – Position d'emprunteur

Le point I donne la distribution initiale des revenus présent et futur. Dans l'exemple pris, les revenus futurs attendus sont plus élevés que ceux présents ; le ménage est un jeune ménage qui se trouve en position d'emprunteur. En ce point, son niveau de satisfaction est mesuré sur la courbe d'indifférence *Ib*. Le lieu des points où les individus égalisent leur consommation d'une période sur l'autre est représenté par la bissectrice. Imaginons que le ménage désire égaliser sa consommation présente et future d'une période sur l'autre. Au point I, lorsque le jeune ménage sacrifie sa consommation future d'un montant $-\Delta x_2$, il exige $+\Delta x_1$ ou cb de consommation présente en contrepartie pour maintenir constante sa satisfaction. Approximons ce ratio au point I par le taux marginal de substitution dans les préférences $\frac{+\Delta x_1}{-\Delta x_2} = \frac{Um(x_2)}{Um(x_1)}$ et comparons celui-ci au taux marginal de substitution à l'échange. En renonçant à une unité de consommation future $-\Delta x_2$, le ménage obtient en contrepartie $+\Delta x_1$ ou ce de consommation présente. En renonçant à 1 euro de consommation future mesuré en valeur présente $\frac{P_2}{(1+i)}$, il obtient P_1 de consommation présente. Le taux marginal de substitution à l'échange s'écrit pour l'emprunteur $\frac{+\Delta x_1}{-\Delta x_2} = \frac{P_2}{P_1(1+i)}$, comme la distance *ce* excède *cb*, le ménage emprunte la somme $B^* = x_1^\star - \frac{R_1}{P_1}$.

Il est impératif de procéder de même avec un ménage en position de prêteur. Cette fois, le ménage (ou l'individu) s'attend à des revenus futurs plus faibles que ceux présents. C'est un ménage d'actifs dont l'âge approche de la retraite. Lui aussi désire stabiliser sa consommation d'une période sur l'autre. En renonçant à $-\Delta x_1$ de consommation présente, il maintient sa satisfaction constante si, en contrepartie,

il obtient un montant de consommation future égale à + Δx_2, distance ab dans la figure 2.

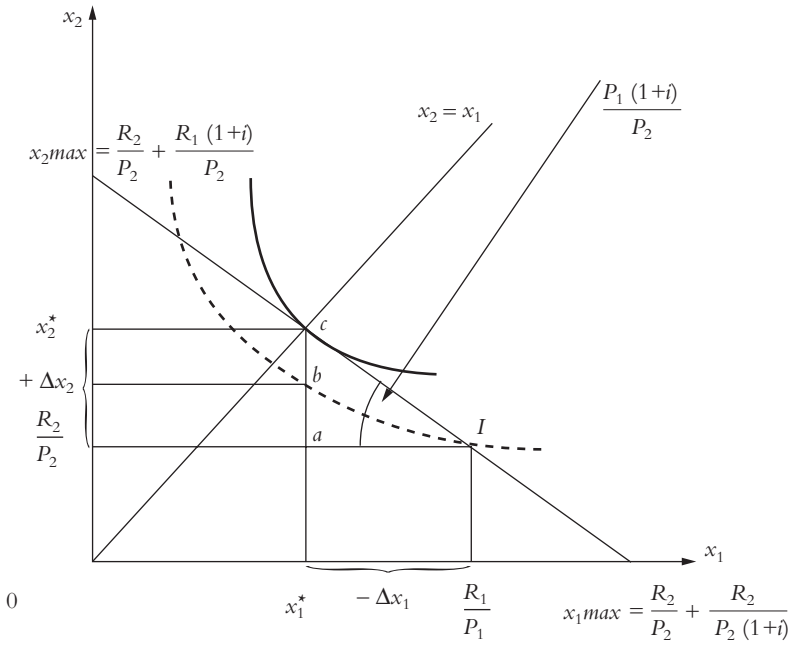


Figure 2 – Position de prêteur

Supposons que – Δx_1 corresponde au sacrifice d'une unité de consommation présente. L'euro économisé peut alors être placé sur le marché des fonds prêtables et rapporter un taux d'intérêt nominal *i*, de telle sorte que dans la période future cet euro vaut $\frac{P_1(1+i)}{P_2}$ et permet de consommer + Δx_2. Dans l'exemple choisi + Δx_2 = la distance *ae*, celle-ci excède le montant exigé *ab* pour maintenir constante la satisfaction, donc le ménage épargne jusqu'à ce que les deux taux soient identiques et prête un montant $S^* = \frac{R_1}{P_1} - x_1^*$.

Le coût d'opportunité de la consommation présente $\frac{P_1(1+i)}{P_2}$ exprimé en valeur future est un concept très important, c'est le prix relatif de la consommation présente en termes de consommation future. Il joue un rôle essentiel, comme tout prix relatif, dans la détermination de la consommation désirée présente et future comme dans les montants désirés de prêts et d'emprunt. Arrêtons-nous un moment sur ce concept.

Nous pouvons transformer ce prix relatif de la façon suivante, en introduisant la variation des prix du panier de biens considéré d'une période sur l'autre. Par définition, le prix du panier de biens en période 2 ou dans le futur est égal à P_2 ; il est égal aussi par définition à $P_2=(1+\pi)$, où π est le taux d'inflation. Le prix relatif de la consommation présente en termes de consommation future est alors égal au ratio $\frac{(1+i)}{(1+\pi)}$ que nous posons égal à $(1 + r)$, où r sera par convention appelé le taux d'intérêt réel monétaire du marché des fonds prêtables. Le taux d'intérêt réel est donc bien le prix relatif de la consommation présente en termes de consommation future.

Quand le taux d'intérêt réel augmente, le prix de la consommation présente augmente relativement à la consommation future, ce qui incite les ménages à retarder la consommation du panier de biens considéré à la période suivante, et donc à la reporter dans le futur (par exemple l'achat d'une automobile ou d'un autre bien durable ou d'équipement, y compris le logement). Cette hausse induit une baisse de la consommation et un ralentissement de l'activité économique Quand ce taux d'intérêt baisse, il incite, au contraire, les ménages à avancer la consommation de leur panier de biens « futur » à la période présente, ce qui stimule la consommation présente et l'activité économique. Par ailleurs, ce taux d'intérêt réel est déterminé par la confrontation de l'offre de prêts et de la demande d'emprunts sur le marché des fonds prêtables et, comme pour tout système de prix, le taux

d'intérêt réel a pour rôle de coordonner les plans des individus sur ce marché des fonds prêtables, c'est-à-dire qu'à un certain taux les emprunteurs trouvent, dans la période, des prêteurs pour pouvoir réaliser leurs plans.

À partir du taux d'intérêt réel $\frac{(1+i)}{(1+\pi)} = (1+r)$, on déduit la relation célèbre d'I. Fisher entre le taux d'inflation et le taux d'intérêt nominal. Développons l'identité précédente : $\frac{(1+i)}{(1+\pi)} = (1+r)$, soit $(1+i) = (1+r)(1+\pi)$ ou $(1+i) = 1+\pi+\pi r$ et $(i) = r+\pi+\pi r$. Négligeons le terme croisé, πr qui est le produit de deux pourcentages (terme qui est donc plus petit que r ou π), nous avons $i \approx r+\pi$, le taux d'intérêt nominal est approximativement égal à la somme du taux d'intérêt réel et du taux d'inflation. Le point important à retenir est le suivant : c'est le taux d'intérêt réel monétaire qui gouverne les comportements de prêts, d'emprunts et de consommation, et non le taux d'intérêt nominal du marché des fonds prêtables.

De cette optimisation, on déduit les fonctions de demande des paniers de biens x_1^*, x_2^* à chaque période, et celles des demandes dérivées de prêts S^* et d'emprunts B^* :

$$x_1^* = x(W, r, G)$$

$$x_2^* = x(W, r, G)$$

$$S^* = s\left(\frac{R_1}{P_1}, \frac{R_2}{P_2}, r, G\right)$$

$$B^* = s\left(\frac{R_1}{P_1}, \frac{R_2}{P_2}, r, G\right)$$

Les consommations présente et future s'accroissent avec la richesse W ; en revanche, le montant des prêts augmente avec une hausse du revenu présent et diminue avec une hausse du revenu futur. *A contrario*, le montant des emprunts diminue avec une hausse du revenu présent et augmente avec une hausse du revenu futur.

Une hausse du taux d'intérêt diminue, sans ambiguïté, la consommation présente pour les emprunteurs et, par conséquent, le montant désiré des emprunts. Il augmente, sans ambiguïté, la consommation future des prêteurs, puisque les effets revenus et substitutions de la hausse du taux d'intérêt vont dans le même sens. En revanche, l'effet d'une hausse du taux d'intérêt est indéterminé sur la consommation présente, et donc aussi sur le montant des prêts car les effets substitution et revenu vont en sens contraire. Pour ne pas avoir une courbe d'offre de prêt coudée, il faut introduire l'effet participation de telle sorte que l'offre agrégée soit croissante en fonction du taux d'intérêt réel.

Voir l'Annexe 1 (Le marché financier) en ligne.

DE LA THÉORIE TRADITIONNELLE DES CHOIX INDIVIDUELS À LA NOUVELLE

L'analyse économique développée jusqu'ici traite le revenu comme une donnée exogène. On parle de revenu non salarial. Mais cette hypothèse est réductrice, car il est toujours possible d'obtenir davantage de revenu en sacrifiant son temps de loisir. Le revenu est donc composé d'un revenu non salarial qui n'exige aucun sacrifice en contrepartie, et d'un revenu salarial qui, lui, au contraire, exige de sacrifier du temps ; or, le temps dont nous disposons dans une journée est limité. La contrainte de budget de l'arbitrage traditionnel :

$$(2) \qquad R = \sum_{i=1}^{n} p_i x_i$$

doit se réécrire :

$$(i) \qquad R + wT_w = \sum_{i=1}^{n} p_i x_i$$

où wT_w mesure le revenu salarial et R le revenu non salarial

(ii) $\quad T = T_w + T_l$

qui exprime la contrainte de temps avec T_w le temps de travail salarié, T_l le temps de loisir et T la contrainte de temps de la période considérée, semaine, mois, cycle de vie. En substituant T_w dans (i) tirée de (ii) on obtient ce que l'on appelle une contrainte de revenu de plein temps :

$$\text{(iii)} \quad R + wT = wT_l = \sum_{i=1}^{n} p_i x_i$$

Le revenu de plein temps est dépensé en biens de consommation mais aussi en loisir. Si l'individu tire une satisfaction du temps passé en loisir, nous avons une fonction de satisfaction dont les arguments sont les biens $(x_1, x_2, \ldots, x_n)$ et le temps de loisir , soit T_l :

(iv) $\quad U(x_1, x_2, \ldots, x_n, T_l)$

L'arbitrage loisir-consommation est obtenu très simplement en maximisant (iv) sous la contrainte (iii) :

$$\text{(v)} \quad \frac{Um(x_1)}{P_1} = \frac{Um(x_2)}{P_2} = \ldots = \frac{Um(x_i)}{P_i} = \ldots \frac{Um(x_n)}{P_n}$$

$$= \frac{Um(T_l)}{w}$$

L'euro sacrifié en prenant une heure de loisir doit rapporter autant de satisfaction supplémentaire que l'euro dépensé dans l'achat d'un panier de biens de consommation. Mais comme le suggère la contrainte de temps, une heure de loisir en plus, c'est une heure de travail salarié en moins. La contrepartie de la demande de loisir est l'offre de travail. Nous ne traiterons pas de cet arbitrage parce qu'il est traité ailleurs, en revanche, nous pouvons généraliser cette contrainte de temps à l'ensemble des activités qui consomment du temps.

Un examen attentif d'un budget temps démontre que le sommeil, les soins personnels, les soins aux enfants, les tâches ménagères, les courses, le bricolage, les repas, sont des activités distinctes du loisir proprement dit tels la lecture, le temps passé à regarder la télévision, les séjours à la campagne ou les soirées passées au théâtre, au cinéma ou dans les enceintes sportives qui consomment du temps. On peut même dire que la consommation des biens et services comme les repas, les soins personnels, ou l'exercice de n'importe quelle activité, prend du temps. La contrainte de temps joue donc un rôle fondamental dans la vie, et n'a pas été prise en compte dans la théorie traditionnelle du consommateur avant l'article fondamental de Gary Becker (1965) qui a donné naissance à la nouvelle théorie de la consommation.

L'idée de fond de cette nouvelle théorie de la demande est simple et particulièrement féconde. Les individus ne tirent pas leur satisfaction des biens et services qu'ils achètent sur le marché, mais d'activités ou de biens plus fondamentaux. Ces activités sont produites par le consommateur lui-même, d'une part à l'aide de biens et services et d'autre part en utilisant son propre temps. Dans cette vision, la demande de biens et services achetés sur le marché ou la demande de loisir sont, en fait, des demandes dérivées de facteurs de production d'une activité plus fondamentale produite par l'individu. Consommer un repas ou prendre des loisirs nécessitent non seulement l'achat de biens sur le marché, mais aussi du temps. Illustrons cette nouvelle théorie avec un exemple simple.

L'arbitrage entre deux loisirs

Prenons l'arbitrage d'un étudiant entre des leçons de tennis et des séances de cinéma. Aller au cinéma ou prendre une leçon de tennis sont des activités qui nécessitent du temps et l'achat de biens et services. Pour aller au cinéma, il faudra acheter un billet pour accéder à la salle où est projeté le film. Il faudra acheter ou louer un DVD et un lecteur si vous désirez voir ce film chez vous. Pour jouer au tennis il

faudra louer le court et payer un droit d'entrée au club, ensuite acquitter le prix des leçons de tennis, sans compter l'équipement : raquette, balles, chaussures, etc.

Posons X_c, X_t les services ou les biens de consommation servant à produire les activités C (cinéma) et T (tennis) : aller au cinéma ou prendre une leçon de tennis.

Posons t_t, t_c le temps passé à consommer : prendre votre leçon de tennis et visionner le film dans une salle de cinéma ou chez vous. C, T mesurent les services rendus par le temps passé à l'activité considérée t_t, t_c lorsque ce temps est combiné avec l'achat de biens ou services X_c, X_t sur le marché.

Nous admettrons pour simplifier les relations suivantes :

(a) $t_c = a \cdot C$

(b) $t_t = d \cdot T$

(c) $X_c = b \cdot C$

(d) $X_t = f \cdot T$

Elles définissent une technologie de consommation à coefficients constants. Nous avons alors les contraintes de budget et de temps suivantes :

(e) $P_c.X_c + P_t.X_t = w.t_w + M$

(f) $t^0 = t_c + t_t + t_w$

où t_w mesure le temps passé à travailler et t^0 est le temps total disponible. Substituons t_w par sa valeur tirée de (f) dans (e) :

(g) $P_c.X_c + P_t.X_t + w.t_t + w.t_c = w.t^0 + M$

Comme $t_t = d.T$, $t_c = a.c$, $X_c = b.C$ et $X_t = f.T$, on peut réécrire (g) de la façon suivante :

(h) $(P_c.b + w.a).C + (P_t.f + w.d).T = w.t^0 + M$

C'est la contrainte de revenu et de temps généralisée.

On peut réécrire cette contrainte en définissant le prix implicite de l'activité tennis ou cinéma par respectivement $(P_t.f + w.d) = \pi_t$ et $(P_c.b + w.a) = \pi_c$, puis le revenu de plein temps par $w.t^0 + M = S$. Nous avons alors $\pi_c.C + \pi_t.T = S$. On maximise une fonction d'utilité $U = u(C,T)$ sous la contrainte de temps généralisée $\pi_c.C + \pi_t.T = S$. Ce qui nous permet d'appliquer tout ce que nous avons appris sur la théorie de la demande pourvu que l'on réinterprète correctement les prix et les revenus. Là encore, les économistes s'efforcent de rapporter cette nouvelle allocation du temps au modèle traditionnel du consommateur.

Le coût d'opportunité des biens de loisirs en termes de biens de consommation n'est pas mesuré par le ratio des prix monétaires relatifs, mais par le ratio des prix implicites des deux activités. Le prix implicite est la somme d'un coût monétaire, $P_c.b$ et d'un coût d'opportunité du temps $w.a$. Revenons à notre arbitrage entre séance de cinéma et leçon de tennis. Dans l'espace des biens et services T, C la droite de budget a désormais pour équation :

$$\text{(i)} \qquad T = \frac{w.t^0 + M}{P_t.f + w.d} - \frac{P_c.b + w.a}{P_t.f + w.d}.C$$

Le coût d'opportunité de la séance de cinéma en termes de leçons de tennis est :

$$\text{(j)} \qquad \frac{P_c.b + w.a}{P_t.f + w.d}$$

L'effet revenu est toujours mesuré par la hausse du revenu *non salarial*, c'est-à-dire par le revenu qui n'exige aucun sacrifice en contrepartie. Si l'on ne précise pas qu'il s'agit d'une hausse du revenu non salarial, mais du revenu, on peut s'attendre à un effet prix. En effet, une hausse du revenu salarial par suite d'une hausse du taux de salaire non seulement accroît le revenu de plein temps, mais modifie le prix relatif des activités très intensives en temps. On observe à la fois un effet revenu et un effet prix relatif. Puisque le cinéma est plus intensif en temps, la

hausse du coût du temps mesurée par w rend plus onéreuse la séance de cinéma par rapport à la leçon de tennis. Plus vous êtes riche, plus la séance de cinéma vous coûtera comparée à la leçon de tennis, moins vous fréquenterez les salles de cinéma. Les consommateurs déplaceront leurs demandes vers des activités substituts peu intensives en temps.

L'arbitrage entre loisirs et consommation n'est plus alors qu'un cas particulier d'une technologie de consommation plus générale. Les services rendus par les loisirs sont produits essentiellement avec du temps, alors que les services rendus par les biens le sont principalement avec des biens ! Une telle réduction dans l'analyse n'est pas tenable puisque les activités de consommation comme de loisir font appel à la fois au temps et à des biens et services achetés sur le marché ou produits chez soi. Cependant, l'esprit de cet arbitrage est toujours utile. L'idée que l'on dépense son revenu de plein temps entre dépenses monétaires pour acheter des biens et services, et revenus perdus en consacrant son temps à d'autres activités que le travail salarié, est féconde. Elle permet de faire apparaître de nouvelles substitutions. Quand le salaire augmente, et ce pour une utilité constante, l'effet de substitution implique que les consommateurs déplacent leurs activités vers celles qui sont intensives en biens et services et peu intensives en temps.

Les déterminants des revenus perdus dépendent :

- de l'intensité temporelle de l'activité ;
- du coût du temps (le coût du temps varie souvent en fonction de l'activité elle même, si celle-ci affecte la productivité de l'individu, et de la période à laquelle on la pratique, si cette activité prend place dans les périodes de vacances) ;
- du progrès technique, souvent biaisé dans le sens d'une économie de temps.

Une augmentation du progrès technique qui réduit les intensités temporelles des activités de consommation implique à la fois un effet prix et un effet revenu réel de plein temps. Les prix relatifs des activités

intensives en temps diminuent, incitant les consommateurs à se tourner vers celles-ci, ce qui pousse les individus à moins travailler. Mais cet accroissement de productivité entraîne un effet revenu réel qui compense cet effet substitution.

Pour mieux percevoir l'intérêt de cette nouvelle approche de la consommation en termes de production d'activité, prenons un exemple : la baisse de la fécondité.

L'exemple de la baisse de la fécondité

La corrélation négative observée entre le nombre d'enfants par famille et le revenu familial a toujours intrigué les démographes mais aussi les économistes, car en termes de théorie de la demande, cela signifie que les enfants sont des « biens inférieurs ». En effet, plus le pays devient riche, plus la population stagne ! Ce qui contredit la théorie de Malthus. Le prix Nobel G. Becker, dans les années 1960, a vivement contesté la théorie de Malthus. En effet, celle-ci prédisait que la hausse du revenu entraînerait une hausse du nombre moyen d'enfants par famille. Cet argument reposait sur deux idées :

- un accroissement du revenu diminue le taux de mortalité infantile, c'est-à-dire permet d'avoir un plus grand nombre d'enfants survivants à l'âge adulte. Si la baisse de la fécondité ne compense pas la baisse du taux de mortalité, le nombre d'enfants moyen par famille augmente ;
- un accroissement du revenu permet aux jeunes de se marier plus tôt et donc d'avoir une période de fécondité plus longue, ce qui incite les femmes à avoir plus d'enfants.

Aujourd'hui, la hausse du revenu est sans effet sur le taux de mortalité, et celui-ci vraisemblablement est sans effet sur le nombre moyen d'enfants, car les individus veulent un nombre donné d'enfants survivants à l'âge adulte. Si un enfant meurt en bas âge, il est aussitôt remplacé. Le taux de fécondité qui est le nombre moyen désiré d'enfants

par femme baisse lorsque le revenu familial augmente. Plus exactement, les statistiques démontrent que l'on observe une relation en J inversé entre le taux de fécondité et le revenu.

La théorie classique du consommateur nous enseigne que la hausse du revenu, toutes choses égales par ailleurs, entraîne en moyenne une hausse des dépenses dans chaque bien selon les goûts des individus mesurés *ex post* par l'élasticité revenu des biens considérés. Certains produits, tels la margarine ou les spectacles de cinéma, sont moins consommés lorsque le revenu croît. Ils sont dits « biens inférieurs ». La tentation est grande alors d'affirmer que les enfants sont aussi des biens inférieurs puisque plus le revenu augmente, plus le nombre d'enfants par famille diminue.

L'erreur vient de ce que les statisticiens, qui mesurent la corrélation entre revenu et nombre d'enfants, prennent en compte le revenu salarial et non salarial ou le revenu familial. Or le revenu salarial ou le revenu de la femme mesure essentiellement le coût d'opportunité du temps à se consacrer aux enfants et ce, d'autant plus que la variance du nombre d'heures de travail dans la population est faible. Le diagnostic de « bien inférieur » doit être prononcé à partir de la relation entre revenu *non salarial* ou le revenu du mari et nombre d'enfants ! Avec un tel test, la relation entre nombre d'enfants et revenu *non salarial* ou revenu du mari n'est plus négative mais positive.

La famille ne tire pas une utilité de la naissance en soi d'un enfant, mais des services ou de l'utilité qu'il va produire ou procurer à ses parents tout au long de sa vie. Si l'on reprend la métaphore de G. Becker entre un bien durable et un enfant, on voit immédiatement que le nombre d'enfants est mis en balance avec la qualité de l'enfant. Il n'est pas nécessairement important d'avoir un grand nombre de réfrigérateurs à la maison, en revanche, il peut être intéressant d'avoir un réfrigérateur d'une grande qualité ou dont les caractéristiques, à la fois comme congélateur et réfrigérateur, rendent un grand service. Or, plus le revenu augmente, plus les familles dépensent de l'argent pour acheter

des biens durables de meilleure qualité tels qu'un logement plus confortable, une voiture plus sûre et plus rapide, ou une machine à laver la vaisselle. Ils agissent de même avec les enfants. Les dépenses consacrées à l'enfant augmentent avec le revenu. Entre deux variétés d'enfants – l'un bien éduqué et l'autre sans éducation – les familles riches préfèrent des enfants bien éduqués. Mais un enfant avec un haut niveau d'éducation coûte cher. Un arbitrage s'instaure entre le nombre d'enfants et la qualité par enfant. L'effet revenu pousse les familles à avoir moins d'enfants.

À cet effet revenu, on peut ajouter un effet prix à la hausse. Élever un enfant a un coût. Outre les dépenses nécessaires à son éducation, il faut sacrifier du temps. Le prix du temps joue un rôle fondamental. Une femme qui reste à domicile pour élever ses enfants prive d'un revenu important l'ensemble de la famille. Plus le revenu perdu sur le marché du travail s'élève, moins elle est incitée à rester au foyer. Or, une fois active, elle ne dispose plus de temps pour s'occuper d'un grand nombre d'enfants et les élever. Elle renonce alors à avoir plusieurs enfants ou un enfant supplémentaire. Les couples où les femmes sont actives, quel que soit le niveau du revenu du chef de famille, ont en moyenne moins d'enfants.

Paradoxalement, la fameuse courbe en J inversé de la fécondité traduit non pas un effet revenu mais un effet prix ! Quand les taux de salaire sont bas et qu'ils augmentent, l'effet prix du salaire de la femme l'emporte sur l'effet revenu réel de la hausse du salaire du mari ; à un taux de salaire élevé, l'effet revenu réel du salaire du mari l'emporte sur l'effet substitution du salaire de l'épouse.

La théorie classique du consommateur n'est donc pas sans utilité pour comprendre un phénomène social aussi important que celui de la baisse de la fécondité dans nos sociétés occidentales.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 3

Système de prix et équilibre

Edmond Malinvaud

Professeur honoraire au Collège de France

Le principal programme de recherche de la théorie microéconomique a concerné l'équilibre général néoclassique. Sa pièce centrale reste la détermination des prix dans une économie régie par les échanges sur des marchés parfaitement concurrentiels. Sous certaines conditions, cet équilibre concurrentiel est efficace au sens de Pareto. Les développements les plus récents de la théorie microéconomique visent à répondre à la question : que reste-t-il de cette propriété d'efficacité dans les économies réelles de marché ? Ce chapitre situe des développements qui sont examinés dans les autres chapitres traitant de la théorie microéconomique.

Ce chapitre est axé sur le principal programme de recherche de la théorie microéconomique, ce qui va le conduire à évoquer les quatre chapitres de la partie : les comportements individuels, l'équilibre général – objet central du présent chapitre –, la concurrence considérée dans l'économie industrielle, les biens publics et les mécanismes de coordination conçus comme structurant les politiques publiques.

Les théories microéconomiques, et plus généralement la théorie économique, ont été conçues et élaborées grâce à un certain nombre d'abstractions qui peuvent aider à la compréhension des phénomènes et problèmes qui en définissent le champ. On commencera ici par les concepts de « système de prix » et d'« équilibre ». Chemin faisant, on

sera amené à rencontrer les définitions des « biens », de la « concurrence » « parfaite » ou « oligopolistique », de l'« efficacité au sens de Pareto », des « défaillances du marché », de l'information des agents.

L'ÉQUILIBRE GÉNÉRAL

Biens, système de prix

La théorie économique, particulièrement la théorie microéconomique, s'intéresse surtout aux « institutions des économies de marché », où un rôle central est joué par les échanges de biens sur les marchés. Des agents, principalement des producteurs et/ou consommateurs, y achètent et vendent des biens à certains prix.

Un « bien » est habituellement défini comme un objet ou service particulier. Le terme évoque l'intérêt qu'un acquéreur peut tirer de la possession de l'objet ou de l'accès au service, le prix d'acquisition pouvant être conçu comme une mesure de l'intérêt en cause. Quand divers biens sont vendus sur un marché, on dit souvent qu'ils s'y échangent. On peut appeler « termes de l'échange » les rapports entre les prix respectifs, interprétés alors comme des rapports entre les valeurs que les acquéreurs attribuent aux divers biens ou services. On peut aussi parler du système de prix établi sur le marché, le terme de système évoquant des relations mutuelles entre les termes de l'échange.

La « théorie des prix » vise à exhiber les facteurs de détermination du système de prix en cause. Elle fait évidemment intervenir les comportements individuels des agents opérant sur le marché, ainsi que le lecteur l'a vu dans le chapitre précédent.

Dans cette théorie, chaque bien est caractérisé par la propriété selon laquelle deux quantités égales du bien sont parfaitement équivalentes. Ainsi, deux objets matériellement identiques sont considérés comme deux biens différents s'ils sont disponibles à deux dates différentes et si

la date de disponibilité importe dans l'analyse. C'est pourquoi l'étude des choix intertemporels, examinée dans le chapitre précédent, fait intervenir les dates de disponibilité et distingue autant de prix pour un objet donné qu'il y a de dates d'échange concevables. Il y a ainsi congruence entre le système de prix, la liste des dates et celle des biens.

Concurrence entre agents et équilibre

La théorie des prix vise à expliquer quels équilibres se réalisent entre les prix dans les économies de marché. Pour cela, il faut préciser les institutions qui régissent les marchés et les comportements des agents qui y opèrent. Une hypothèse s'avère à cet égard particulièrement commode, celle de la « concurrence parfaite ».

Pour bien en percevoir la portée et en comprendre la définition, une citation de l'économiste anglais John Hicks (cf. bibliographie) s'avère opportune ici. Il écrit à la fin de son introduction :

> *« Une limitation de notre analyse va vite être très évidente… Nous allons retenir d'un bout à l'autre l'hypothèse de concurrence parfaite ; ceci veut dire que nous allons presque toujours négliger l'influence sur l'offre qui peut résulter de calculs faits par les vendeurs quant à l'influence sur les prix que leurs propres ventes pourraient avoir (de même pour l'influence sur la demande). En fait, beaucoup d'offres et de demandes sont probablement quelque peu influencées par de tels calculs ; il se peut que ces influences soient très importantes. Mais, il est très difficile de bien tenir compte de cette influence, dans tout autre que les problèmes les plus simples. Ainsi, bien que l'analyse de cet ouvrage serait certainement améliorée si plus d'attention était portée à la concurrence imparfaite, j'ai pensé préférable de laisser cela de côté pour le moment. Je ne crois pas, personnellement, que les résultats les plus importants de cette œuvre seraient beaucoup affectés par cet oubli, mais je m'en tiens ici à dire que le sujet devrait évidemment être étudié le moment venu. »*

La même limitation affecte la quasi-totalité des ouvrages consacrés à l'équilibre général. Les deux termes de ce concept demandent à être précisés. Par « équilibre du marché » d'un bien donné, on entend une situation dans laquelle le prix est compatible avec la demande de tous les acheteurs opérant sur ce marché comme avec l'offre de tous les vendeurs. On parle souvent d'un « équilibre partiel ». Un équilibre général suppose la considération simultanée de tous les marchés envisagés dans une analyse portant sur tout un ensemble de biens. On s'intéresse aux cas où tous les équilibres partiels sont réalisés sur tous les marchés. Caractériser l'équilibre général est alors indiquer quels déterminants assurent la réalisation simultanée de tous les équilibres partiels.

Un tel objectif est évidemment ambitieux. On comprend qu'il ait intéressé une longue lignée d'économistes mathématiciens depuis Léon Walras (1834-1910) jusqu'à Gérard Debreu (1921-2004) et bien d'autres.

Efficacité de l'équilibre et objectifs de la théorie

La question vient naturellement à l'esprit de savoir à quoi sert la théorie de l'équilibre général. Mais avant de tenter d'y répondre, il convient de compléter ce qui vient d'être dit. Car, avec une assez grande généralité, l'équilibre général de concurrence parfaite a une propriété d'efficacité qui a été bien formulée et établie par l'économiste italien Vilfredo Pareto (1848-1923).

Afin de définir cette propriété, il faut au préalable dire ce que l'on entend par un « état de l'économie » où producteurs et consommateurs opèrent. Comme on l'a vu dans le chapitre précédent, la théorie classique a retenu depuis ses débuts deux hypothèses :

- premièrement, le *comportement rationnel des agents*, les producteurs maximisant leurs profits et les consommateurs maximisant chacun une fonction de l'utilité des quantités de biens qu'ils consomment et des quantités de travail qu'ils fournissent, les uns et les autres sous

les contraintes techniques ou physiques qui s'imposent à eux. Ainsi sont déterminées leurs offres et leurs demandes individuelles sur les marchés des biens et des diverses catégories de travail ;

- deuxièmement, la *loi de l'offre et de la demande* exige que les marchés s'équilibrent pour tous les biens et toutes les catégories de travail. Sur chacun de ces marchés, le total des offres individuelles égalise le total des demandes.

Un état de l'économie se caractérise par les prix des divers biens, y compris les catégories de travail, et par les offres et demandes individuelles. Cet état est dit *possible* s'il résulte de comportements rationnels et respecte la loi de l'offre et de la demande sur tous les marchés. Un état possible de l'économie est ainsi, quasiment par définition, un équilibre général.

La contribution de Pareto fut de montrer que, sous certaines hypothèses sur lesquelles nous reviendrons dans notre seconde partie, un équilibre général de concurrence parfaite pouvait aussi être dit *optimal*, d'où l'expression « optimal au sens de Pareto », plus tard remplacée par « efficace au sens de Pareto ». Quoi qu'il en soit, un tel état est non seulement possible, mais aussi tel qu'on ne puisse pas trouver un autre état possible de la même économie (mêmes fonctions d'utilité et mêmes fonctions de production) qui soit jugé meilleur par la fonction d'utilité d'un des consommateurs sans être jugé moins bon par celle d'aucun autre consommateur. Cette propriété est évidemment intéressante, mais elle passe sous silence la possibilité qu'un autre état possible de la même économie puisse être jugé plus juste (c'est en quoi parler d'optimalité est en fait abusif).

La question vient naturellement à l'esprit de savoir quels peuvent être les objectifs de la théorie de l'équilibre général, notamment parce que cette question a été source de confusions quand la théorie en cause a été perçue par certains comme visant à justifier le capitalisme. En fait, elle a surtout son intérêt comme ayant fourni aux économistes une multiplicité de concepts, qui sont tous essentiels comme instruments pour les réflexions sur les économies de marché.

Des théories de l'équilibre général ont été conçues afin d'éclairer des domaines variés où des interdépendances entre prix jouent un rôle important, tels sont les cas par exemple du commerce international ou de l'économie publique.

La section qui suit va le montrer à propos du commerce international. De son côté, l'économie publique attribue, comme nous le verrons dans la dernière section du chapitre, une importance méritée aux calculs économiques propres à orienter le choix des équipements publics. Elle fait souvent référence à un résultat en somme adverse de celui établi par Pareto. Étant donné un état A efficace au sens de Pareto, peut-on lui associer un système de prix qui lui soit conforme, en ce sens que, confronté aux prix en question, aucun agent, producteur ou consommateur, ne gagnerait à dévier du comportement que A lui attribue ? La théorie de l'équilibre général donne une réponse positive à cette question, à la condition que les données de l'économie, fonctions de production et d'utilité, satisfassent une propriété qui sera examinée ici dans la section « Les défaillances du marché », mais simplement qualifiée pour le moment de « convexité », à l'intention des lecteurs qui l'auraient déjà rencontrée. Le système des prix en cause est alors qualifié de « dual » de l'équilibre. Il peut servir à des calculs économiques dans un monde de décisions décentralisées concernant la gestion des services publics ou la réglementation des activités privées.

Théorie néoclassique du commerce international

L'étude économique du commerce international est née de controverses qui mettaient en exergue des aspects différents des situations dans lesquelles les pays étaient supposés se trouver respectivement les uns par rapport aux autres.

Un pays peut, en effet, soit disposer de ressources naturelles plus ou moins favorables, soit avoir accès à des technologies plus ou moins productives, ou encore héberger des ménages plus ou moins avides de consommer certains biens rares. Ces pays pouvaient envisager de tirer

parti de politiques commerciales plus ou moins protectionnistes. La théorie du commerce international envisagea, dès lors, de supposer le « libre-échange », c'est-à-dire l'abolition de tout contrôle des prix et des échanges aux frontières.

Poussant l'hypothèse à l'extrême, des théoriciens en vinrent à concevoir un monde dans lequel les marchés des biens et des facteurs de production engloberaient l'ensemble de tous les pays, et où la théorie de l'équilibre général s'appliquerait directement, par exemple au cas où K pays pourraient chacun produire n biens à partir de m facteurs de production, puis échanger entre eux sur des marchés où régnerait la concurrence parfaite.

En fait, l'histoire de la théorie du commerce international (cf. bibliographie, F. Etner) est assez complexe jusqu'à ses derniers développements. Le souci de répondre à certaines préoccupations réelles amena à imaginer des contextes dans lesquels l'auteur comparait le libre-échange à l'autarcie. Dans son ouvrage fondamental publié en 1817, David Ricardo établit le principe des « avantages comparatifs » en prenant l'exemple de l'Angleterre et du Portugal, le premier pays ayant une productivité relativement élevée dans la production des draps, le second dans celle du vin. C'est en exportant des draps et en important du vin que l'Angleterre retirait le plus grand avantage du commerce extérieur en termes de monnaie. Traitant du même sujet en 1848, John Stuart Mill prit pour référence non plus la monnaie mais le revenu national réel, égal au volume de la consommation qu'il caractérisa par la moyenne géométrique des volumes de biens consommés : celle-ci, identifiée à l'utilité des biens, devait croître si l'autarcie était remplacée par le libre-échange.

Au 20^{e} siècle, la théorie dite de Heckscher-Ohlin apporta un prolongement à ces propositions, auquel nombre d'économistes mathématiciens contribuèrent (cf. bibliographie). Grâce à un petit nombre d'hypothèses, le système des prix dominant le commerce international se révélait avoir des propriétés intéressantes. Outre l'hypothèse du

libre-échange des biens, régi par la concurrence parfaite, il faut citer la supposition de facteurs de production propres à chaque pays et de productions s'effectuant avec des rendements d'échelle constants. Divers théorèmes furent démontrés aux termes desquels le commerce international influencé par les avantages comparatifs induisait *une tendance à l'égalisation des prix des facteurs de production entre pays*, donc aussi à l'égalisation des revenus qu'ils généraient et à divers autres effets caractérisés.

Il y a ainsi une similitude avec certains résultats du modèle de l'équilibre général néoclassique tel que formulé notamment par G. Debreu (cf. bibliographie).

Il apparut cependant que le commerce international ne se révélait empiriquement pas conforme à ces propositions théoriques. Assez naturellement, la théorie abandonna le modèle de Heckscher-Ohlin. Ainsi l'ouvrage très lu de E. Helpman et P. Krugman (cf. bibliographie) prêta plus d'importance aux économies d'échelle et aux structures des marchés des produits dans un contexte favorisant la concurrence monopolistique, dont il sera brièvement question dans la section ci-après.

AUTRES VISIONS DU MARCHÉ

Alternatives à la concurrence parfaite

Il serait erroné de croire que les économistes aient accepté sans réserve les hypothèses retenues par la concurrence parfaite. La plupart d'entre eux savaient au contraire que les institutions et structures réelles des économies de marché en différaient notablement. Dès les années 1950, deux réactions se manifestèrent : soit examiner de près les hypothèses de la spécification retenue par l'école néoclassique ; soit rechercher une autre spécification du secteur des entreprises, c'est-à-dire construire ce qui fut appelé l'« économie industrielle ».

La première réaction se manifesta de la part de deux membres de l'école néoclassique, contemporains de Hicks et Debreu, dans un livre très lu à l'époque (*cf.* bibliographie). Kenneth Arrow et Frank Hahn procédèrent à une investigation détaillée du corps d'hypothèses retenu dans le modèle standard. Quant à l'économie industrielle, elle poursuivit une voie de recherche qui avait été engagée dès le milieu du 19e siècle avec les analyses du monopole (une entreprise face à un marché) et de l'oligopole (quelques entreprises face à un marché).

Considérant l'hypothèse de la concurrence parfaite, structurante dans la théorie des prix, et les multiples hypothèses qui ont été proposées pour formaliser la concurrence imparfaite, l'alternative la plus fructueuse vis-à-vis des objectifs de cette théorie semble être aujourd'hui celle de la « concurrence monopolistique » incorporée dans divers modèles mathématiques de l'équilibre général.

Elle suppose non seulement qu'il existe une seule entreprise productrice de chaque bien échangé sur chaque marché actif, mais aussi que la concurrence joue entre les entreprises pour établir quels marchés existeront, de sorte que la diversification des produits s'en trouvera déterminée et efficace à l'équilibre. Au total, cette hypothèse établit un pont fructueux entre la théorie microéconomique et la théorie macroéconomique, ses agrégats et ses marchés.

La démarche qui vient d'être décrite est assez semblable à celle de l'économie industrielle, par opposition à celle qui fut retenue par les artisans de la théorie de la concurrence imparfaite. Selon la voie ouverte par Cournot (1801-1877) et conduisant à traiter d'abord du monopole puis de l'oligopole, l'accent est bien mis sur les producteurs en tant qu'entreprises, dites encore firmes. Mais leur liste est fixée, l'apparition possible de nouveaux producteurs étant ignorée. La concurrence porte alors sur le partage d'un marché où les demandes sont en quelque sorte prédéterminées.

Au contraire, selon une voie plus inspirée par l'observation de la révolution industrielle du 19e siècle, l'essentiel réside dans les comportements stratégiques des firmes, certaines déjà installées sur les marchés

et d'autres susceptibles d'y entrer. Cette voie se garde bien de négliger les barrières, soit naturelles soit légales, à l'entrée (*cf.* la distinction entre les deux approches avec Paolo Sylos Labini et Jean Tirole dans la bibliographie).

Les défaillances du marché

Revenons maintenant sur l'efficacité au sens de Pareto. L'énoncé rigoureux de la propriété qui a été évoquée dans la section « Concurrence entre agents et équilibre » s'appelle aujourd'hui « Deuxième théorème fondamental de l'économie de bien-être ». Par économie de bien-être, il faut entendre précisément le modèle retenu par l'économie néoclassique, encore qualifié « modèle de Arrow-Debreu ». Par premier théorème, il faut entendre : « Tout équilibre concurrentiel est efficace au sens de Pareto. » Le deuxième théorème peut alors s'énoncer simplement : « Sous l'hypothèse de convexité de l'économie, tout état efficace au sens de Pareto est décentralisable en équilibre, car on peut lui associer un système de prix avec lequel cet état est un équilibre concurrentiel. »

Qu'entendre alors par hypothèse de convexité ? La convexité est évidemment une propriété géométrique, mais qui se traduit en termes analytiques. Pour un producteur qui opère avec n biens, consommés en tant que facteurs de production (les « inputs » ou « entrants ») ou produits pour être vendus (les « outputs »), l'opération de production peut être figurée par un vecteur de l'espace à n dimensions, les inputs apparaissant comme des composantes négatives et les outputs comme composantes positives. Les contraintes techniques imposent à ce vecteur d'appartenir à un « ensemble des possibilités de production ». Or, précisément, l'hypothèse dont il est question maintenant retient que cet ensemble est convexe.

La figure 1 ci-après s'applique au cas où n'interviendraient que deux biens, celui de l'output y, dont la quantité est comptée verticalement, et celui de l'input x mesuré horizontalement. La zone hachurée repré-

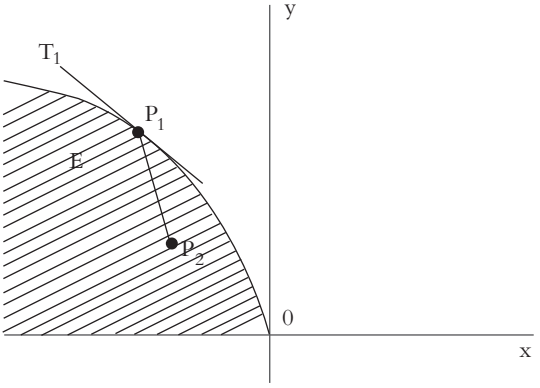


Figure 1 – Convexité de l'ensemble *E* des possibilités de production

sente l'ensemble E des possibilités de production. Les points P_1 et P_2 figurent deux combinaisons productives possibles. L'ensemble E de telles combinaisons est convexe sur la figure : tout point sur le segment de droite P_1 P_2 correspondrait à une autre combinaison possible. La « fonction de production » correspondant à cette figure s'écrirait $y = f(x)$, la fonction f étant définie par la frontière nord-est de l'ensemble E. Elle produirait la supposition bien naturelle que le producteur ne gaspille aucune part de l'input x ; le point de coordonnées (x, y) se situant sur la frontière de E. Notons en passant que la productivité marginale correspondante serait égale à la dérivée $f''(x)$ et que la convexité de E se traduirait alors par *la décroissance de cette productivité marginale*, la dérivée seconde $f'(x)$ étant négative ou nulle.

Le lecteur n'a pas de difficulté à imaginer comment cette figure se transposerait dans l'espace à trois dimensions qui correspondrait au cas où la production impliquerait deux inputs x_1 et x_2 avec leurs productivités marginales respectives $f_1'(x_1,x_2)$ et $f_2'(x_1,x_2)$.

Dans le deuxième théorème de l'économie de bien-être, la convexité concerne aussi les préférences des consommateurs et leurs fonctions d'utilité. La figure 2 se prête à des considérations analogues à celles

permises pour la figure 1. Pour la commodité de l'exposé, considérons le cas d'un individu qui consomme une quantité x_1 d'un bien consommable et fournit une quantité x_2 de travail (mesurée verticalement sur la figure). Ses préférences concernant le couple (x_1,x_2) sont représentées sur la figure 2 par deux courbes d'indifférence U^A et U^B ; un couple (x_1,x_2) figuré par le point M_3 de U^B lui permettant de travailler moins ou de consommer plus que ce que permettent les couples représentés par les points M_2 ou M_1 de U^A. L'hypothèse de convexité, appliquée à la convexité de la fonction d'utilité $U(x_1,x_2)$, s'exprime par la convexité de l'ensemble des combinaisons (x_1,x_2) qui sont préférées ou au moins équivalentes à celles représentées par les points de la courbe d'indifférence U^A.

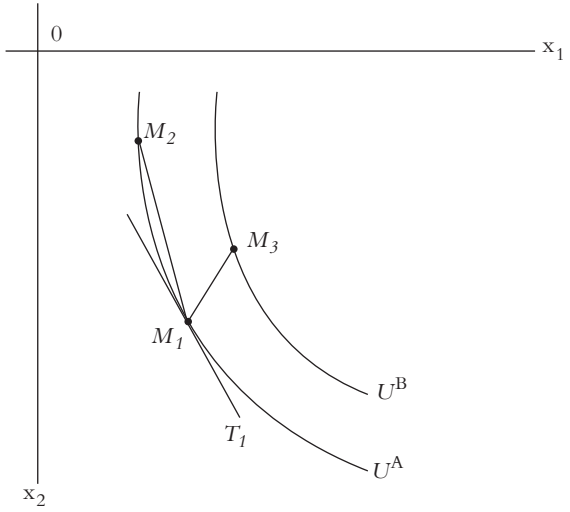


Figure 2 – Convexité de la fonction d'utilité $U(x_1,x_2)$ d'un consommateur

Notons encore que le système des prix grâce auquel l'état efficace peut être décentralisé est révélé aussi bien par l'ensemble des possibilités de production que par la courbe d'indifférence à l'équilibre constaté : le rapport des prix p_2/p_1 est révélé aussi bien par la tangente T_1P_1 à l'ensemble des productions possibles que par la tangente T_1M_1 à la courbe d'indifférence U^A.

Les hypothèses retenues dans le deuxième théorème de l'économie de bien-être sont cependant fortes. On le constate si on examine les économies de marché réelles dans leur complexité. On parle de « défaillances du marché » pour dénommer d'une manière générale les cas dans lesquels l'équilibre concurrentiel n'est pas efficace au sens de Pareto. Les écarts par rapport à la convexité, notamment du fait des économies d'échelle dans la sphère productive, en fournissent de premiers exemples. Une autre source importante de telles défaillances (*cf.* bibliographie, Bernard Salanié, 1er titre) concerne les *effets externes*.

On dit qu'il y a un *effet externe*, ou une *externalité*, quand les actions d'un agent influencent *directement* les possibilités de choix (donc l'ensemble des possibilités de production ou les préférences) d'un autre agent. Les *biens publics*, tels les services collectifs, fournissent un cas extrême de telles externalités : ils sont produits par des agents publics et consommés par de nombreux agents privés – le plus souvent sans que la consommation par l'un n'affecte guère la consommation par d'autres. Les pollutions entrent, à sens inverse, dans la même catégorie : ceux qui en supportent les coûts diffèrent de ceux qui les produisent.

Politiques publiques, information et mécanismes de coordination

Dans de nombreux cas, les défaillances du marché peuvent en théorie être palliées par des politiques publiques appropriées. Étudier de tels cas constitue un développement important de la théorie microéconomique. Nous allons nous en tenir ici au cas de la fourniture des consommations collectives, un cas qui fait bien apparaître la nécessité de prolongements théoriques.

Soit une économie contenant n individus ($i = 1,\ldots,n$) et deux biens, l'un dit « privé » consommable par tous, x_i étant la quantité consommée par i. Soit X la quantité totale disponible pour de telles consommations. Soit aussi le surplus x :

$$[1] \quad x = X - \sum_{i=1}^{n} xi$$

Supposons qu'il serve à produire, en quantité z, un bien public disponible pour tous, la fonction de production étant :

$$[2] \quad z = f(x)$$

L'utilité obtenue par l'individu i est :

$$[3] \quad U_i(x_i,z) \quad (i = 1,\ldots, n)$$

où U_i est une fonction croissante de x_i et de z.

Un état possible de l'économie est caractérisé par $n + 1$ quantités $x,x_2,\ldots,x_n,z$ qui satisfont les deux équations [1] et [2]. Si les fonctions f et U_i ont toutes la propriété de convexité, tout état possible est efficace au sens de Pareto dès lors qu'il existe un prix p_z et n prix p_i $(i = 1,\ldots,n)$ tels que :

$$[4] \quad \frac{p_i}{p_z} = \frac{U_z}{U_i} \quad (i = 1,\ldots,n)$$

$$[5] \quad p_z = \sum_{i=1}^{n} p_i$$

p_z peut alors être interprété comme le prix du bien public et p_i, égal à $p_z U'_z/U'_i$, comme l'utilité marginale qu'a le bien public pour l'individu i relativement à sa consommation de bien privé x_i. Si le système des prix $(p_z,p_1,\ldots,p_n)$ était observable, l'équilibre efficace pourrait être dit « décentralisable ».

La question se pose alors de savoir si un mécanisme de décentralisation pourrait être institué pour conduire à ce résultat. C'est concevable (*cf.* bibliographie, Bernard Salanié, 1er titre) si un organisme central peut obtenir connaissance des prix relatifs individuels p_i/p_z. Mais cela peut paraître artificiel, car l'information requise ne viendra pas spontanément, à moins que les individus consommateurs y aient intérêt.

Il existe toutefois, dans le cadre de la théorie des contrats, un prolongement significatif concernant justement les mécanismes de révélation, que l'on peut encore qualifier être des mécanismes de coordination puisqu'ils sont organisés en vue de la révélation d'informations utiles à la coordination d'activités exercées par des agents distincts (*cf.* bibliographie, Bernard Salanié, 2e titre).

Ce prolongement lui-même pourrait apparaître à certains comme assez exotique. Nous avons cependant vu ici depuis le début que la recherche de conceptualisations nouvelles était une spécialité importante de la théorie microéconomique, au risque qu'elle paraisse parfois exotique, comme le furent d'ailleurs certaines de ses applications à la théorie du commerce international que nous avons rencontrées dans la section « Théorie néoclassique du commerce international ».

Il faut savoir d'ailleurs que reconnaître les défaillances des marchés, souvent liées aux difficultés de transmission de l'information, a constitué un mobile fort pour l'extension de la théorie microéconomique vers des champs que l'école néoclassique avait longtemps négligés, de même que son économie industrielle avait longtemps négligé les percées de la « théorie des jeux » qui lui ont cependant été fort utiles.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 4

Concurrence et économie industrielle

THIIERRY PÉNARD

Professeur d'économie à l'Université de Rennes 1, CREM

L'économie industrielle s'intéresse à l'organisation des marchés et des entreprises, pas seulement dans l'industrie, mais aussi dans les services et le commerce. La question centrale en économie industrielle est de comprendre comment une entreprise peut obtenir du pouvoir de marché. Précisément, le pouvoir de marché renvoie à la capacité d'une entreprise à obtenir des niveaux de marges élevés sur les produits ou services qu'elle vend. Les réponses apportées à cette question présentent un intérêt pour les décideurs d'entreprise en charge de la stratégie et du marketing, mais aussi pour les décideurs publics. L'économie industrielle permet, en effet, d'éclairer les politiques de régulation des marchés (politique de la concurrence, réglementations sectorielles), les politiques industrielles, d'innovation et d'aménagement du territoire.

DE LA CONCURRENCE PURE ET PARFAITE À LA CONCURRENCE IMPARFAITE

Les premiers travaux d'économie industrielle se sont développés dans les années 1930 et après la Seconde Guerre mondiale, autour du cou-

rant structuraliste, appelé aussi école de Harvard, dont les figures emblématiques sont Joe Bain et Edward Mason. Ces travaux sont essentiellement empiriques et s'articulent autour du paradigme Structure-Comportement-Performance (SCP). Selon ce paradigme ou schéma d'analyse, les conditions de base et la structure d'un marché (nature de l'offre, de la demande, caractéristiques des produits, conditions d'entrée…) déterminent les comportements des entreprises (investissements, politique de prix, publicité…) qui eux-mêmes déterminent les performances de ce marché (profits, qualité des produits…). Les économistes se réclamant de l'école de Harvard privilégient les études de cas et études sectorielles pour inférer l'impact des structures de marché sur les comportements et la profitabilité des entreprises.

Ces travaux vont faire l'objet de critiques à partir des années 1970, pour leur manque de fondements théoriques et de rigueur sur le plan méthodologique. Avec l'essor de la théorie des jeux et des techniques statistiques et économétriques, l'économie industrielle connaît un profond renouvellement dans les années 1980-1990. La théorie des jeux devient la boîte à outils de référence, qui va servir aux économistes industriels à modéliser les comportements et interactions stratégiques des entreprises sur les marchés et à établir des propositions testables. La démarche consistant à construire des modèles théoriques et à les tester empiriquement est au cœur de la *nouvelle économie industrielle*, que l'on appelle également théorie de l'organisation industrielle.

La *nouvelle économie industrielle* offre aussi une relecture stratégique du paradigme SCP, en mettant l'accent sur le caractère endogène des structures de marché. C'est l'idée que la structure d'un marché peut être modifiée de façon stratégique par les entreprises. Par exemple, une fusion aura pour effet de renforcer la concentration de l'offre sur le marché.

Enfin, la *nouvelle économie industrielle* se démarque ouvertement du modèle de concurrence pure et parfaite, qui ne constitue pas une description réaliste et pertinente des marchés. Les conditions d'une concurrence pure et parfaite sont les suivantes :

- une atomicité de l'offre et de la demande, qui se traduit par une absence de pouvoir de marché des vendeurs et des acheteurs ;
- une homogénéité des biens offerts sur le marché, qui se traduit par une substituabilité parfaite des vendeurs ;
- une libre entrée sur le marché pour les vendeurs et les acheteurs, et une libre circulation des facteurs de production ;
- des vendeurs et des acheteurs parfaitement informés.

La réalité des marchés montre un tout autre visage. Les entreprises disposent la plupart du temps d'un pouvoir de marché, en tant qu'offreurs ou acheteurs, qui se traduit par une capacité à fixer leurs prix. Par ailleurs, les entreprises concurrentes proposent rarement des biens parfaitement identiques, les différences pouvant être liées aux caractéristiques des biens ou à la localisation des entreprises sur le marché. Les entreprises n'entrent pas non plus librement sur les marchés. Il existe des barrières ou des coûts à l'entrée et à la sortie du marché. Enfin, l'information est loin d'être parfaite sur les marchés, certains acteurs étant mieux informés que d'autres. Pour ces différentes raisons, le référentiel de l'économie industrielle n'est pas la concurrence pure et parfaite, mais la concurrence imparfaite. Les modèles de concurrence imparfaite sont particulièrement adaptés pour analyser des situations d'oligopole (un nombre restreint d'offreurs sur le marché) ou de monopole (un seul offreur) : des situations que l'on retrouve dans de nombreux secteurs d'activité (du secteur automobile à celui des télécommunications, en passant par l'agroalimentaire ou la banque).

Dans la suite de ce chapitre, nous allons revenir sur les différentes caractéristiques et enseignements de la concurrence imparfaite.

POUVOIR DE MARCHÉ

Les stratégies de prix et de quantités

La concurrence sur un marché se manifeste tout d'abord par des choix en matière de prix et de production. Ces choix se font de manière

interdépendante, c'est-à-dire que chaque entreprise prend ses décisions en anticipant les décisions des autres entreprises. Il est en effet important pour une entreprise de tenir compte de ce que les concurrents font, parce que son profit dépend non seulement de ses propres choix de prix et de production, mais aussi des prix et de la production choisis par ses concurrents. Par exemple, si une entreprise anticipe que ses concurrents vont baisser leurs prix, elle a intérêt à baisser les siens pour conserver ses clients. Elle peut, de même, avoir intérêt à augmenter sa production si elle anticipe que les autres vont limiter leur production sur le marché.

Dans les modèles de concurrence imparfaite, le marché est en situation d'équilibre concurrentiel dès lors qu'aucune entreprise ne souhaite modifier sa stratégie compte tenu de la stratégie adoptée par les autres entreprises. En d'autres termes, les prix ou les quantités que l'entreprise choisit correspondent à la meilleure réponse possible aux choix de ses concurrents. Cette situation est formellement qualifiée d'équilibre de Nash.

Les deux modèles de base utilisés par les économistes industriels pour étudier la concurrence imparfaite sont le modèle de Cournot (1838) et le modèle de Bertrand (1881). La concurrence à la Bertrand est une concurrence qui porte sur les prix. Les entreprises sont supposées offrir des produits homogènes et ne pas avoir de contraintes de capacités. Si les entreprises ont les mêmes coûts de production, la situation d'équilibre se caractérise alors par des profits nuls pour les entreprises. Cette situation résulte des comportements stratégiques des entreprises qui les poussent à fixer des prix toujours inférieurs (même légèrement) à ceux des concurrents afin d'attirer l'ensemble de la demande (puisque les produits sont homogènes, donc parfaitement substituables). Au final, les entreprises se retrouvent toutes à fixer le même prix, qui est égal à leur coût marginal de production, c'est-à-dire au coût qu'elles supportent pour produire et distribuer une unité additionnelle. Même si les entreprises sont peu nombreuses sur le marché, elles se retrouvent par conséquent dans l'incapacité d'en tirer le moindre pouvoir de marché,

les marges obtenues sur chaque unité vendue étant nulles. Cette situation, qualifiée de paradoxe de Bertrand, montre que l'on peut avoir une absence de relation causale entre le nombre d'entreprises sur le marché (le degré de concentration de l'offre) et leur profitabilité. Dans une concurrence à la Bertrand, les profits seront les mêmes, que le marché comporte deux, dix ou mille entreprises.

La situation est un peu différente avec une concurrence à la Cournot dans laquelle les entreprises se concurrencent sur les quantités offertes sur le marché. Dans ce modèle de concurrence, les entreprises offrent des biens homogènes, et le prix du marché est celui qui permet de rendre compatibles les quantités offertes et les quantités demandées. Ainsi, le prix du marché baisse (*augmente*) en cas de hausse (*baisse*) des quantités offertes. Il en résulte qu'une entreprise a intérêt à réduire son volume de production, si elle anticipe que ses concurrents vont augmenter leurs volumes de production, afin de limiter la baisse du prix de marché et de préserver ses niveaux de marge. À l'inverse, si elle anticipe que ses concurrents vont diminuer leurs volumes de production, elle a intérêt à produire plus, afin d'augmenter ses parts de marché. Dans une concurrence à la Cournot, la meilleure réponse de chaque entreprise est donc de faire l'inverse de ce que les autres entreprises font, et à l'équilibre, toutes les entreprises obtiennent des profits positifs.

Le modèle de concurrence à la Cournot permet d'éclairer la relation entre structure et performance de marché puisque, à l'équilibre, le pouvoir de marché des entreprises (mesuré par leur taux de marge) tend à diminuer avec le nombre d'entreprises sur le marché. Il existe donc une relation positive entre le degré de concentration sur un marché et la profitabilité des entreprises dans le cadre d'une concurrence à la Cournot, relation que l'on ne retrouve pas dans une concurrence à la Bertrand.

Une des questions est de savoir quand il est plus pertinent d'utiliser un modèle de Cournot ou un modèle de Bertrand pour analyser un marché. Du point de vue empirique, le modèle de Cournot est plus

adapté pour analyser des secteurs d'activité qui nécessitent des investissements élevés en capacités et qui ne se prêtent pas à des révisions rapides à la hausse comme à la baisse des niveaux de production. On peut penser, par exemple, à des marchés comme l'automobile, l'acier ou la chimie. À l'inverse, les secteurs qui ne connaissent pas de contraintes de capacités (comme le secteur des services) ou qui se caractérisent par des capacités fortement excédentaires, relèvent plus d'une concurrence à la Bertrand.

Les barrières à l'entrée

Le pouvoir de marché d'une entreprise n'est pas seulement contraint par la concurrence des entreprises présentes sur le marché, mais aussi par la concurrence potentielle d'entreprises qui peuvent entrer sur le marché. Les possibilités d'entrée sur un marché renvoient à l'importance des barrières à l'entrée, c'est-à-dire aux coûts que doit supporter une entreprise pour s'installer sur un marché. Plus ces coûts sont élevés, et moins les entreprises en place sont menacées par l'entrée de nouveaux concurrents. Ces coûts d'entrée constituent également des coûts de sortie, car ils sont généralement irrécupérables ou perdus si l'entreprise décide de sortir du marché (*sunk costs*).

On distingue deux sortes de barrières à l'entrée : des barrières naturelles et des barrières stratégiques. Les barrières naturelles sont liées à la nature des produits ou à leur mode de production. Le marché présentera des barrières naturelles à l'entrée importantes si la production nécessite de maîtriser des compétences technologiques avancées, ou si la production se caractérise par de fortes économies d'échelle (c'est-à-dire si les coûts moyens des entreprises diminuent avec leur taille ou leur niveau de production). L'entrant devra par conséquent investir fortement en recherche et développement et en capacités de production pour pouvoir être aussi compétitif ou efficace que les entreprises en place. Ces dernières peuvent également accroître de manière artificielle les barrières à l'entrée, par exemple :

- en multipliant les variétés ou les marques pour couvrir tout le marché, et ne laisser aucune niche profitable pour les entrants ;
- en surinvestissant en publicité pour accroître les dépenses publicitaires que devront consentir les entrants ;
- en multipliant les dépôts de brevets pour pouvoir ensuite réclamer des redevances aux entrants qui utiliseront les technologies brevetées, ou pour les empêcher d'accéder à ces technologies ;
- en contrôlant les fournisseurs ou les sources d'approvisionnement et les clients, soit en les intégrant, soit en signant avec eux des contrats exclusifs. Il s'agit d'une stratégie de forclusion dont l'objectif est d'augmenter les coûts des concurrents existants ou potentiels.

Les barrières à l'entrée sont un des facteurs structurels clés pour expliquer les comportements et les performances des entreprises. En l'absence de barrières à l'entrée et à la sortie, le marché est considéré comme un marché contestable au sens de Baumol, Panzar et Willig (1982) ; c'est-à-dire que l'exercice d'un pouvoir de marché par les entreprises en place peut être contesté par des entreprises extérieures, qui, en entrant librement sur le marché, peuvent proposer des prix plus bas et capter de cette façon une partie des profits des entreprises en place. Sur un marché contestable, les entreprises sont donc contraintes par la concurrence potentielle à pratiquer des prix bas, et même une entreprise en situation de monopole peut détenir un très faible pouvoir de marché. En revanche, tout ce qui contribue à empêcher la libre entrée ou à augmenter les coûts de sortie renforce le pouvoir de marché des entreprises en place. C'est pour cette raison que les autorités de la concurrence accordent beaucoup d'attention aux barrières à l'entrée sur les marchés et s'efforcent d'éliminer toutes les barrières artificielles. Ces mêmes autorités sont aussi vigilantes quant à la concentration de l'offre sur ces mêmes marchés.

Concentration et politique de la concurrence

Les fusions entre des entreprises présentes sur un même marché ont pour effet d'augmenter la concentration de l'offre. Pour quelles raisons les entreprises fusionnent-elles ? Une fusion peut avoir pour objectif de faire baisser les coûts, et par conséquent d'accroître l'efficacité productive. Elle peut également permettre d'acquérir et de combiner des compétences complémentaires et d'accroître la capacité d'innovation des entreprises (c'est-à-dire l'efficacité dynamique). Dans les deux cas, l'objectif pour les entreprises qui fusionnent est de créer des synergies ou de la valeur. Une fusion peut aussi permettre d'accéder à de nouveaux marchés géographiques. Enfin, cela peut permettre de relâcher la concurrence et d'accroître le pouvoir de marché des entreprises qui fusionnent. Toutefois, ce dernier effet bénéficie aussi aux entreprises qui ne sont pas parties prenantes de la fusion. Il est d'ailleurs possible, si la fusion ne génère pas suffisamment de synergies en matière d'innovation et de coûts, qu'il soit plus profitable pour une entreprise de rester en dehors de la fusion que d'être dans la fusion. En l'absence de synergie, les entreprises qui fusionnent cessent de se concurrencer et sont moins agressives sur le marché qu'avant. Cela se traduit par des parts de marché pour la nouvelle entité qui sont inférieures à la somme des parts de marché de ces mêmes entreprises avant fusion. Mais si la fusion génère d'importantes synergies, les entreprises en dehors de la fusion se retrouveront face à un nouveau concurrent plus efficace, qui pourra pratiquer des prix plus bas, offrir de meilleurs services et par conséquent leur prendre des parts de marché.

Une fusion peut ainsi avoir des effets ambigus sur la concurrence et l'efficacité économique sur les marchés. Cela justifie la mise en place en Europe d'un contrôle des opérations de concentration. La procédure de contrôle s'applique essentiellement sur les marchés caractérisés par un niveau de concentration déjà élevé. Les entreprises doivent notifier leurs projets de fusion et fournir aux autorités de la concurrence des informations permettant d'apprécier l'effet de cette fusion sur la concurrence future. L'objectif des autorités est d'empêcher les

entreprises qui fusionnent de créer ou de renforcer une situation de position dominante individuelle (ou même collective sous la forme d'une collusion avec les concurrents restants sur le marché).

Les autorités vont procéder à un bilan concurrentiel pour déterminer si la fusion est susceptible d'avoir des effets anticoncurrentiels. Les critères pris en compte sont, entre autres, le degré de concentration de l'offre, la puissance de la demande qui peut contrebalancer le pouvoir de marché des offreurs, les barrières à l'entrée et la concurrence potentielle. Si, au terme de cette analyse, les autorités concluent à l'absence d'effets anticoncurrentiels, la fusion est autorisée. Dans le cas contraire, les autorités procèdent à un bilan économique pour déterminer si la fusion peut avoir des effets positifs en termes d'efficacité ou de progrès économique, comme la création de nouveaux services, une amélioration de la qualité des produits, des réductions de coûts ou la sauvegarde d'emplois. Si les effets positifs sont jugés insuffisants au regard des effets anticoncurrentiels, la fusion peut être interdite, comme dans le cas du projet de fusion Legrand-Schneider sur le marché des équipements électriques, refusé par les autorités européennes en 2001. Si, à l'inverse, les effets positifs font plus que contrebalancer les effets négatifs, alors la fusion est acceptée avec éventuellement l'obligation de céder certains actifs, comme dans le cas de la fusion Suez-GDF qui s'est accompagnée de cessions en Belgique pour ne pas renforcer la position déjà dominante de Suez sur le marché belge de l'énergie.

DIFFÉRENCIATION

Une des façons de relâcher la concurrence pour les entreprises est de rendre leurs produits moins substituables. Lorsque les offres sont différenciées, les consommateurs ne considèrent plus les prix comme unique critère de décision. Les caractéristiques des offres et les préférences des consommateurs sur ces caractéristiques vont aussi entrer en ligne de compte, ce qui modifie la nature et l'intensité de la concur-

rence entre les entreprises. Les économistes industriels distinguent deux principales formes de différenciation : la différenciation horizontale et la différenciation verticale. Pour saisir la distinction, prenons deux biens A et B proposés sur le même marché à des prix identiques :

- si les consommateurs sont parfaitement indifférents entre A et B, les deux biens sont considérés comme homogènes ;
- si les consommateurs préfèrent de manière unanime l'un des biens à l'autre (par exemple le bien A au bien B), alors les deux biens sont considérés comme différenciés verticalement ;
- si les consommateurs n'ont pas les mêmes préférences sur les deux biens (certains classent le bien A devant le bien B, et d'autres le bien B devant le bien A), alors les deux biens sont considérés comme différenciés horizontalement.

Comment les entreprises parviennent-elles à se différencier verticalement ou horizontalement, et ont-elles toujours intérêt à le faire ?

Différenciation horizontale

La différenciation horizontale renvoie initialement à une différenciation spatiale des entreprises (Hotelling, 1929). Un marché est un espace géographique dans lequel les consommateurs et les entreprises sont caractérisés par leur localisation. Dans sa décision d'achat, le consommateur doit prendre en compte non seulement le prix du bien, mais aussi le coût d'opportunité de se déplacer chez le vendeur (ou de se faire livrer). Ce coût est croissant avec la distance qui sépare l'acheteur et le vendeur (frais de transport, coût d'opportunité du temps de transport). Entre deux vendeurs proposant le même bien à un prix identique, mais qui ne sont pas localisés au même endroit, le consommateur choisira le vendeur le plus proche. Des différences dans les localisations des entreprises entraînent des différences dans les préférences des consommateurs. Chaque vendeur dispose ainsi d'un certain

pouvoir de marché sur les consommateurs qui sont proches de lui, et cette clientèle sera d'autant plus captive que les autres vendeurs sont éloignés et que les coûts de déplacement sont élevés.

Une autre forme de différenciation horizontale consiste pour les entreprises à jouer sur les caractéristiques ou le positionnement du produit (ses fonctionnalités, son design, ses options…). Il s'agit d'une différenciation fondée sur l'hétérogénéité des préférences ou des goûts des consommateurs (Lancaster, 1966). Les consommateurs diffèrent sur les caractéristiques du produit idéal sur le marché. Certains peuvent préférer une fonctionnalité qui n'est proposée que dans le bien A, et d'autres une fonctionnalité proposée seulement dans le bien B. Certains peuvent préférer le design moderne du bien A, d'autres le design plus traditionnel du bien B. Les consommateurs subissent, au final, une désutilité (un coût subjectif) à acheter un bien éloigné du bien qu'ils considèrent comme idéal. Ils ont intérêt à se tourner vers le bien dont ils se sentent le plus proches.

Les choix des entreprises en matière de localisation spatiale ou de positionnement de leurs offres vont être motivés à la fois par la recherche d'un niveau de marges élevé et par la conquête de nouvelles parts de marché. Ces deux objectifs peuvent entrer en contradiction. Une politique de marge suppose de s'éloigner de ses concurrents, pour disposer d'une clientèle plus captive, alors qu'une politique de parts de marché implique de se rapprocher des concurrents, pour leur prendre des clients. Mais, si toutes les entreprises adoptent cette dernière stratégie, les entreprises vont se retrouver avec un positionnement identique (c'est-à-dire une offre homogène) et une concurrence plus intense.

Quel résultat peut-on attendre d'une concurrence stratégique en différenciation horizontale ? Lorsque les coûts de déplacement ou la désutilité des consommateurs à s'éloigner de leur bien idéal sont élevés, alors l'effet *politique de marge* l'emporte sur l'effet *course à la part de marché*, et les entreprises ont tendance à se différencier au maximum. Il

est possible toutefois d'observer sur certains marchés des phénomènes d'agglomération lorsque les prix sont rigides ou nuls. Dans ce cas, l'intérêt de se différencier disparaît et les entreprises vont se concurrencer sur les parts de marché. C'est le phénomène que l'on retrouve sur le marché de la télévision gratuite financée par la publicité, le prix étant nul pour les téléspectateurs. Les chaînes sont plutôt dans une stratégie de mimétisme dans leur programmation, l'objectif étant d'obtenir le maximum d'audience.

Différenciation verticale

Des biens sont différenciés verticalement lorsque les consommateurs sont unanimes dans leurs classements de ces biens. Cette forme de différenciation tient à des différences dans la qualité des biens : à prix identique, les consommateurs préfèrent tous un bien de qualité supérieure à un bien de qualité inférieure. Dans ces conditions, pour que des produits verticalement différenciés coexistent sur un marché, il faut que les produits de qualité inférieure soient vendus moins chers que les produits de meilleure qualité.

Comment les entreprises choisissent-elles leur positionnement vertical ? Ce choix est guidé tout d'abord par la disposition à payer des consommateurs pour la qualité et le coût pour produire cette qualité. En situation de monopole, une entreprise sera amenée à choisir le niveau optimal de qualité qui lui permet de maximiser ses profits (recettes – coûts). Ce niveau de qualité peut être très élevé si le surcoût d'une qualité supérieure est plus que compensé par les recettes additionnelles. Si les consommateurs sont hétérogènes dans leurs préférences, l'entreprise peut également décider de proposer des variantes ou versions de qualité différentes : une version de base vendue à un prix bas, et une version haut de gamme vendue à un prix plus élevé.

En situation de concurrence, chaque entreprise doit aussi tenir compte des choix de ses concurrents et est incitée à proposer une qualité un peu différente des autres pour ne pas s'exposer à une concurrence trop

vive. Dans le cas d'une entreprise entrant sur le marché, son intérêt est de rechercher une position ou une niche qui n'est pas occupée par les entreprises en place. C'est la stratégie adoptée par les compagnies *low cost* comme Ryanair, qui sont entrées sur le marché du transport aérien avec un niveau de qualité inférieur à celui des compagnies historiques. Cette stratégie leur a permis d'avoir des coûts plus faibles et de pratiquer des prix plus bas, mais aussi de ne pas entrer en concurrence directe avec les compagnies en place offrant une qualité supérieure à un prix plus élevé.

Comment les entreprises parviennent-elles à se différencier verticalement ? Elles peuvent tout d'abord se différencier en proposant des biens objectivement de meilleure qualité, en termes de fiabilité ou de durée de vue, de confort, de services complémentaires (conseils et service après-vente…). Mais cette différenciation objective entraîne des surcoûts de production, liés à l'utilisation de composants plus chers, à un temps de production plus long, à une main-d'œuvre plus nombreuse et plus qualifiée… Une entreprise peut de même se différencier verticalement de façon subjective ou artificielle, en recourant au marketing et à la publicité. À travers la création d'une marque ou d'une image de marque, l'entreprise peut parvenir à convaincre les consommateurs que son produit est de meilleure qualité que les produits concurrents et le vendre plus cher, même si les produits sont objectivement de même qualité. Cette stratégie de différenciation peut être rentable, car les surcoûts associés à une différenciation subjective (en termes de dépenses publicitaires ou de marketing) sont souvent plus faibles que ceux associés à une différenciation objective (surcoûts de production). Un exemple de différenciation artificielle est fourni par les grandes marques et les marques de distributeurs comme Carrefour, Leclerc ou Casino. Les grandes marques sont généralement vendues beaucoup plus cher que les marques de distributeurs, alors que leurs produits peuvent sortir des mêmes usines. La différence de qualité objective justifie-t-elle les écarts de prix observés ? Pas toujours, mais le marketing et la publicité autour des

grandes marques ont pour effet d'accentuer artificiellement la différenciation et d'augmenter la disposition à payer des consommateurs pour les produits de grande marque.

Les stratégies de différenciation verticale peuvent aussi être un moyen de créer ou de renforcer les barrières à l'entrée sur un marché. C'est le cas lorsqu'une entreprise en place propose de nombreuses variantes ou versions de ses produits pour couvrir l'ensemble du marché, et ne laisser aucune niche profitable aux entrants. Par exemple, une entreprise installée peut multiplier les offres comportant différents niveaux de qualité de service, différentes options ou fonctionnalités. On retrouve ces mêmes comportements stratégiques en différenciation horizontale : une entreprise peut multiplier les localisations ou les variantes pour empêcher l'entrée de nouveaux concurrents. Si les stratégies de différenciation constituent un moyen *a priori* légal de relâcher la concurrence ou de dissuader l'entrée, tel n'est pas le cas des stratégies de collusion ou de cartellisation.

COLLUSION

Une collusion ou un cartel est une entente entre des entreprises qui se fait au détriment de leurs clients et/ou de leurs fournisseurs. La collusion peut prendre, par exemple, la forme d'un accord sur des prix élevés ou sur un partage des marchés.

Intérêt et limites de la collusion

L'intérêt d'une collusion ou d'un cartel est de relâcher la concurrence et d'obtenir des profits plus élevés que ceux obtenus dans le cadre d'une concurrence normale, même imparfaite. Les exemples ne manquent pas dans le secteur du BTP (répartition des appels d'offres de marchés publics sur la construction du pont de Normandie ou de la ligne du TGV Nord à la fin des années 1980), des banques (entente

sur les crédits immobiliers au début des années 1990), de la téléphonie mobile (entente sur un gel des parts de marché entre 2000 et 2002)… Mais, heureusement pour les consommateurs, plusieurs obstacles rendent difficile la généralisation des ententes à l'ensemble des marchés.

Tout d'abord, l'instabilité peut venir de l'intérieur du cartel. Les entreprises sont, en effet, dans un dilemme de la concurrence, appelé aussi dilemme du prisonnier : collectivement, elles ont intérêt à coopérer pour relâcher la concurrence, mais individuellement, chacune est tentée d'adopter un comportement opportuniste (voir encadré).

Dilemme du prisonnier et dilemme de la concurrence

Le dilemme du prisonnier décrit une situation stratégique dans laquelle des joueurs ou des agents économiques sont amenés, chacun de leur côté, à prendre des décisions qui sont collectivement sous-optimales. L'histoire initiale est celle de deux voleurs qui ont commis un hold-up et qui se font arrêter par la police. Les policiers ne disposent toutefois pas de preuves suffisantes et décident de les interroger séparément pour obtenir des aveux. Chaque suspect se voit offrir deux alternatives : se taire ou dénoncer son complice. Si les deux complices gardent le silence, ils ne peuvent être condamnés que pour un délit mineur comme la détention illégale d'une arme et ne resteront pas longtemps en prison (un an). Si l'un dénonce et l'autre se tait, alors le premier est relâché et le second écope d'une longue peine de prison (huit ans). En revanche, si les deux se dénoncent, ils sont envoyés en prison, mais la sentence est moins sévère que si un seul est dénoncé (six ans). Si les utilités des prisonniers sont une fonction décroissante du nombre d'années de prison, alors on peut représenter les gains des prisonniers sous la forme de la matrice suivante. En ligne, nous avons les décisions possibles du prisonnier 1, et en colonne celles du prisonnier 2. Dans chacune des cases du tableau, le premier chiffre correspond à l'utilité du prisonnier 1 et le second chiffre à l'utilité du prisonnier 2.

		Prisonnier 2	
		dénoncer	se taire
Prisonnier 1	dénoncer	(– 6 ; – 6)	(0 ; – 8)
	se taire	(– 8 ; 0)	(– 1 ; – 1)

Si chaque prisonnier cherche à maximiser son utilité (ou minimiser le temps passé en prison), le choix « dénoncer » apparaît comme la meilleure stratégie. En effet, si le prisonnier 1 s'attend à ce que son complice se taise, il est relâché s'il le dénonce et il est condamné à un an s'il se tait. S'il s'attend à ce que son complice le dénonce, sa meilleure réponse est aussi de le dénoncer, car il est condamné à six ans contre huit ans s'il se tait. Quel que soit le choix de son complice, le suspect 1 obtient une plus grande utilité en choisissant de le dénoncer. Le raisonnement est similaire pour le suspect 2. La situation prévisible est donc que les deux se dénoncent, alors qu'il aurait été collectivement préférable que les deux se taisent.

Le dilemme du prisonnier renvoie à un dilemme entre l'intérêt collectif et les intérêts privés. On retrouve ce dilemme sur les marchés, où les entreprises auraient collectivement intérêt à coopérer (en augmentant leurs prix) ou en se différenciant, mais individuellement sont incitées à adopter un comportement agressif (en fixant des prix bas).

Une entreprise au sein d'un cartel peut être incitée à fixer son prix en dessous du prix collusif appliqué par les autres, de manière à récupérer l'essentiel des ventes. Or, comme les accords de collusion sont illégaux, il n'est pas possible pour les membres d'un cartel de signer un contrat juridiquement exécutoire qui les protégerait de tels comportements. Les accords de collusion doivent être auto-exécutoires, c'est-à-dire que les entreprises ne peuvent compter que sur elles-mêmes pour faire respecter la collusion. La solution passe alors par des menaces mutuelles de représailles futures sous forme de guerres de prix, de boycott ou de pressions physiques envers les entreprises qui ne respecteraient pas les termes de l'accord. Le cartel sera stable ou soutenable si pour chaque entreprise le coût à long terme des représailles est supé-

rieur au gain de court terme de dévier de l'accord collusif. Cette condition aura d'autant plus de chance d'être satisfaite que les entreprises ont une préférence élevée pour le futur (ce qui a pour effet d'accroître le coût des représailles en cas de déviation de l'accord collusif).

L'instabilité d'un cartel peut aussi venir de l'extérieur du cartel. L'entrée de nouvelles firmes, attirées par les profits collusifs, peut déstabiliser ce cartel, surtout si ces nouvelles firmes restent en dehors de l'entente. En pratiquant des prix plus bas, elles peuvent prendre des parts de marché aux membres du cartel. De ce point de vue, des barrières à l'entrée élevées constituent généralement un préalable à la mise en place d'un cartel.

Une dernière source d'instabilité est l'action des autorités de la concurrence qui peuvent, par des enquêtes, mettre à jour et démanteler les ententes. Une des tendances récentes a été le renforcement des moyens consacrés à la détection et la sanction des cartels en Europe. Les entreprises ayant eu des comportements collusifs peuvent se voir infliger une amende atteignant jusqu'à 10 % de leur chiffre d'affaires annuel. Par ailleurs, un nouveau dispositif, le programme de clémence, permet d'exempter de toute amende la première entreprise qui dénonce un cartel et apporte des preuves matérielles suffisantes. L'intérêt de ce programme, appliqué depuis quelques années en Europe, est de mettre les membres d'un cartel dans une situation de dilemme du prisonnier. Les entreprises savent, tout comme les deux suspects dans le dilemme du prisonnier, qu'elles ont intérêt à dénoncer avant d'être dénoncées. Ce programme vise donc à déstabiliser de l'intérieur des cartels qui sont de plus en plus difficiles à détecter, tant les entreprises prennent des précautions pour se concerter.

Les facteurs et pratiques favorables à la collusion

Certains marchés sont-ils plus propices aux cartels que d'autres ? Est-il possible de dresser un portrait-robot des marchés sujets à la collusion ? Les facteurs structurels favorables à la collusion sont ceux qui facilitent la

convergence de vue des entreprises sur un accord collusif et permettent de diminuer les incitations à dévier de cet accord (en facilitant la détection des déviations ou en augmentant le coût des représailles). Il ressort des travaux théoriques et empiriques sur les cartels, que les marchés concentrés, composés d'entreprises symétriques (c'est-à-dire proches en termes de coûts, de capacités de production, de localisation…) sont propices à la collusion. Il est, de fait, plus facile de s'entendre à trois entreprises qu'à dix, ou lorsque les entreprises ont des vues communes sur le marché. De plus, une offre concentrée permet de mieux observer les comportements des autres membres du cartel et de détecter plus rapidement les déviations. Enfin, les incitations à dévier sont réduites lorsque les entreprises sont peu nombreuses à se partager les profits collusifs.

Dans les facteurs favorisant la collusion, une demande dispersée (un grand nombre d'acheteurs) oblige les entreprises à rendre leurs prix publics ou transparents, ce qui facilite la détection des entreprises qui ne respectent pas l'accord collusif. À l'inverse, une demande très volatile ou irrégulière rend plus difficile la détection des déviations. Toujours dans les facteurs qui favorisent la stabilité des cartels, des excès de capacité donnent aux entreprises des moyens de représailles sévères et crédibles. De même, l'existence de contacts multimarchés permet de punir plus sévèrement une entreprise qui dévierait d'un cartel, en déclenchant des guerres de prix sur l'ensemble des marchés sur lesquels les membres du cartel sont en contact.

Les entreprises peuvent, par leurs pratiques, renforcer le caractère collusif du marché, en augmentant les barrières à l'entrée, la concentration de l'offre ou la transparence sur les prix. Par exemple, les membres du cartel peuvent menacer de racheter les entreprises qui ne veulent pas rentrer dans le cartel. Les entreprises peuvent s'échanger des informations sur les prix et les ventes récentes afin de s'assurer que personne n'a dévié de l'accord. Une autre pratique du même ordre consiste à mettre en place une clause du meilleur prix qui garantit aux clients qu'ils seront remboursés de la différence de prix si jamais ils trouvent moins cher ailleurs. Ces clauses qui, à première vue, semblent

bénéficier aux consommateurs peuvent se retourner contre eux, car elles permettent de surveiller à moindre coût les prix des concurrents et de rendre inopérante toute baisse de prix (puisque l'entreprise s'engage à s'aligner sur le prix le plus bas). Les autorités de la concurrence ont eu l'occasion, à plusieurs reprises, de sanctionner des clauses de meilleur prix dont l'objet était manifestement anticoncurrentiel.

INFORMATION IMPARFAITE

Sur les marchés, les consommateurs comme les producteurs ont rarement accès à l'ensemble de l'information pertinente. Ces asymétries d'information ont-elles un impact sur les comportements d'achat et les stratégies des entreprises ?

Des consommateurs imparfaitement informés

Les consommateurs sont rarement sur des marchés parfaitement transparents. En règle générale, ils ne connaissent pas la liste de tous les vendeurs présents sur un marché (proposant le produit ou le service recherché), ni la liste des prix pratiqués par ces vendeurs. De plus, ils peuvent aussi avoir des difficultés pour observer la qualité des produits ou des services proposés par ces vendeurs. On parle d'un bien de recherche lorsque le consommateur doit procéder à une recherche d'informations sur la liste des vendeurs et les prix qu'ils pratiquent, mais qu'il observe parfaitement la qualité du bien avant de l'avoir acheté. On parle, en revanche, d'un bien d'expérience lorsque la qualité du bien n'est connue du consommateur qu'une fois le produit acheté ou testé. Les biens culturels (livres, CD), les produits alimentaires ou cosmétiques relèvent des biens d'expérience.

Lorsque le consommateur est dans l'incertitude sur la qualité ou le contenu d'un produit, certaines entreprises peuvent abuser de cette asymétrie d'information en essayant de faire passer des produits de qualité médiocre pour des produits de bonne qualité. Ce type de com-

portement opportuniste est à l'origine du phénomène de sélection adverse qui est un dysfonctionnement du marché qui conduit à chasser les bons produits, voire à détruire le marché. En effet, si le consommateur est incapable de distinguer un bon d'un mauvais produit parce que les entreprises produisant de la mauvaise qualité font tout pour brouiller l'information, il n'acceptera jamais de payer un prix élevé. Les vendeurs proposant des produits de bonne qualité vont alors se retirer du marché, et le marché ne comportera plus que des produits de qualité médiocre.

Pour les entreprises proposant de la qualité haute, la solution consiste à adopter une stratégie de signalement. Pour que cette stratégie soit efficace ou crédible aux yeux des consommateurs, il faut qu'elle ne soit pas imitable par des firmes de basse qualité ; c'est-à-dire qu'elle soit beaucoup plus coûteuse à entreprendre pour une firme proposant de la mauvaise qualité que pour une firme proposant de la bonne qualité. Prenons le cas d'un bien d'expérience pouvant faire l'objet d'achats répétés. Si les consommateurs découvrent que le produit est de bonne qualité lors de leur premier achat, ils renouvelleront leur achat, alors que s'il s'agit d'un produit de mauvaise qualité, ils cesseront de l'acheter. Dans ce cas, une stratégie de signalement crédible peut consister à vendre à perte lors du lancement du produit ou à distribuer des échantillons gratuits, ou à faire une opération satisfait ou remboursé. Seule une entreprise proposant de la bonne qualité peut se permettre d'adopter une telle stratégie, car elle sait que les pertes qu'elle supporte lors du lancement du produit seront couvertes par les ventes dans les périodes suivantes. Cette stratégie ne peut pas être imitée par une entreprise proposant de la basse qualité, car cette dernière sait que sa présence sur le marché sera éphémère.

De même, la mise en place d'une campagne intensive de publicité ou l'offre d'une garantie sont des moyens de signaler la qualité du produit, car seule une entreprise certaine de la fiabilité de son produit a intérêt à le faire. La publicité n'est donc pas seulement un moyen de se différencier et de relâcher artificiellement la concurrence, mais c'est aussi un

moyen de réduire les asymétries d'information entre les consommateurs et les entreprises. Dans le premier cas, la publicité est persuasive (puisqu'elle cherche à modifier les préférences des consommateurs), alors que dans le second cas, elle est informative (Nelson, 1974).

Voir l'Annexe 1 (Dispersion des prix et information imparfaite) en ligne.

Des entreprises imparfaitement informées

Les entreprises peuvent également observer imparfaitement les caractéristiques et comportements des consommateurs. Ces asymétries d'information peuvent exposer les entreprises à un risque moral, c'est-à-dire à des comportements abusifs ou opportunistes de la part de leurs clients. Par exemple, un consommateur qui bénéficie d'une garantie sur un produit peut prendre moins soin du produit car il sait qu'en cas de dégradation ou de vol, il sera dédommagé. Les problèmes de risque moral se retrouvent particulièrement sur le marché de l'assurance, où les assurés peuvent être tentés de relâcher leur vigilance. Pour réduire le risque moral, les entreprises peuvent proposer à leurs clients des contrats incitatifs qui les amènent à rester vigilants ou à ne pas adopter des comportements abusifs. Par exemple, les compagnies d'assurances peuvent introduire des systèmes de franchise, de couverture partielle ou de bonus malus, qui incitent les assurés à ne pas provoquer d'accidents. En matière de garantie, une entreprise peut prévoir des clauses suspensives en cas d'usage non approprié du bien ou du service.

Les asymétries d'information peuvent aussi porter sur les comportements et caractéristiques des autres entreprises. Des problèmes de risque moral peuvent survenir dans le cadre d'accords de coopération interentreprises, que ces accords soient légaux (accord de R&D, alliances…) ou illégaux (cartel). Certains partenaires peuvent relâcher leur niveau d'effort ou de contribution par rapport à ce qui était prévu

dans l'accord, si leurs décisions sont imparfaitement observables par les autres partenaires. Par exemple, si les prix et les ventes des membres d'un cartel sont difficiles à surveiller, certains peuvent en profiter pour baisser secrètement leurs prix et accroître de cette façon leurs ventes. La généralisation de ces comportements opportunistes tend à fragiliser ou déstabiliser les accords de coopération.

L'absence de transparence peut de même contribuer à renforcer les barrières à l'entrée et le pouvoir de marché des firmes installées. En effet, des stratégies visant à dissuader l'entrée qui, en situation d'information parfaite, seraient jugées comme non crédibles ou inefficaces, peuvent le devenir lorsque les entrants observent imparfaitement les caractéristiques du marché ou des entreprises en place. Prenons l'exemple d'une stratégie de prix limite ou de prix prédateur qui consiste pour les entreprises en place à annoncer un prix suffisamment bas de manière à ce qu'aucun concurrent potentiel ne puisse espérer un profit positif en cas d'entrée sur le marché. Une telle stratégie peut-elle être jugée crédible par les entrants ? Non, si l'entrant a une information parfaite sur les entreprises en place, car un prix bas n'a pas une grande valeur d'engagement et a toutes les chances d'être révisé en cas d'entrée. Il est, de fait, coûteux pour les entreprises en place de maintenir des prix faibles une fois que l'entrée est effective. Ces dernières ont plutôt intérêt à s'entendre avec le nouvel entrant et à relever leurs prix. À l'inverse, en cas d'asymétrie d'informations sur les coûts de l'entreprise en place, une stratégie de prix limite peut devenir crédible et dissuader l'entrée sur un marché. C'est le cas lorsque l'entrant ne sait pas si l'entreprise en place est efficace (dans ce cas, l'entrée est non profitable) ou inefficace (dans ce cas, l'entrée est profitable). En pratiquant un prix suffisamment bas, l'entreprise en place peut parvenir à signaler qu'elle est efficace et dissuader l'entrée, même lorsque cette entrée est profitable. Une stratégie de prix limite s'inscrit plus largement dans les stratégies de prédation qui consistent pour une entreprise en place à renoncer à court terme à des profits afin d'éliminer les concurrents du marché ou de dissuader l'entrée, et ainsi d'obtenir des profits plus

élevés à long terme. Ce type de stratégies peut être considéré comme un abus de position dominante, tout comme la forclusion (contrôle des fournisseurs ou des débouchés à travers l'intégration verticale ou la signature de contrats exclusifs). Les autorités de la concurrence peuvent sanctionner de tels comportements par des amendes.

CONCLUSION

L'économie industrielle offre des outils et des modèles permettant de décrypter les stratégies concurrentielles des entreprises et de mieux comprendre comment les entreprises interagissent avec leurs concurrents, mais aussi avec leurs clients, leurs fournisseurs ou leurs autorités de régulation. Nous avons vu que l'objectif principal d'une entreprise est de préserver ou renforcer son pouvoir de marché et qu'elle dispose pour cela de différents moyens comme la différenciation de ses produits, la mise en place d'un cartel ou l'acquisition d'autres entreprises. Le marché français de la téléphonie mobile illustre bien les stratégies que l'on peut rencontrer sur les marchés de concurrence imparfaite. Tout d'abord, les opérateurs de téléphonie mobile sont en concurrence sur les prix de leurs forfaits, à travers les promotions ou les subventions d'accès qu'ils proposent aux nouveaux abonnés. Par ailleurs, chaque opérateur propose une gamme de forfaits qui sont verticalement différenciés en termes d'heures de communications et de qualité de services. Les opérateurs font aussi largement appel à la publicité et au marketing pour se différencier en termes d'images (différenciation artificielle). Enfin pour relâcher la concurrence, SFR, Orange et Bouygues Télécom se sont échangés des informations sensibles sur leurs ventes récentes et se sont entendus pour geler leurs parts de marché. Ces pratiques mises en œuvre entre 1997 et 2002 ont d'ailleurs fait l'objet d'une condamnation par les autorités de la concurrence en France.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 5

État, impôt et politique publique

Alain Trannoy

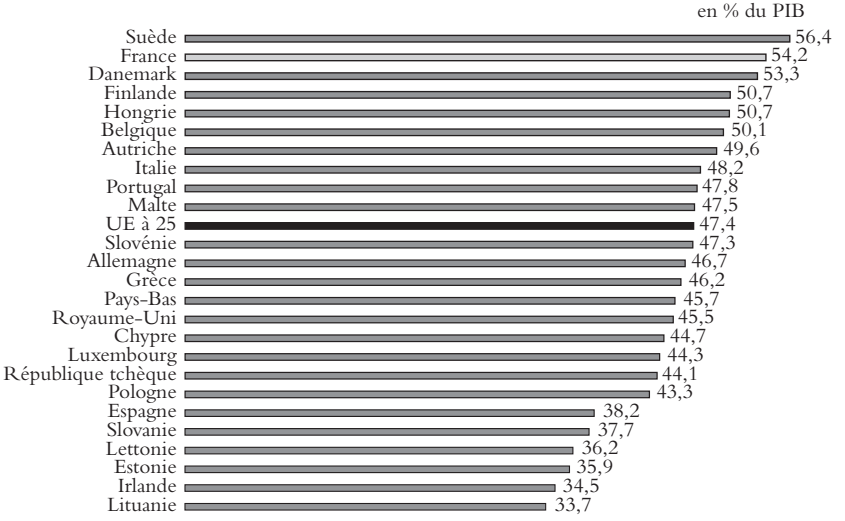
Directeur d'études à l'EHESS

Les fonctions régaliennes de l'État – justice, police, défense – sont un exemple de bien public pur, sujet au phénomène de passager clandestin, lequel permet d'expliquer l'émergence de l'État comme acteur économique complémentaire à l'initiative privée. L'action de l'État s'est par la suite étendue à la production de biens et services qui peuvent être qualifiés de biens publics impurs ou de biens privés soumis à des effets externes. La préoccupation distributive peut généralement être invoquée pour ces biens, tout comme elle est centrale pour le choix des impôts permettant à l'État de faire face à ses dépenses. Comme il existe peu d'impôts qui n'induisent pas une perte d'efficacité économique, le défi de la fiscalité consiste à retenir des outils qui permettent d'améliorer – dans le sens d'une plus grande justice sociale – la distribution du revenu qui résulte du fonctionnement des marchés, tout en minimisant les pertes d'efficacité économique.

L'État semble omniprésent dans la vie économique alors qu'il est absent de la théorie économique décrivant le fonctionnement des marchés (*cf.* le chapitre 4). Et pourtant, l'un des premiers économistes importants, David Ricardo, a tenu à faire figurer le thème de l'impôt dans le titre de son principal ouvrage. Avec l'apparition de la théorie moderne des prix due à Walras, Marshall et quelques autres, et systématisée par Arrow et Debreu,

s'est instillée peu à peu l'idée que l'État devrait s'occuper de ce dont le marché ne peut valablement s'occuper. Une sorte de principe de subsidiarité s'est au moins intellectuellement imposée au bénéfice du marché, où la place assignée à l'État serait finalement résiduelle.

La figure ci-dessous dément toutefois cette vision en quelque sorte subalterne de la place de la puissance publique, car même si des différences sensibles apparaissent dans la part des dépenses publiques rapportées au PIB, des pays réputés libéraux comme l'Irlande (les États-Unis sont dans le même cas de figure) consacrent quand même 35 % de leurs ressources à des dépenses publiques.



Source : Eurostat.

Figure 1 – Part de la dépense publique en % du PIB dans les pays de l'Union européenne en 2005

Ce pourcentage s'élève à 55 % dans les pays scandinaves comme la Suède, les pays de l'Europe de l'Ouest s'étageant entre 40 % et ce plafond. La France est le champion de la zone euro en 2005 avec 54 %.

Au total, il est difficile de ne pas conclure que les pays industrialisés connaissent une situation d'économie mixte où l'initiative privée et les marchés coexistent avec un rôle important dévolu à l'État.

Notre première tâche est donc de comprendre quels sont les rôles que celui-ci doit assurer sous peine de ne pas permettre aux marchés de fonctionner d'une manière adéquate. Ce raisonnement introduit le concept de bien public pur, dont les caractéristiques s'opposent trait pour trait à celui de bien privé. Ces mêmes caractéristiques empêchent le marché de produire d'une manière efficace ces biens spécifiques légitimant par là même le rôle de l'État. La présence de *passagers clandestins* atteste de l'impossibilité pour l'initiative privée de prendre le relais pour la fourniture de biens publics purs. On tient là un exemple de dilemme du prisonnier, démontrant que le défaut de coordination des agents peut être fatal à l'efficacité économique.

Par ailleurs, si l'unanimité prévalait quant à l'étendue souhaitable du domaine de l'action de l'État, on ne constaterait pas de telles différences entre des pays de développement comparables. Ces divergences portent sur la provision de biens et services publics qualifiés d'impurs qui sont en quelque sorte intermédiaires entre les biens publics purs et les biens privés. Ces biens peuvent en général être produits par le secteur privé, mais il n'est pas sûr qu'il les produise au niveau efficient, en raison de la non-prise en compte d'effets externes. Il s'ajoute également des préoccupations redistributives qui importent d'une manière considérable, le financement par l'impôt permettant de rendre l'accès au service public indépendant de la situation financière de l'usager.

Quel que soit l'arbitrage retenu concernant l'étendue du domaine public, le financement par l'impôt prévaut, accompagné du dilemme presque inévitable entre des instruments fiscaux favorisant une répartition des revenus plus équitable et ceux minimisant la perte d'efficacité économique induite par la taxation.

TROIS BIENS NÉCESSAIRES AUX MARCHÉS : JUSTICE, POLICE, DÉFENSE

Selon les historiens qui se sont penchés sur l'origine du marché, il semble que celui-ci soit le fruit d'une mutation de l'échange de dons entre voisins, amis ou parenté. Dès que l'échange cesse d'être bilatéral pour devenir multilatéral, il implique que les droits de propriétés soient bien définis. Le vendeur transmet la propriété du bien qu'il cède à l'acheteur. Si l'acheteur a un soupçon sur l'origine du bien dont est en possession le vendeur, par exemple vol ou rapine, il hésitera à l'acheter car il craindra la contestation ultérieure de son acquisition par autrui. Un doute sur les droits de propriété du vendeur mine donc les possibilités d'échange, comme en témoigne, à l'époque actuelle, la difficulté d'écouler des antiquités ou des objets d'art volés.

L'acheteur, comme le vendeur, a tout intérêt à ce que les droits de propriété sur les différents objets soient certifiés par une autorité externe. Le notaire en constitue un bon exemple pour les transactions dans le domaine foncier ou immobilier, pour lesquelles il représente le passage obligé pour vérifier les titres de propriété du vendeur et authentifier la transmission de droits de propriété. Toutefois, nommé par le garde des Sceaux, le notaire tire son pouvoir de l'État et il exerce, de ce fait, une mission de service public.

Ce n'est pas un hasard si l'instauration d'un système juridique protégeant les droits de propriété a constitué une priorité lors de l'ouverture à l'économie de marché des anciennes démocraties populaires situées à l'Est du rideau de fer. La définition de droits de propriété nécessite la création de tribunaux où les litiges concernant ces droits, mais aussi les contrats, peuvent être dénoués. L'existence de cours de justice placées sous la responsabilité d'une autorité souveraine est indissociable du pouvoir de police permettant l'exécution des décisions de justice.

Toute atteinte à la sécurité du transport des marchandises et des hommes qui se déplacent d'un marché à un autre se répercute négativement sur les possibilités de transaction, comme en témoigne la chute des Bourses de valeur après l'attentat du 11 septembre 2001. Le développement des échanges sur Internet n'a vraiment démarré que lorsque la sécurité des transactions financières a paru assurée. Bien sûr, la guerre constitue un exemple global d'insécurité particulièrement défavorable au développement des échanges. La défense d'un territoire contre des agressions extérieures participe aussi au climat de sécurité nécessaire à l'essor d'une économie de marché dans un espace donné.

Nous venons de rencontrer trois grandes fonctions (justice, police et défense), qualifiés de régaliennes, car assurées autrefois par un souverain dans un espace politique donné, et transférées par la suite aux États démocratiques qui leur ont succédé. Il faudrait ajouter à ces fonctions celle de battre monnaie, dont l'un des rôles est de servir d'intermédiaire des échanges, mais dont la spécificité impose une analyse détaillée (*cf.* chapitre 7).

LES BIENS PUBLICS

Il est utile de réfléchir aux caractéristiques communes que possèdent ces trois types de biens et qui les distinguent radicalement des biens dits « privés » comme les biens alimentaires. Ces derniers possèdent deux caractéristiques essentielles qu'il convient de rappeler.

Premièrement, les individus sont rivaux dans leur utilisation. Dans l'exemple des biens alimentaires, ce qui est consommé par l'un ne peut l'être par l'autre. Deuxièmement, le coût d'exclure un individu du bénéfice du bien est très faible, voire nul. L'instauration d'un système de prix constitue en lui-même une technologie d'exclusion particulièrement peu coûteuse. Une fois qu'un individu s'est porté acheteur d'un bien, aucun autre individu ne peut l'acquérir. Par là même, l'acte d'achat exclut tous les autres consommateurs possibles de l'utilisation du bien.

Non-rivalité dans l'utilisation des biens

Il est aisé d'observer, au regard des caractéristiques du bien privé, que les biens en question (justice, police et défense) se situent aux antipodes. D'abord, les individus ne sont pas rivaux dans l'utilisation du bien. Prenons l'exemple de la défense nationale. Le fait qu'un territoire soit protégé des agressions extérieures implique que tous ses habitants se trouvent protégés *ipso facto*. L'arrivée d'un habitant supplémentaire ne réduira en aucune façon la protection dont bénéficient les premiers arrivants.

Une autre façon de traduire cette disposition consiste à énoncer que le coût additionnel pour offrir la sécurité extérieure à un habitant supplémentaire est nul. En conséquence, il n'est pas *souhaitable* d'exclure un individu de la jouissance du bien en raison d'une application immédiate d'un raisonnement à la marge, à laquelle le microéconomiste a souvent recours. Afin de savoir si une action est désirable à la marge, c'est-à-dire une action dont les incidences sont minimes eu égard à la situation initiale, nous comparons le bénéfice social dit marginal au coût social marginal de cette action. Dans le cas d'espèce, le coût marginal est nul et le bénéfice marginal, positif, se résume à la satisfaction que ressent un habitant supplémentaire de vivre en paix dans un territoire donné.

C'est un exemple particulièrement évident d'application de l'analyse dite *coûts-bénéfices* qui se révèle si importante comme outil du calcul économique public. Cette analyse fait l'objet d'une utilisation poussée dans les décisions concernant la réalisation d'infrastructures publiques (routes, chemins de fer, aéroports, etc.), même si son domaine d'application est en principe beaucoup plus vaste.

Nous avons eu recours à l'exemple de la défense car il servait l'intérêt de la démonstration. Il vaut encore, à un degré moindre, pour les services de justice et de police. La différence provient de ce que le coût d'un usager supplémentaire, sans être prohibitif, n'est pas strictement nul. Des phénomènes de congestion peuvent se produire, tels que de possibles engorgements des commissariats de police après un

week-end marqué par de nombreuses atteintes à la sécurité des biens et des personnes, ou le délai particulièrement long entre l'instruction d'une affaire criminelle et le procès. L'utilisation des services de police ou de justice par un usager peut réduire la disponibilité de ceux-ci pour un autre individu. Néanmoins la différence avec un bien privé reste sensible. En général, si le service est dimensionné aux besoins de la population, ces phénomènes de congestion sont purement passagers, et le raisonnement selon lequel il n'est pas souhaitable d'exclure un individu de la consommation du bien reste valable.

Possibilité d'exclusion du bénéfice du bien

La seconde caractéristique d'un bien public est de nature différente, même si elle peut être aisément confondue avec la première. Indépendamment du fait que l'exclusion d'un individu de la consommation du bien puisse se révéler socialement souhaitable, elle porte sur la question de savoir si elle est simplement possible. Les bénéfices des services publics se caractérisent généralement par le fait qu'ils sont spatialisés. C'est l'évidence même dans le cas de la Défense nationale. Exclure un individu des bénéfices de ce service ne peut alors s'opérer que par son expulsion du territoire national, une solution qui, en raison de son caractère extrême, ne peut être mise en œuvre. Le coût social d'exclusion est alors très élevé, sans commune mesure avec le coût moyen par usager du bien en question. La Défense nationale constitue donc un exemple type d'un bien dont un usager ne peut être exclu.

En revanche, il est possible de laisser des individus sans protection à l'écart du système juridique d'un pays et, pour s'en tenir à la France, certaines populations en ont fait l'amère expérience dans les périodes les plus sombres de notre histoire.

Les biens publics purs

Lorsqu'un bien possède parfaitement les deux qualités précédentes – l'exclusion d'un consommateur supplémentaire *n'est ni possible ni souhaitable*

– le bien est qualifié de *bien public pur* selon la terminologie proposée par Paul Samuelson dans un article fondateur de l'économie publique.

Il existe une large gamme de biens qui sont intermédiaires entre le bien public pur et le bien privé, son antithèse. Ainsi, contrairement à la Défense nationale qui peut être qualifiée de service public pur, la police et la justice présentent des caractéristiques de biens publics impurs. D'une manière générale, ceux-ci sont tous caractérisés par un certain degré de non-rivalité et la présence d'*effets externes*, c'est-à-dire que l'utilisation du service par un individu peut influer sur le niveau de satisfaction d'autres individus. La vaccination d'un habitant supplémentaire contre une maladie contagieuse renforce la protection de toute la communauté contre cette maladie et constitue un exemple d'effet externe.

Bien que peu nombreux dans la réalité, les biens publics purs se révèlent fondamentaux pour comprendre la difficulté que pose au marché la fourniture des biens publics.

LE PHÉNOMÈNE DU PASSAGER CLANDESTIN

La question posée est de savoir si le marché peut produire un bien public pur de manière efficace. La propriété d'impossibilité d'exclusion engendre des problèmes formidables pour la solution de marché. Pourquoi un consommateur accepterait-il de payer un prix pour un bien qu'il peut consommer gratuitement ?

Le fait qu'un individu n'ait pas intérêt à acquitter un droit d'entrée pour un service dont il ne peut être exclu porte le nom de « phénomène de passager clandestin ». Une entreprise privée ne peut alors parvenir à couvrir les coûts de fabrication du bien public, et cela condamne à la faillite l'entreprise qui songerait à tenter l'aventure. Ce raisonnement particulièrement simple amène donc à la conclusion provisoire que le secteur privé n'est pas en mesure d'offrir ce type de bien qui offre un exemple type de défaillance du marché. Un autre acteur doit donc se substituer à l'initiative privée.

Ce raisonnement a été contré par l'observation empirique faite par Ronald Coase, selon laquelle trois quarts des phares anglais au début du 19e siècle résultaient d'une initiative privée. Le phare est certainement un exemple emblématique d'un bien public pur. Il est apparu comme tel à première vue à nombre d'économistes éminents pour lesquels il ne semblait ni souhaitable ni possible d'exclure un navire de l'usage d'un phare.

Or, à bien y regarder de plus près, le phare se révèle d'autant plus utile au navire que celui-ci se rapproche d'une côte, ce qui signifie en général qu'il vient de sortir d'un port, qu'il va s'y abriter ou décharger sa cargaison. La possibilité de prélever un droit d'utilisation du phare se révèle ainsi crédible pour tous les bateaux qui partent ou arrivent du port le plus proche du phare. C'est exactement cette possibilité qui a été utilisée par des entrepreneurs privés en Angleterre au 18e siècle. L'utilisation du phare, bien public pur, se révèle très complémentaire de celle d'un autre bien public soumis à congestion et dont il est facile d'exclure l'usage, le port le plus proche. La sanction, en cas de refus d'acquitter le droit d'usage du phare, est de ne pouvoir quitter le port.

Bien sûr, l'entrepreneur privé a besoin de l'accord de l'autorité qui contrôle le port pour appliquer la sanction, faute de quoi le phénomène du passager clandestin resurgit. C'est pourquoi les entrepreneurs privés, avant de construire un nouveau phare, faisaient avaliser par le roi d'Angleterre non seulement le fait d'exiger des navires fréquentant le port le plus proche ce droit supplémentaire, mais également le barème du droit en question en fonction du tonnage, afin que son paiement revête une obligation légale. L'exemple du phare permet ainsi de relativiser non le phénomène du passager clandestin, mais la conclusion qui en résulte quant à l'impossibilité pour le secteur privé de produire des biens publics purs.

LA SOUSCRIPTION AUX BIENS PUBLICS

Une autre particularité du mécanisme en question doit être ici soulignée et commentée. Lorsque l'entrepreneur présentait sa requête au roi, il

l'accompagnait d'une pétition de propriétaires de navires se déclarant prêts à payer le droit d'usage du phare. Bien que l'engagement fût formel à ce stade, ce procédé n'est pas sans rappeler le mécanisme dit de *souscription*, où des agents privés acceptent de payer pour la fourniture d'un bien public, de souscrire, sans être assurés qu'ils en soient les seuls bénéficiaires. Dans l'exemple du phare, des navires qui longeaient la côte en bénéficiaient sans contribuer à sa construction ou à son entretien.

La lutte contre le réchauffement climatique constitue également un bon exemple pour comprendre le mécanisme de souscription. Cette lutte représente sans aucun doute un bien public pur à l'échelle de la planète, alors que, face à ce problème, il n'existe clairement pas de gouvernement à l'échelle mondiale. Les efforts réalisés par les différents États résultent d'une action décentralisée qui fait l'objet en partie d'un effort de coordination, par exemple sous la forme du protocole de Kyoto. Certains pays adoptent un comportement de passager clandestin avec pour effet de les rendre bénéficiaires des efforts effectués par d'autres pour changer de mode énergétique sans changer le leur. Cette attitude égoïste n'empêche pas d'autres pays de vaillamment s'inscrire dans un programme ambitieux de réduction de leurs émissions de gaz à effet de serre.

De ces deux exemples que l'on pourrait multiplier, germe l'idée qu'il est faux de déduire de la présence de passagers clandestins l'absence de production de biens publics purs en l'absence d'État. La souscription constitue un mécanisme de fourniture décentralisée d'un bien public pur sans que la présence de l'État se révèle nécessaire.

La non-optimalité des comportements décentralisés

La vraie question est davantage de savoir si cette souscription possède les bonnes propriétés d'efficacité que la théorie économique prête au marché de concurrence pure et parfaite dans le cas des biens privés (*cf.* chapitre 4). Pour être concret, si l'on se réfère aux deux exemples précédents, la quantité de phares ou la réduction d'émissions de gaz à

effet de serre fournie par les mécanismes décrits plus haut sont-elles adéquates au regard du critère d'efficacité de Pareto ? En produit-on trop, trop peu ou juste ce qu'il faut ?

À cette question, la réponse obtenue par la théorie économique est sans appel si l'on admet que les agents économiques ne poursuivent que leur strict intérêt personnel et qu'ils ne sont mus par aucun sentiment altruiste ou la perception d'un intérêt général supérieur. La fourniture à l'équilibre de souscription – il s'agit bien d'un équilibre (au sens de Nash), chacun optimisant le montant de sa souscription en tenant compte de la souscription fournie par les autres – aboutit à une situation sous-efficace : une quantité insuffisante de bien public est produite à l'équilibre de souscription.

Une autre propriété intéressante a été obtenue, à savoir que plus le nombre de bénéficiaires est important et plus la sous-production est importante. En somme, lorsque le nombre de participants est faible, le mécanisme de souscription aboutit à une sous-production manifeste, mais celle-ci devient dramatique lorsque le nombre d'agents devient très élevé. Ce résultat avait été anticipé par l'économiste et sociologue Mancur Olson, qui a été le premier à insister sur l'importance du mécanisme de souscription. L'application de ce résultat au réchauffement climatique amène à constater qu'il est heureux que la négociation se déroule au niveau des 180 pays du globe plutôt qu'entre les 6 milliards d'humains.

Cette non-optimalité d'un comportement décentralisé ne doit pas complètement nous surprendre. En effet, il s'agit là d'un exemple supplémentaire du fameux dilemme du prisonnier dans les jeux non coopératifs, dans lesquels le défaut de coordination d'agents se comportant d'une manière stratégique se révèle coûteux sur le plan collectif.

Il faut toutefois souligner que les expériences en laboratoire effectuées par les économistes et les psychologues ne confirment pas entièrement cette sombre prédiction théorique lorsque des personnes, souvent des étudiants, sont placées dans une situation où ils doivent souscrire à un

micro-bien public à l'échelle de l'expérience de laboratoire. Les quantités souscrites dépassent assez souvent la quantité prédite par la théorie, suggérant que des phénomènes d'altruisme et de dépassement de soi sont à l'œuvre communément dans la fourniture de bien public.

Le phénomène d'éviction

Un remède à cette sous-production pourrait consister en ce que l'État complète à hauteur du déficit de la quantité privée par rapport à la quantité efficace. Cette solution n'est pourtant pas opérante en raison d'une propriété très particulière du mécanisme de souscription, connue sous le nom de phénomène d'*éviction* : une unité produite de plus par l'État incite les souscripteurs privés dans leur ensemble à souscrire pour une unité de moins, l'État étant perçu alors simplement comme un souscripteur supplémentaire. La fourniture publique entraîne l'éviction de la fourniture privée. Ce phénomène aurait été à l'œuvre en particulier dans la création artistique. Cet effet d'éviction contribue au constat d'une omniprésence de l'État dans ce domaine alors que cela n'a pas toujours été le cas par le passé.

LES SOLUTIONS ÉTATIQUES : UN PROBLÈME D'INFORMATION

Quoi qu'il en soit, même si l'on admet qu'il demeure une certaine incertitude sur l'importance de la sous-production à laquelle peut aboutir une solution décentralisée non coordonnée pour la fourniture de biens publics, il n'en reste pas moins que l'on est loin du résultat de la main invisible selon lequel des marchés de concurrence pure et parfaite aboutissent à une allocation efficace des ressources pour les biens privés. La conclusion qui semble s'imposer est que l'État doit prendre le relais pour la fourniture de biens publics purs.

En réalité, cette conclusion est quelque peu hâtive, dans la mesure où, si l'État peut sans aucun doute financer des biens publics grâce au prélèvement de l'impôt, il est plus ardu de prouver qu'il va réussir à

fournir la ou l'une des quantités efficaces du bien public. En effet, la solution décentralisée ne produisant pas d'une manière efficace le bien public, le problème posé est de savoir si l'État a les moyens de faire mieux. Par quel mécanisme est-il assuré de fournir la quantité efficace et non une quantité trop importante ou, à l'inverse, insuffisante ? Cette question et les réponses que l'on peut y apporter n'ont été comprises dans leur globalité qu'assez récemment, lorsqu'au tournant des années 1970 les économistes ont mieux intégré le rôle crucial que jouaient les problèmes d'information dans les questions d'allocation des ressources.

La règle de Samuelson

Pour le comprendre, il nous faut tout d'abord énoncer la règle optimale d'allocation des ressources dans le cas des biens publics purs obtenue par Samuelson. Il s'agit, là encore, d'une règle d'égalité d'un coût marginal et d'un bénéfice marginal.

On retrouve donc sur l'un des côtés de la balance le coût de production d'une unité supplémentaire de bien public en termes d'unité de bien privé. L'autre plateau de la balance est occupé par le bénéfice marginal social, c'est-à-dire la disposition de la société à payer pour une unité de bien public en plus. La société est composée par l'ensemble des consommateurs dans le cas d'espèce du bien public pur. En vertu de quoi, le bénéfice marginal social se calcule comme la somme des dispositions marginales de tous les membres de la société à payer pour le bien public.

La règle de Samuelson énonce donc que la quantité du bien public est efficace lorsque la somme des dispositions à payer pour une unité supplémentaire équilibre juste le coût supplémentaire de production de cette unité.

Pour mettre en œuvre cette règle, l'État a besoin de connaître les deux termes de cette égalité. Il ne doit pas éprouver trop de problèmes à obtenir de l'information sur le coût à la marge du bien public dans la

plupart des cas de figures, tout du moins s'il en assure lui-même la production. En revanche, il va éprouver davantage de difficultés à cerner le bénéfice marginal car seul chaque consommateur connaît sa propre disposition marginale à payer.

Prenons une fois encore l'exemple des phares. Bien sûr, il n'est pas trop difficile de déduire que l'utilité d'un phare supplémentaire est faible, voire nulle, pour un particulier qui ne pratique pas le nautisme. Pour les plaisanciers, les pêcheurs, les armateurs nationaux, on peut supposer en revanche que la disposition marginale à payer est positive. Mais elle peut différer d'un individu à un autre. Une certaine hétérogénéité des préférences ne peut être ignorée, qui rend très délicate une supputation du bénéfice marginal social. Si l'État commet une erreur sur cette dernière grandeur, il n'est pas en mesure, tout comme le marché, de calculer et par conséquent de fournir la quantité efficace de bien public.

La connaissance des préférences des agents

L'État doit donc chercher coûte que coûte à obtenir la révélation par les individus de leur disposition marginale à payer pour le bien public. Celle-ci a le statut d'une information privée que seuls les agents connaissent. L'exercice n'a rien d'évident, dans la mesure où l'agent peut craindre que la révélation d'une disposition marginale à payer relativement élevée influe positivement sur le montant de son impôt. La difficulté réside donc dans l'invention d'un mécanisme qui incite l'agent à révéler ses préférences.

Ce type de mécanisme existe-t-il ? Curieusement oui, ce sont les mécanismes dits de Clark-Groves, qui sont dérivés du mécanisme des enchères au second prix (l'enchérisseur ne paye que le second prix annoncé).

Ces mécanismes ont suscité un intérêt considérable dans les années 1970-1980, à telle enseigne que l'on s'est demandé si cette découverte n'était pas porteuse d'une véritable révolution concernant la fourniture

des biens publics. Il a fallu cependant déchanter car il a été montré que le coût du processus de révélation était important et qu'il s'en suivait que le budget du mécanisme n'était pas équilibré la plupart du temps. Au total, il n'est pratiquement pas possible d'obtenir un mécanisme de fourniture des biens publics qui soit simultanément révélateur, efficace et équilibré.

Il faut donc constater que l'État butte sur un problème de révélation des préférences et n'est pas assuré, à l'instar du marché, de produire la quantité efficace de bien public. La seule différence provient de la connaissance du sens du biais pour le marché, une sous-production presque à coup sûr, alors que l'on ne dispose pas de résultats généraux sur le sens du biais pour l'État.

En tout état de cause, la seule solution concrète au financement des biens publics purs, lorsque l'État en a la charge, semble reposer sur le prélèvement de l'impôt. Cette solution présente l'inconvénient de ne pas révéler d'information sur la disposition marginale à payer des individus, *a contrario* du financement au moyen d'un prix.

Toutefois, il faut garder à l'esprit que pour les biens publics qui présentent une possibilité d'exclusion, et ils sont nombreux, l'équivalent du prix existe sous la forme d'un droit d'usage dont il n'est pas difficile de trouver des illustrations concrètes, droit d'entrée dans un musée, droit d'inscription à une université, péage pour accéder à une route. Cependant, même pour ces types de biens, il est fréquent que le budget soit aussi couvert par le recours à l'impôt.

L'IMPÔT

Prélever l'impôt n'est pas une opération aisée car les gens préfèrent naturellement s'y soustraire. Même s'ils perçoivent qu'il est utile au financement des biens publics, ils préféreraient ne pas y être contraints ou le voir réglé par quelqu'un d'autre. Par conséquent, il ne faut pas s'attendre à ce que le produit de l'impôt entre dans les caisses de l'État

sans effort. Il lui faut, à tout le moins, la possibilité de recourir à la contrainte pour être en mesure de le recouvrir correctement. Dans l'État minimal, le monopole de la force légale lui est justement dévolu à travers les forces de police et l'armée.

Le monopole de la force légale

Il vaut la peine de s'arrêter sur cette situation qui semble parfois aller de soi mais qui, en raison du monopole en matière fiscale qu'elle implique, a suscité ou suscite encore des rivalités de la part d'organisations purement privées. Que l'on songe aux différentes mafias qui, en Italie du Sud, contestent ce monopole de l'État en prélevant sur les commerces ou services un droit d'installation, le *pizzo* ! À une échelle moindre, les bandes criminelles implantées dans certains quartiers des villes contestent également le monopole de la force légale de l'État et prélèvent sur certains commerces illicites des taxes qui échappent à l'État par définition. D'une manière générale, la plupart des guerres civiles peuvent être vues comme une compétition entre plusieurs groupes pour posséder le monopole de la force légale et ainsi s'approprier le pouvoir de taxer et les rentes afférentes.

Un des problèmes majeurs d'un certain nombre de pays en voie de développement, plus particulièrement en Afrique, résulte du fait que ces pays n'ont pas encore réussi à fonder un État qui dispose du monopole de la force légale. La présence dans le sous-sol de richesses minières constitue à cet égard un handicap, car l'expropriation des revenus du sol est plus facile que celle des produits du travail. Il semble avéré que l'avènement d'un État moderne sans compétiteur est plus difficile dans les pays riches en ressources naturelles.

Démocratie et tyrannie

Cette question est relativement indépendante de celle de l'apparition de la démocratie, au sens habituel du terme. L'Angleterre, l'Espagne et

la France, par exemple, ont réussi assez tôt, autour du 16e siècle, à faire cesser la contestation de la force légale de l'État sur leur sol, sans devenir pour autant des démocraties dès cette époque. On touche là l'un des paradoxes les plus profonds de l'économie publique, qui a été bien anticipée par des auteurs comme Thomas Hobbes, John Locke et Jean-Jacques Rousseau. La dévolution du monopole de la force à un agent, l'État, peut favoriser le progrès économique et humain à travers la fourniture de biens publics, si l'esprit dans lequel l'État émerge est celui du contrat social cher à ces auteurs.

Toutefois cette dévolution peut aussi engendrer une période de tension parfois extrême, de par l'importance des enjeux de pouvoir. Une fois le pouvoir de coercition créé et chaque homme dépossédé de la possibilité d'y résister, le retour en arrière est pratiquement impossible. Ce pouvoir peut tout aussi bien devenir un instrument de paix civile entre les hommes qu'un instrument de tyrannie terrifiante dont l'histoire ancienne et l'actualité offrent maints exemples.

Cette tentation tyrannique concerne au tout premier chef l'impôt qui peut devenir confiscatoire au détriment d'un groupe et au profit d'un autre. En effet, une particularité importante du paiement de l'impôt est d'être un jeu à somme nulle. En d'autres termes, ce qui n'est pas payé par l'un doit être payé par un autre. Cela constitue plutôt une singularité en économie car, dans l'échange, tous les participants gagnent puisqu'ils s'y engagent de façon volontaire.

La création de l'impôt et la coercition qu'il représente n'ont pas été acceptées sans garde-fous dans nombre de sociétés, bien avant qu'elles puissent être considérées comme démocratiques. Ainsi, sous l'Ancien Régime, les États généraux étaient convoqués pour consentir un nouveau prélèvement en hommes ou en argent et décider de sa répartition entre les différents ordres. Le fameux cri de ralliement des indépendantistes américains – « L'impôt sans représentation est une tyrannie » – prouve, s'il en était besoin, l'importance du détonateur fiscal dans l'avènement de la révolution américaine.

Le consentement à l'impôt

Assurer le consentement de l'impôt au travers d'une représentation démocratique représente une garantie pour éviter un trop grand arbitraire dans la répartition de la charge fiscale entre les différents groupes et individus constituant l'ensemble des contribuables. D'ailleurs, le suffrage censitaire, mis en place dans les premiers temps du balbutiement de la démocratie au 19e siècle, établissait un lien direct entre faculté contributive et appartenance au corps électoral.

À partir du moment où est posé le sain principe que l'impôt résulte d'un vote dans une assemblée représentative, se pose le problème dit du *choix collectif*, qui soulève des questions difficiles dont l'un des exemples est le théorème d'impossibilité d'Arrow. Les enjeux redistributifs sous-jacents au choix de l'impôt engendrent des différences d'intérêt et donc de point de vue qu'il nous faut maintenant introduire.

ENJEUX DISTRIBUTIFS

Les intérêts des contribuables sont antagonistes car l'impôt organise une redistribution du revenu résultant du fonctionnement des marchés. Cet antagonisme ne s'exprime pas directement mais se dissimule sous la forme d'un choix entre des types d'impôt qui frappent à des degrés divers les différentes catégories d'agent.

L'impôt foncier, dont on retrouve la trace dans les plus vieilles civilisations, et auquel les auteurs classiques comme Ricardo ont consacré d'importants développements, frappe les propriétaires dont une partie de la rente foncière va être ainsi accaparée par l'État. En revanche, les locataires en sortent plus ou moins indemnes.

L'impôt sur les transactions

L'impôt sur les transactions a généralement été mis en place un peu plus tard, transactions d'abord sur le marché des biens sous forme de droits de douane, droits d'accise ou, plus récemment, de TVA,

transactions sur le marché du travail sous forme de cotisations sociales et d'impôt sur le revenu du travail, enfin transactions sur les marchés financiers et du capital sous forme de taxe sur les intérêts, loyers et dividendes. Quelle que soit la personne qui acquitte une taxe sur les échanges, acheteur ou vendeur, les deux sont en général perdants ; le côté du marché le moins inélastique supportant le plus lourdement le fardeau de la taxe dans un régime de concurrence pure et parfaite.

Si les transactions revêtent un caractère anonyme, il est difficile de pratiquer un taux de taxe variant selon le montant des échanges en raison des multiples possibilités d'arbitrage introduites par cette différence de taux.

En revanche, si les transactions ne sont pas anonymes, cas par exemple des transactions en matière de logement ou de travail, il est possible de rendre variable le taux *marginal* de l'impôt.

La progressivité de l'impôt

L'impôt est dit « progressif » si le taux *moyen* de l'impôt progresse avec le montant imposable. Un impôt progressif sur le revenu fait porter une charge beaucoup plus importante sur les individus situés dans les déciles de revenu les plus élevés.

À titre d'exemple, il a été estimé que la redistribution verticale opérée en France par l'ensemble des impôts directs, complété de quelques prestations, revenait à transférer 5 % du revenu des 20 % les plus riches aux 20 % les plus pauvres, les déciles intermédiaires étant peu affectés (voir B. Salanié).

Au total, selon le type d'impôt, la redistribution fiscale avantage plus ou moins tel ou tel groupe d'individus. Cette redistribution peut être envisagée plus ou moins favorablement selon le jugement que l'on porte sur la distribution des revenus issue du fonctionnement des marchés.

Ici, rentrent en ligne de compte des jugements de valeur d'ordre normatif que l'économiste peut étudier sur le plan de leur stricte rationalité (voir M. Fleurbaey). Il s'agit de réfléchir aux critères de justice et d'équité à l'aune desquels on doit évaluer non seulement la redistribution de l'impôt, mais également l'ensemble des politiques publiques.

L'efficacité économique des différents impôts

Le choix de l'impôt offre donc un terrain où les intérêts des différents agents s'affrontent. Il ne faudrait pas en conclure que tous les impôts se valent sur le plan de l'efficacité économique. La plupart des impôts induisent une distorsion des décisions par rapport à celles qui seraient choisies dans une situation de *laisser-faire*.

Toutefois, certains impôts induisent une distorsion plus importante que d'autres. Par exemple, il n'est pas conseillé sur le plan de l'efficacité de taxer des biens dont la demande ou l'offre sont très élastiques au prix. À cet égard, l'impôt foncier frappant un facteur dont la quantité est fixe, la terre, induit des pertes d'efficacité moindres que d'autres impôts. Il a de chauds partisans comme Léon Walras ou Henry George, en particulier dans la perspective du financement des biens publics locaux offerts par des collectivités locales.

Le défi du choix collectif en matière fiscale consiste à retenir des outils qui permettent d'améliorer la distribution du revenu qui résulte du fonctionnement des marchés dans le sens d'une plus grande justice sociale, tout en minimisant les pertes d'efficacité économique.

Au total, la complexité des tâches dévolues à l'État nécessite une articulation complexe entre différents acteurs qui se prévalent d'une mission de service public. La déclinaison territoriale de la puissance publique en entités gigognes – Union européenne, État national, régions, départements, communes – qui toutes fournissent des biens publics et organisent une redistribution des ressources, en est une manifestation patente, tout comme l'existence de la Sécurité sociale et des différentes agences publiques, au premier rang desquelles il faut placer la Banque centrale européenne.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 6

Monnaie, banques et marchés financiers

MICHEL AGLIETTA

Professeur à l'université Paris X Nanterre

La monnaie est un système de règles. Elle définit la mesure de la valeur, sélectionne les moyens de paiements acceptables et organise le règlement des dettes. Dans une économie monétaire de marchés, les banques sont spéciales parce qu'elles créent la monnaie en accordant des crédits. Ce processus est inhérent au dynamisme économique parce qu'il permet à l'investissement de se libérer de la limite d'une épargne préalable. Il engendre l'accumulation du capital. Dans la finance globalisée, la puissance d'expansion du crédit est amplifiée par l'essor des marchés financiers. Mais cette finance qui fait des banques des auxiliaires des marchés est portée aux excès du crédit. Elle n'est pas autorégulatrice. Au contraire, elle est vulnérable aux crises. C'est pourquoi la finance requiert une régulation publique qui est celle de la banque centrale. Entité publique indépendante, la Banque centrale a la double mission de préserver la stabilité financière et la stabilité des prix dont dépend la confiance dans la monnaie. Pour affirmer son autorité, la Banque centrale peut être conduite à faire des arbitrages entre ses objectifs, lesquels sont devenus difficiles en raison de la globalisation financière qui rend les économies étroitement interdépendantes.

LA FINANCE EST L'INTELLIGENCE DE L'ÉCONOMIE

La finance définit l'horizon temporel des activités économiques, tant pour les individus que pour les entreprises. Car le taux d'intérêt est le pivot de l'arbitrage entre consommer aujourd'hui et consommer plus tard. Il donne aussi le seuil de rentabilité des projets des entreprises, tandis que la Bourse oriente l'allocation des ressources financières cherchant à s'investir. L'évaluation financière des valeurs et des risques repose sur un socle plus fondamental encore qui est la monnaie.

LA MONNAIE ET LES BANQUES

Nous vivons dans des économies de marchés. On peut se représenter l'économie comme un ensemble de biens qui doivent trouver preneurs grâce à des échanges. On ne peut pas échanger directement des biens entre une multitude d'individus qui ne se connaissent pas, qui ne savent rien de leurs désirs réciproques. Il faut un intermédiaire des échanges qui soit reconnu par tous, dont l'utilité pour chacun est d'être accepté par les autres. C'est la monnaie. L'échange d'un bien quelconque contre monnaie est un paiement.

Pourtant cette représentation est incomplète, donc inadéquate, car pour que les biens existent, il faut les produire. Il est préférable de se représenter l'économie comme un ensemble de producteurs séparés qui forment une division du travail. Les producteurs doivent d'abord acquérir le travail humain, les équipements et les inputs intermédiaires pour produire les biens qu'ils vont chercher à vendre. Ils doivent par conséquent acquérir leurs ressources en s'endettant avant de pouvoir disposer du montant des ventes. Une économie marchande est avant tout un réseau de liens de dettes. *C'est un commerce de promesses sur des biens futurs, avant d'être un échange de biens présents*. Le commerce de promesses est la raison d'être de la finance. Celle-ci n'est pas un supplément à l'économie de marchés, elle en est constitutive.

Mais si l'on ne suppose pas que les biens préexistent à leur production, on ne peut pas supposer non plus que la monnaie soit un objet qui préexiste aux échanges. Qu'est-ce que la monnaie ?

La monnaie est un système de règles

La monnaie est une abstraction humaine, comme l'écriture dont elle est contemporaine. Elle a évolué au cours des millénaires avec les activités humaines organisées en sociétés. Dans des économies formées de réseaux de dettes entre des agents privés, il faut pouvoir évaluer les dettes avant de conclure les contrats d'endettement. Il faut aussi pouvoir énoncer les prix d'offre auxquels les producteurs espèrent écouler leurs marchandises. Il faut en conséquence instituer une unité de mesure à laquelle tous se réfèrent. C'est *l'unité de compte commune* qui est unique dans un espace monétaire (l'euro, le dollar, la livre sterling, etc.). Les nombres exprimés dans cette unité sont des valeurs. La monnaie est donc la mesure de la valeur. C'est la première règle du jeu économique.

La deuxième règle est celle qui régit l'émission des dettes. Parce qu'elles représentent des promesses de valeur, toutes les dettes ne se valent pas du point de vue de leur capacité à payer, car ces dettes mettent à disposition des producteurs des moyens de paiements pour acquérir leurs ressources. Il importe de distinguer les dettes qui sont restreintes à la relation entre ceux qui les ont conclues et les dettes qui sont capables d'être acceptées en paiement par des tiers. La qualité d'être acceptée par les tiers est *la liquidité des dettes*. La liquidité est la règle qui fait de certaines dettes des moyens de paiements largement acceptés dans la société fondée sur une unité de compte. Les agents économiques qui émettent ce type de dettes en contrepartie des crédits qu'ils accordent à leurs emprunteurs sont les banques. Les dettes les plus liquides sont les dépôts des banques contre lesquels on peut tirer des chèques ou utiliser des cartes de débit.

Mais cela ne suffit pas à garantir l'usage de la monnaie comme un bien commun à tous, dont on puisse disposer en toute confiance. Car la liquidité des dettes émises par les banques est toute relative. Les banques

peuvent faire faillite si elles ont fait de mauvais crédits. Les dépôts émis en contrepartie des crédits n'ont alors plus de valeur. Il faut une troisième règle qui vérifie la liquidité des dettes proposées comme moyens de paiements. Cette règle est celle qui établit l'équivalence entre les moyens de paiements bancaires et la forme de liquidité ultime, parce qu'elle est unanimement et irrévocablement acceptée. C'est la monnaie émise par la Banque centrale. On appelle *règlement* le processus par lequel les moyens de paiements sont rendus équivalents, et constituent un flux homogène de monnaie qui irrigue l'économie comme le sang irrigue le corps humain.

Ainsi, lorsqu'on a des billets de la Banque centrale dans sa poche (ou des pièces de monnaie divisionnaire), le paiement que l'on fait avec ces billets se confond avec le règlement. Il est final. Si, au contraire, on paye par chèque, on transmet un ordre de paiement qui ne vaut pas règlement pour le bénéficiaire. Il induit une dette entre les banques des deux protagonistes. Le paiement devient final lorsque cette dette interbancaire est réglée, ce qui a lieu chaque jour par transfert de monnaie centrale entre les banques dans le système de compensation et règlement interbancaire tenu par la Banque centrale.

L'ensemble des trois règles constitue le système monétaire, qui est en quelque sorte la grammaire de la valeur. Il englobe la conservation de l'unité de compte et la régulation du système des paiements, lequel est hiérarchisé : les banques commerciales sont les banques de second rang, la Banque centrale est la banque de premier rang.

Pourquoi les banques sont-elles spéciales ?

La spécificité des banques se trouve dans la création monétaire : *les crédits font les dépôts*. Lorsqu'une banque accorde un crédit, elle ne prête pas un dépôt préexistant. Elle crée indissolublement un dépôt nouveau au bénéfice de l'emprunteur. La quantité de monnaie dans l'économie est donc augmentée (voir encadré). En sens inverse, lorsqu'un crédit est remboursé à une banque, une quantité équivalente de monnaie est détruite.

La création de monnaie bancaire

Création de monnaie par avance de découvert :

Agent emprunteur		Banque	
Compte courant bancaire : + 100	Emprunt : + 100	Prêt : + 100	Compte courant de l'argent : + 100

Création de monnaie par vente/achat d'un titre financier :

Agent vendeur de titres		Banque	
Titre financier : – 100 Compte courant bancaire : + 100		Portefeuille de titres : + 100	Compte courant de l'argent : + 100

Dans les économies monétaires, le pouvoir de commander la création de monnaie permet à la dépense d'être le moteur de l'économie. La dépense financée par injection nouvelle de monnaie transforme cette avance en revenu pour le vendeur de biens ou de travail (salaire). Les bénéficiaires de revenus suscitent de nouvelles dépenses, et ainsi de suite dans un processus multiplicateur du crédit initial. Ce processus n'explose pas mais converge vers un niveau agrégé de valeur produite dans une période de temps donné ; c'est le produit intérieur brut (PIB), parce que ce circuit a des fuites dans l'épargne, la partie du revenu qui n'est pas dépensée à chaque *round* de dépense. Au niveau d'équilibre du PIB, le montant agrégé de l'épargne est égal à celui de l'investissement qui a été financé par crédit. Mais la décision d'investir qui a déclenché la création monétaire n'est pas contrainte par les intentions d'épargne au même moment. Ce sont les revenus formés par

l'investissement qui déterminent le montant d'épargne qui lui est égal. On peut dire que *l'investissement fait l'épargne au niveau macroéconomique.*

On comprend bien l'importance cruciale de la création de monnaie nouvelle par les banques dans une économie capitaliste en croissance. C'est elle qui permet aux projets nouveaux d'investissement porteurs de progrès technique, qui sont des paris sur l'avenir dans un univers incertain, de se réaliser. C'est le pouvoir de commander la création monétaire qui assigne aux banques un rôle spécifique. Néanmoins, ce pouvoir n'est pas illimité puisque la troisième règle monétaire impose aux banques la contrainte de règlement de leurs dettes réciproques et la contrainte de conversion de leurs dépôts en cash. Cette contrainte se reflète dans leur bilan, dans la relation entre leur actif et leur passif. Cette relation unit deux fonctions hétérogènes : à l'actif, les banques font des crédits en tant que firmes privées cherchant à maximiser leurs profits ; au passif, chaque banque fait partie du système monétaire et doit par conséquent assurer l'équivalence de la monnaie qu'elle crée en monnaie centrale. Il s'ensuit que les banques supportent, au même titre qu'une autre firme, des risques qu'elles ne peuvent assumer individuellement, du moins dès qu'elles atteignent une taille importante. *Elles bénéficient d'une franchise et, en contrepartie, elles subissent une réglementation spécifique.*

Cette franchise bancaire consiste dans une sécurité des dépôts qui bénéficie à toute la société. C'est une garantie qui résulte de l'accès privilégié des banques commerciales aux fonds de la Banque centrale. En contrepartie de cette franchise, les banques doivent respecter différentes contraintes qui sont distinctes selon les systèmes monétaires : des réserves obligatoires en dépôts à la Banque centrale, des titres sûrs (ratio de liquidité) à leur actif, des fonds propres minimaux à leur passif (ratio prudentiel). Ces contraintes sont destinées à limiter le pouvoir de crédit des banques pour préserver leur stabilité financière. Mais les crises bancaires qui parsèment toute l'histoire du capitalisme jusqu'à nos jours montrent amplement que cette régulation prudentielle n'est pas parfaitement efficace lorsque l'appât du gain incite les banques à épouser les anticipations euphoriques qui s'emparent de la communauté financière tout entière.

En effet, les banques ont structurellement des bilans fragiles. Leur position, à la croisée de l'intermédiation financière et de la fourniture des services de la monnaie, les conduit à cumuler des risques que l'on ne retrouve pas rassemblés chez les autres agents économiques. Leur actif est fait en majeure partie de crédits illiquides, tandis que l'essentiel de leur passif doit être parfaitement liquide (les dépôts sont convertibles un pour un en monnaie centrale). Les banques sont beaucoup moins capitalisées que tout autre agent économique, c'est pourquoi il leur est imposé des ratios de capital minimaux. Lorsque les banques empruntent des ressources financières marginales en sus de leurs dépôts pour nourrir leurs crédits, elles payent des taux d'intérêt différents et souvent plus variables que ceux qu'elles reçoivent sur leurs crédits. Elles subissent en conséquence un risque de taux d'intérêt.

Mais ces précautions ne suffisent pas. Du fait de leur faible capitalisation, même en respectant les ratios minimaux, les banques peuvent subir des pertes très supérieures à leurs fonds propres. Ces pertes dépassent donc l'engagement de leurs actionnaires. Elles doivent être absorbées par la collectivité. Cela veut dire que le coût social d'une crise bancaire est très supérieur à son coût privé. Cela justifie non seulement la réglementation bancaire, mais aussi l'intervention publique en cas de crise. Les mécanismes de l'assurance privée sont incapables de maîtriser une crise. C'est le sens même du risque systémique. C'est pourquoi les banques commerciales dépendent étroitement de l'action de la Banque centrale. Elles fournissent le principal canal de transmission à l'économie de la politique de cette dernière. Il faut que le canal soit robuste pour que la politique monétaire soit efficace.

La Banque centrale : sauvegarde du système bancaire et contrôle de la monnaie

La Banque centrale émet la liquidité ultime. On pourrait croire qu'elle est acceptée en tout temps au sein de l'espace monétaire qu'elle régit. Pourtant, les expériences traumatiques de l'hyperinflation, où le pouvoir

d'achat de la monnaie est anéanti, révèlent qu'il n'en est rien. Le respect de la première règle monétaire, c'est-à-dire la conservation de l'unité de compte dans le temps, appelée *ancrage nominal*, est un souci permanent de la Banque centrale pour préserver la confiance.

Mais la deuxième règle monétaire est un souci tout aussi crucial. L'économie ne peut fonctionner correctement que si elle est irriguée par la liquidité sans à-coups. Une trop grande abondance de liquidités peut faciliter le renouvellement de dettes douteuses dont la solvabilité est suspecte. Une trop grande rareté de monnaie centrale rend les règlements bancaires, conformément à la troisième règle monétaire, difficiles ou excessivement coûteux – ce qui perturbe le crédit. L'activité économique en est déprimée. La liquidité devient un enjeu social. Les débiteurs sont plutôt en faveur d'un régime laxiste qui fournit une liquidité abondante et à bas coûts (faibles taux d'intérêt). Les créanciers, au contraire, sont concernés par la rémunération de leur épargne. Ils ne veulent pas être payés « en monnaie de singe ». *Lorsque ces conflits s'exacerbent, c'est la stabilité financière qui est en question.*

La Banque centrale a une double responsabilité : éviter que l'inflation ne corrompe la mesure des valeurs et prévenir l'instabilité financière qui déclenche des crises très coûteuses pour l'économie. La conduite de la Banque centrale ne peut être improvisée, bien qu'elle soit largement discrétionnaire. Elle poursuit une politique qui s'inspire d'une doctrine monétaire.

Stabilité financière

L'action de la Banque centrale dans les temps troublés de l'instabilité financière s'inspire de la doctrine *du prêteur en dernier ressort*. Cette doctrine s'est constituée au 19e siècle, en même temps que la Banque centrale s'affirmait comme la banque des banques pour conjurer les paniques bancaires. Ces paniques résultent de ruées sur les dépôts par crainte que les banques manquent de liquidités en monnaie centrale. Elles proviennent aussi de la paralysie du marché interbancaire par

réticence des banques à faire circuler les liquidités. *La Banque centrale prête en dernier ressort lorsqu'elle fournit aux banques dans le marché monétaire des montants de sa propre monnaie potentiellement illimités à des conditions qu'elle décide de manière souveraine*. Cette action est exorbitante de la logique marchande de l'échange des créances. Elle est poursuivie dans le but d'endiguer le risque systémique. Elle est ambiguë parce qu'elle peut alimenter l'imprudence des comportements bancaires si elle fournit une assurance collective aux banques que celles-ci peuvent anticiper. C'est pourquoi la Banque centrale doit laisser ses intentions imprévisibles et ne se préoccuper que de la sauvegarde de la stabilité financière globale, et non du destin de telle ou telle institution financière particulière.

Stabilité des prix

Que signifie l'ancrage nominal, ou encore la stabilité des prix ? Il y a des millions de prix individuels dans une économie développée, qui sont exprimés en unités monétaires. Certains dirigent des transactions immédiates, d'autres des transactions dans le temps. Lorsque les prix bougent dans le temps, les agents économiques ne peuvent prendre les bonnes décisions que s'ils parviennent à interpréter l'information dans la variation des prix. S'agit-il de l'effet d'une modification des conditions réelles de la production ou de la demande des biens individuels ? Ou bien s'agit-il de l'incidence d'indexations nominales des prix pour compenser la perception d'une diminution de la valeur de l'unité monétaire, c'est-à-dire de l'augmentation du niveau général des prix ? La première circonstance fait des prix des vecteurs d'efficacité économique. La seconde provoque le brouillage du calcul économique, et peut conduire à des surenchères entre les agents économiques cherchant à préserver leur part du revenu global par des indexations croisées qui engendrent une spirale inflationniste.

Mais comment définir le niveau général des prix ? C'est un indice, mais il y a un nombre indéterminé d'indices qui peuvent le représenter. *Le niveau général des prix est donc une convention statistique. Celle qui s'impose comme mesure de la variation du pouvoir d'achat de la monnaie est celle que la Banque centrale prend pour objectif de stabilité.*

Encore faut-il définir ce qu'est cet objectif. Son caractère conventionnel se reflète dans la multiplicité de ses formulations selon les Banques centrales et surtout dans la diversité de leurs modes d'action. On pourrait croire que la stabilité des prix veut dire que l'indice du niveau général des prix visé par la Banque centrale doit rester constant. Mais l'économie se meut dans l'incertitude. Il y a des chocs surprenants dont certains sont récessifs. Si l'inflation est trop faible ou nulle en moyenne, un choc récessif fait tomber dans la déflation. Or la baisse des prix est très dangereuse pour les agents endettés. Une inflation trop basse augmente donc le risque de l'instabilité financière. Par ailleurs, la Banque centrale ne doit pas réagir à n'importe quel symptôme de hausse des prix. Il existe des chocs inflationnistes, provenant par exemple de poussées des prix énergétiques ou alimentaires, qui sont réversibles. Si la Banque centrale ne prend pas en compte l'effet de retour des prix, elle risque de sur-réagir et d'alimenter l'instabilité des prix qu'elle prétend combattre.

En tirant les leçons de leurs erreurs, les grandes Banques centrales sont parvenues à se doter d'une doctrine pragmatique : *le ciblage flexible de l'inflation*. L'idée est de fournir aux agents économiques un point focal sur lequel ils puissent coordonner implicitement leurs propres anticipations. La politique monétaire dans la conjoncture se place sous la contrainte d'une règle d'action à moyen terme. Cette règle consiste à maintenir en moyenne, sur la durée d'un cycle des affaires, l'inflation dans une plage ni trop basse ni trop haute. Les écarts par rapport à la valeur centrale de cette plage se combinent à un indicateur du degré d'utilisation des capacités de production pour permettre à la Banque centrale de doser l'impulsion qu'elle veut donner au taux d'intérêt à court terme (voir encadré).

La politique monétaire comme arbitrage entre l'excès d'inflation et le sous-emploi

Soit p^* le taux d'inflation cible (le centre de la plage d'inflation tolérée à moyen terme par la Banque centrale) et y^* le niveau du PIB de plein-emploi (mesuré en log). La réalisation du couple (p^*, y^*) est

l'optimum pour la Banque centrale. Mais elle ne dispose que d'un seul instrument monétaire (son taux d'intérêt directeur). Soit i^* le niveau optimal du taux d'intérêt.

Pour atteindre sa cible, la Banque centrale est sensible aux écarts entre les grandeurs observées dans le dernier trimestre connu et les grandeurs optimales : $(y_t - y^*)$ et $(p_t - p^*)$.

Son arbitrage s'exprime par une fonction de réaction aux écarts, appelée *règle de Taylor*, qui permet de fixer le taux d'intérêt à court terme :

$$i_t = i^* + a(y_t - y^*) + b(p_t - p^*)$$

où a et b sont des coefficients qui expriment les préférences relatives de la Banque centrale dans l'intensité de ses réponses aux deux types de déséquilibres.

L'ESSOR DES MARCHÉS FINANCIERS

Depuis un quart de siècle, les marchés financiers ont pris une grande importance dans les économies capitalistes. Les marchés d'actions établissent des prix dont dépendent les rendements des capitaux investis dans les entreprises. Les marchés d'obligations permettent à celles-ci de trouver des financements alternatifs au crédit bancaire, réduisant ainsi le coût de l'endettement. Cependant, les marchés sont volatils et peuvent être emportés par des bulles spéculatives suivies de krachs. Les marchés de contrats dérivés (marchés à terme, d'options, de swaps) ont été créés aussi bien pour gérer les risques que pour spéculer sur eux. Contribuent-ils à la stabilité ou à l'instabilité dans l'économie ?

De nombreuses questions se posent sur les vertus et les vices des marchés financiers. Les optimistes mettent l'accent sur leur efficience, les pessimistes soulignent les comportements moutonniers qui sont responsables des excès générateurs de crises financières. C'est pourquoi, au-delà du fonctionnement des marchés, il faut s'intéresser aux acteurs qui ont profité de leur essor. Au cours des vingt dernières

années, marchés et acteurs prépondérants sont devenus globaux ; ce qui ouvre des perspectives mais pose des problèmes de régulation non résolus.

Efficience des marchés et vulnérabilité aux crises

Dire qu'un marché est efficient, c'est dire qu'il organise l'information de manière que toutes les données pertinentes pour anticiper les prix futurs sont rendues disponibles dans les mêmes conditions à tous les opérateurs. À tout moment, le marché établit un prix d'équilibre qui incorpore toutes les informations disponibles. Il est impossible à quiconque de « battre le marché ». Par conséquent, le mouvement du prix sur un marché efficient est une martingale. Le prix se modifie lorsqu'une nouvelle information imprévisible fait irruption, et qu'elle est unanimement et immédiatement exploitée par tous les intervenants. Les valeurs successives du prix sont indépendantes.

Remarquons que la définition de l'efficience ne dit rien sur le contenu de l'information dans le prix courant. Il n'y a aucun lien nécessaire entre l'efficience d'un marché et la valeur économique « fondamentale » de l'actif qui y est échangé. Il y a efficience dès qu'une croyance collective, appelée *convention de marché*, s'autoréalise. La raison se trouve dans les comportements des opérateurs. Un marché boursier, par exemple, n'est pas dominé par des investisseurs qui détiennent les titres sur de longues périodes pour toucher les dividendes. Le prix se fait par l'interaction de spéculateurs professionnels qui achètent les titres pour les revendre ou qui vendent les titres à découvert en espérant faire une plus-value, selon leurs anticipations de la valeur des actions dans le futur proche. Hormis l'arrivée d'informations nouvelles, qui est la source de la variation des prix, la force d'impulsion dépend de la psychologie collective de ceux qui interprètent l'information, c'est-à-dire de ce chacun pense que les autres pensent de l'effet des informations.

La convention de marché est formée par l'anticipation de chacun sur la croyance collective. C'est pourquoi elle est autoréférentielle. C'est le

produit de l'intersubjectivité des opérateurs. Dans certaines circonstances, cette croyance collective peut être très stable. Dans d'autres, elle est emportée par des vagues d'euphorie suivies de renversements violents d'opinion. Ce sont les bulles spéculatives et les krachs, tous deux également imprévisibles. Les krachs déclenchent des crises financières lorsque les achats de titres ont été financés par des dettes dont le remboursement dépendait de la poursuite de la hausse des cours.

Il en résulte que les comportements microéconomiques interdépendants sur les marchés du crédit et sur les marchés des actifs financiers entraînent des cycles financiers qui sont mus par le renforcement réciproque de l'expansion du crédit et des hausses spéculatives des prix sur les marchés des actifs – boursiers ou immobiliers le plus souvent. Les bulles sur ces marchés sont rationnelles parce que conformes à l'hypothèse d'efficience. L'instabilité est ainsi intrinsèque à la finance de marché.

La gestion d'actifs : marchés dérivés et investisseurs institutionnels

Parce que les marchés financiers sont instables, la gestion du risque est impérative. Toute l'innovation financière depuis plus de trente ans est tournée vers la gestion du risque financier. Supposons, par exemple, une banque française qui fait un prêt à cinq ans à taux fixes en euros à une entreprise industrielle. Parce qu'à la date du prêt, la politique monétaire américaine est plus dynamique que celle de la BCE, les taux américains sont plus avantageux que les taux européens. La banque choisit donc de financer le prêt par un emprunt à trois mois sur le marché de l'eurodollar. Ce faisant, la banque cumule plusieurs risques : un risque de taux d'intérêt – puisqu'elle emprunte à taux variable et prête à taux fixe –, un risque de change – puisqu'elle prête en euros et emprunte en dollars –, un risque d'échéance – puisqu'elle prête long et emprunte court et devra renouveler son emprunt à des conditions inconnues actuellement –, enfin un risque de crédit – si l'emprunteur fait défaut et ne peut donc honorer le service de sa dette.

Tous ces risques se cumulent dans le même prêt. Mais l'innovation financière permet de dissocier les facteurs élémentaires de risque énoncés ci-dessus et de les incorporer séparément dans des contrats dérivés. Ceux-ci peuvent être achetés par différents agents financiers qui se spécialisent sur la gestion de tel ou tel type de risque (voir encadré). *Les marchés des contrats dérivés permettent de transférer les risques selon le principe général de la division du travail.* Ainsi, notre banque française peut-elle se débarrasser du risque que le dollar s'apprécie au moment de rembourser son emprunt en achetant une option de change. Elle peut couvrir le risque de taux en concluant un swap taux fixe/taux variable avec une autre banque qui est spécialisée dans l'arbitrage de taux, et couvrir le risque d'échéance par un swap 3 mois/5 ans avec une compagnie d'assurances qui a un passif à long terme (contrats d'assurance-vie) et qui a donc besoin de prêter long. Enfin, la banque peut même vendre son risque de crédit en achetant une protection sous la forme d'un dérivé de crédit. Au bout de ces couvertures en cascades, la banque est finalement rémunérée d'une commission pour avoir initié le crédit industriel. Tous les risques ont été disséminés.

Les contrats dérivés

Contrat à terme : une partie accepte d'acheter quelque chose à une autre à une date future spécifiée et à un prix spécifié.

Contrat d'option : une partie accepte d'accorder à une autre le droit sans obligation de lui acheter ou de lui vendre quelque chose dans l'avenir à un prix d'exercice spécifié jusqu'à un horizon spécifié (options américaines) ou à une date spécifiée (options européennes). Les contrats d'achat sont appelés « *call* » et les contrats de vente « *put* ». Les produits optionnels complexes sont des combinaisons de *call* et de *put*.

Contrats de swaps : ce sont des enchaînements de contrats à terme par lesquels sont échangés des flux de paiements futurs sur des périodes qui peuvent aller jusqu'à dix ans ou plus. La valeur nette d'un swap est nulle à la date où il est conclu. Elle est égale aux paiements non compensés avec la variation des taux sur la branche variable au cours de la vie du swap. C'est donc un risque de contrepartie si le swap doit être renouvelé.

Swap de dérivé de crédit (« *credit default swap* ») : c'est l'achat d'une protection contre une prime payée au cours de la vie d'un prêt, donnant droit au paiement de la valeur faciale du prêt par le vendeur de protection à l'acheteur de protection lorsqu'un événement de crédit spécifié dans le contrat (faillite de l'emprunteur, restructuration ou fusion de l'entreprise emprunteuse, etc.) se produit.

Le transfert des risques par les banques permet d'améliorer la robustesse de leurs bilans. Il n'est possible que s'il existe des institutions financières qui ne subissent pas la contrainte de liquidité des banques, et qui sont capables et désireuses d'assumer ces risques dans une perspective de diversification de leurs portefeuilles d'actifs financiers. Ces acteurs munis de ressources d'épargne que leur confient les ménages sont les investisseurs institutionnels. À leur tour, ils confient le soin d'allouer cette épargne à des gestionnaires d'actifs. *On passe ainsi de systèmes financiers fondés sur l'intermédiation bancaire à des systèmes fondés sur l'intermédiation de marché dans lesquels les banques deviennent des courtiers du crédit.*

L'industrie de la gestion d'actifs a pris une importance énorme. Elle est alimentée par plusieurs sources d'épargne : les plans d'épargne collective pour la retraite, les produits d'assurance, les fonds communs de placements (FCP), les sociétés d'investissement de l'épargne privée, qui sont en partenariat (*hedge funds* et *private equity*) mais qui accueillent aussi les apports en capital des investisseurs institutionnels. L'épargne collective et l'assurance sont gérées de manière à garantir les revenus futurs des ayants droit. Les FCP et les sociétés d'investissement transmettent le risque aux épargnants finaux.

Le principe de la gestion d'actifs est la diversification des risques par la détention de portefeuilles d'actifs de marché dont les rendements aléatoires sont peu corrélés, voire non corrélés ou corrélés négativement, les uns avec les autres. Les gestionnaires d'actifs agissent comme des délégués des investisseurs institutionnels qui leur assignent des mandats de gestion, dans lesquels ils expriment des

objectifs, fonction de leurs engagements vis-à-vis des épargnants. Les horizons de placement et les proportions d'actifs risqués dans les portefeuilles dépendent des durations des engagements des investisseurs institutionnels (vingt ans et plus pour un fonds de pension, huit ans pour une compagnie d'assurance-vie).

Les sociétés d'investissement et les FCP constituent l'autre grande catégorie de la gestion d'actifs. Dans ces types de placements, ce sont les épargnants qui assument les risques *in fine*. C'est ce type d'acteurs qui apporte la liquidité aux marchés financiers en se portant acquéreurs des contrats dérivés. Mais ce sont eux aussi qui, en utilisant les mêmes outils de gestion et en prenant des positions similaires, sont largement responsables des comportements mimétiques conduisant aux bulles et aux krachs des marchés.

L'ère de la finance globale

La globalisation financière a commencé dans les années 1970 avec l'effondrement du système monétaire de Bretton Woods et les deux chocs pétroliers. Elle s'est imposée sous l'impulsion de l'essor des marchés financiers. Un système financier international, administré par les gouvernements et où les mouvements de capitaux étaient sous contrôle, a été remplacé par un système mû par les marchés et où les mouvements de capitaux ont été libérés.

La globalisation financière s'est propagée en trois étapes. Elle est étroitement liée aux réformes monétaires qui ont vaincu l'inflation.

La première étape va du début des années 1980 à l'effondrement du système soviétique au tournant des années 1990. Elle a entraîné le changement du pouvoir financier des débiteurs vers les créanciers, mais aussi des managers d'entreprise vers les actionnaires, avec le reflux de l'inflation et la montée des taux d'intérêt réels. Les innovations financières ont proliféré pour réduire le coût des crédits et pour gérer les risques dus à la volatilité des marchés.

La deuxième étape couvre la période qui va du début des années 1990 aux crises asiatique et russe (1997-1998). C'est l'époque dite du consensus de Washington, vécue comme la projection du capitalisme occidental sur le monde entier. Les pays appelés « émergents » étaient sommés d'adopter les institutions occidentales prétendument universelles et de s'ouvrir, sans restriction, à l'entrée des capitaux par la voie des marchés financiers internationaux. Cette époque d'arrogance a déchaîné des excès d'investissement et une spéculation immobilière en Asie, le vol systématique de la propriété publique par les mafias en Russie, la consommation parasitaire des classes aisées en Amérique latine. Elle s'est brisée sur les crises financières gigantesques de la fin du 20e siècle.

La troisième étape est celle de l'affirmation de la diversité des capitalismes au début du 21e siècle, mais aussi celle de dangereux déséquilibres financiers. Les pays émergents ont rétabli leur souveraineté sur leurs économies. Ils se sont débarrassés de la tutelle du Fonds monétaire international (FMI). De pays déficitaires, ils sont devenus excédentaires ; de débiteurs en dollars, ils sont devenus créanciers des États-Unis. La globalisation financière fonctionne à l'envers de la période précédente. Mais elle n'est pas devenue plus stable. Par l'augmentation démesurée de leurs déficits extérieurs, les États-Unis concentrent les déséquilibres de l'économie mondiale d'une manière qui leur a permis de consommer au-dessus de leurs moyens et de financer une gigantesque spéculation immobilière jusqu'à la crise qui a éclaté à l'été 2007.

On peut avoir un jugement mitigé sur la globalisation. D'un côté, certains s'extasient sur la flexibilité des marchés financiers qui absorbent les mouvements de balancier dans les positions créditrices et débitrices des pays. De l'autre, l'endettement augmente inexorablement plus vite que le PIB mondial pour soutenir une hausse intenable du prix des actifs financiers, sans que des ajustements correcteurs ne soient mis en œuvre.

FINANCE ET POLITIQUE

Au-delà des dispositifs de contrôle prudentiels à efficacité limitée, la finance doit être gouvernée par la monnaie, mais la régulation de la finance globale rencontre l'obstacle de la séparation des monnaies dans des espaces de souveraineté. Or, plus les marchés étendent leur emprise sur les économies, plus la stabilité financière devient un bien commun qui dépasse les frontières des nations. Et il n'y a aucune raison que les acteurs financiers puissants qui émanent de différents types de capitalismes aient des intérêts compatibles. Il va falloir mettre en cause le postulat libéral de la toute-puissance régulatrice des marchés. Le monde d'interdépendances multiformes et hétérogènes est celui de la confrontation des puissances. Le fonctionnement d'une finance globale dans ce monde implique une concertation interétatique selon des institutions de gouvernance internationale rénovées.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 7

Économie du travail

Bruno Van der Linden

Fonds de la recherche scientifique –
Université catholique de Louvain

L'économie du travail étudie l'ensemble des comportements et des transactions associés à une activité professionnelle. Elle accorde une place particulière à l'échange de services du travail contre un salaire et donc à l'étude du salariat. Elle s'intéresse, en amont de cet échange, aux comportements d'*offre de travail* émanant de la population et à ceux de *demande de travail* émanant des entreprises. Parmi les comportements relatifs à l'offre de travail, l'économie du travail s'intéressera notamment aux choix d'éducation et de formation continue, et aux comportements migratoires.

Les principaux acteurs en économie du travail sont les individus et les ménages d'une part, et les entreprises d'autre part. D'autres acteurs jouent cependant un rôle essentiel : les partenaires sociaux et l'État en tant que régulateurs des échanges (par la législation du travail et les conventions paritaires) et responsables de l'enseignement obligatoire, de la formation continue, de systèmes d'assurance contre divers risques sur le marché du travail, etc.

Le marché du travail est complexe à analyser. La concurrence et l'information y sont souvent imparfaites. Les risques encourus par les travailleurs (risques d'accident du travail, de dévalorisation inopinée de leur savoir-faire) sont très difficiles à diversifier. Il y a donc, de leur

part, une demande d'assurance contre les risques présents sur le marché du travail. En outre, le travail est en lui-même un « bien » fort particulier. Il occupe, en effet, une place cruciale dans nos sociétés.

Au fil du temps, l'économie du travail s'est éloignée d'hypothèses fort simples à propos des objectifs des acteurs économiques. Dans les limites de cette introduction, il n'est pas possible de faire écho de ces travaux récents. On s'en tiendra donc à des hypothèses standards de la discipline.

> ***Voir le Document 1 (Quelques références…) de l'Annexe 1 en ligne.***

Ce chapitre commence par les déterminants de l'offre et de la demande de travail. Il aborde ensuite la question de l'équilibre sur ce marché. La négociation collective, phénomène majeur en Europe, fait l'objet de la section suivante. Ce chapitre s'achève par une analyse du chômage et de ses déterminants. L'analyse est limitée à une seule période de temps, elle ignore les aléas (l'incertitude) et n'aborde pas le rôle des prélèvements obligatoires.

L'OFFRE DE TRAVAIL

Un modèle de base

L'analyse théorique standard de l'offre de travail s'appuie sur l'hypothèse que les individus maximisent un objectif, appelé *fonction d'utilité* et noté *U*. Les déterminants de l'utilité peuvent être abordés de plusieurs manières. La plus simple consiste à supposer que deux grandeurs affectent positivement l'utilité : la consommation (sans distinction entre les divers biens et services) et le loisir. Notons-les respectivement *C* et *L*. À mesure qu'une quelconque de ces grandeurs augmente, l'utilité s'élève mais à un rythme qui faiblit. Consommation et loisir

sont en outre, dans une certaine mesure, substituables. L'individu a une dotation en temps L_0 (la longueur de la période diminuée d'une durée forfaitaire consacrée au sommeil et aux travaux ménagers). Aussi, le temps de travail, rémunéré à un salaire unitaire donné w, n'est autre que $L_0 - L$. Le loisir est donc ici perçu comme le temps passé à ne pas travailler. Le *coût d'opportunité* du loisir est dès lors le salaire auquel l'individu renonce. Il se peut que l'individu dispose d'un revenu I indépendant de son temps de travail (des revenus de la propriété, par exemple). Si le prix unitaire de la consommation est noté P, la *contrainte budgétaire* de l'individu impose que les dépenses soient au plus égales aux revenus :

$$P \times C \leq w \times (L_0 - L) + I$$

Par la suite, on simplifiera quelque peu l'écriture en normalisant P à l'unité : w devient alors le salaire (unitaire) réel.

Bien des contraintes pourraient limiter l'espace de choix de l'individu : par exemple, un seuil minimal de consommation nécessaire à la survie ou une durée maximale de travail. Dans un souci de simplicité, supposons que seule la contrainte budgétaire limite le choix de C et de L. En l'absence de phénomène de satiété, l'individu utilise tout son budget : la contrainte budgétaire devient alors une égalité.

L'individu choisit donc ses quantités de consommation et de loisir de manière à élever autant que possible son utilité $U(C, L)$ sous sa contrainte budgétaire. Pour l'étude de ce choix, il est commode de définir l'équivalent de courbes de niveau sur une carte géographique. Il s'agit de combinaison de C et L procurant le même niveau d'utilité : on parle de *courbes d'indifférence*. La figure 1 les représente dans un espace à deux dimensions, où le temps de loisir et le temps de travail sont en abscisse et la consommation en ordonnée. Compte tenu des hypothèses ci-dessus, ces courbes d'indifférence présentent plusieurs propriétés illustrées sur cette figure : elles sont en particulier décroissantes et convexes par rapport à l'origine des axes. Cela signifie que l'individu est disposé à réduire sa consommation en échange de plus de

temps de loisir mais que la réduction de consommation qui le laisse indifférent diminue à mesure que le loisir augmente. La (valeur absolue de la) pente de la courbe d'indifférence porte le nom de *taux marginal de substitution entre la consommation et le loisir*. Le niveau d'utilité s'élève à mesure que la courbe d'indifférence occupe une position plus élevée à la figure 1. Le choix optimal, le point E de coordonnées (L^*, C^*), est donc caractérisé par la tangence entre la courbe d'indifférence et la contrainte budgétaire (la droite AB de pente w si on regarde l'axe horizontal de droite à gauche). Autrement dit, à l'optimum, le taux marginal de substitution égale le *salaire réel w*.

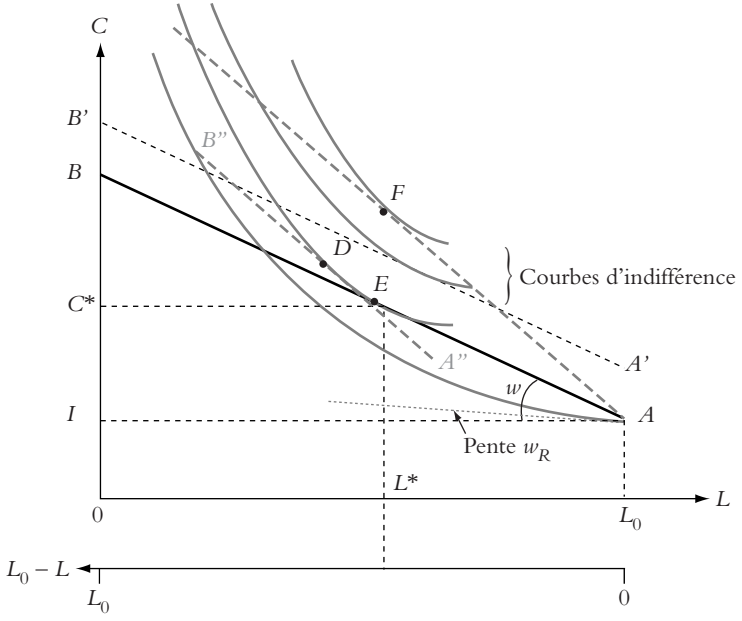


Figure 1– Décision de participation et arbitrage entre consommation et loisir

Intéressons-nous à présent à l'effet de modifications du revenu non salarial I et du salaire w. Si le revenu I augmente, l'ensemble de la contrainte budgétaire glisse vers le haut de la figure 1 (la droite AB devient A'B').

À salaire, prix et fonction d'utilité inchangée, le niveau de loisir augmente. Cet effet, appelé *effet revenu*, s'observe car le loisir est un bien qualifié de *normal* (autrement dit, quand l'individu devient plus riche, il désire plus de temps de loisir, propriété que confirment les analyses menées sur la base de comportements observés, encore appelées *études empiriques*).

Considérons à présent une hausse du salaire réel de w. L'effet net sur le choix de l'individu se décompose en deux étapes. D'abord, à utilité inchangée (en restant donc sur la courbe d'indifférence passant par E), le coût d'opportunité du loisir augmente. Dès lors, l'individu souhaite augmenter son temps de travail. Cela se vérifie sur la figure 1. Partant de l'optimum (L^*, C^*), confronté à un salaire plus élevé (la droite de budget en traits interrompus A''B''), l'individu réduit son temps de loisir (comparez les *optima* E et D). On parle d'*effet de substitution*. Ensuite, la hausse de salaire engendre un surcroît de richesse : la droite de budget du consommateur n'est en réalité pas A''B'' mais AF, qui lui est parallèle. L'effet revenu conduit à passer de D à F. Comme l'effet revenu entraîne une hausse du temps de loisir désiré, l'effet net d'une hausse de salaire sur l'offre de travail est en général de signe indéterminé. Dans l'exemple de la figure 1, les effets de substitution et de revenu se compensent.

Jusqu'ici, nous avons supposé que l'individu désire participer au marché du travail. Or, ce n'est pas vrai pour n'importe quel taux de salaire réel w. Disposant du revenu I, il peut choisir d'allouer tout son temps disponible au loisir (auquel cas, $L = L_0$ et son choix se porte sur le point A de la figure 1). La courbe d'indifférence qui passe par A a une pente. Cette pente fournit le salaire minimal, appelé *salaire de réserve* et noté w_R, sous lequel l'individu préfère rester en dehors du marché du travail (ou encore *inactif*). Dans l'exemple de la figure 1, le salaire de réserve est très petit car l'individu représenté a un taux marginal de substitution très faible au voisinage de $L = L_0$.

En résumé, ce cadre théorique simple prédit l'allure suivante du nombre d'heures de travail (offertes) en fonction du salaire horaire réel. Sous le salaire de réserve, l'offre est nulle. Ensuite, comme le

confirment les études empiriques, l'effet de substitution l'emporte d'abord sur l'effet revenu : le temps de travail augmente donc avec le salaire. Au-delà d'un seuil salarial, l'inverse se produit. On aboutit ainsi à une courbe en forme de cloche. L'*élasticité de l'offre* mesure le pourcentage de variation des heures de travail offertes lorsque le salaire réel varie d'un pour cent. Depuis la révolution industrielle, on a assisté à une division par deux de la durée annuelle du travail et à une hausse vertigineuse des salaires. Ce cadre d'analyse suggère qu'à cette échelle historique, l'effet revenu l'a largement emporté.

La prise en compte de l'information imparfaite

Les modèles d'offre de travail abordés jusqu'ici supposent que toute offre de travail trouve un employeur sans délai. On est ainsi soit inactif, soit employé, jamais chômeur. Or la prospection d'emploi est une activité consommatrice de temps.

Pour rendre compte de ce phénomène, introduisons l'information imparfaite. Admettons, pour la simplicité, que le salaire soit la seule caractéristique pertinente d'un emploi. Une personne sans emploi ne connaît pas le salaire associé à chaque offre d'emploi correspondant à ses compétences. Admettons qu'elle connaisse uniquement la distribution du niveau de salaire. La prospection d'emploi consiste alors à rechercher des offres d'emploi et, face à chaque proposition d'embauche, à décider s'il y a lieu de l'accepter ou d'y renoncer pour poursuivre la prospection. La règle de décision est en soi simple : si le gain espéré d'une poursuite de la recherche dépasse son coût, la proposition d'embauche est rejetée et la prospection se poursuit. Sinon, l'offre d'embauche est acceptée.

Ayant reçu une offre de salaire w, le coût de prolonger la prospection est le coût direct de la recherche d'offres (envoi de lettres de candidature, déplacements…) augmenté du *coût d'opportunité* de l'offre (à savoir le salaire auquel on renonce) et diminué des gains en cours de recherche (par exemple, une allocation de chômage). Le coût de la

recherche *croît* donc avec le niveau salarial w de l'offre. Si la quête de propositions d'embauche se poursuit, le gain espéré augmente avec la probabilité de découvrir une nouvelle offre associée à un salaire supérieur à w. Ce gain croît aussi avec l'écart attendu entre ces offres salariales supérieures et le niveau w de l'actuelle proposition d'embauche. Intuitivement, plus le salaire de la proposition w est élevé, plus le gain d'une prolongation de la prospection est *bas*.

La confrontation d'une relation croissante en w mesurant le coût de la recherche et d'une relation décroissante en w mesurant son gain définit le *salaire de réservation*. Cette notion reçoit l'interprétation suivante : toute offre supérieure (respectivement, inférieure) à ce salaire de réservation est acceptée (respectivement, rejetée).

Ce cadre théorique permet l'étude des déterminants microéconomiques de la sortie du chômage. La probabilité d'embauche est le produit de deux probabilités : celle de recevoir une offre (fonction de la vigueur de la demande de travail) et celle que cette offre d'embauche soit plus lucrative que le salaire de réservation. Il y a un vif débat sur le rôle des incitations monétaires à la sortie du chômage. Face à la complexité de la question, la littérature empirique, bien qu'assez vaste, ne converge pas sur des conclusions unanimes. Le niveau des allocations de chômage paraît réduire, souvent modérément, la sortie du chômage vers l'emploi. Au voisinage de la fin de droit, on identifie une accélération souvent substantielle de la sortie, mais le contexte (allocations d'assistance, économie informelle…) a toute son importance. En général, les auteurs concluent que la probabilité d'acceptation d'une offre est très élevée.

Voir le Document 2 (À propos de l'offre de travail…) de l'Annexe 1 en ligne.

LA DEMANDE DE TRAVAIL

Un modèle de base

Considérons une entreprise qui maximise son profit. Pour la simplicité, elle produit un bien, en quantité Υ, à l'aide d'un unique *facteur de production* : les services du travail. L'entreprise peut demander (et utiliser sans délai) n'importe quelle quantité de travail à un prix unitaire qui s'impose à elle : le salaire w. Sa technologie est caractérisée par un *ensemble de production*. Pour chaque niveau Υ, cet ensemble caractérise les quantités de travail permettant la production de cette quantité. Pour être efficace et minimiser ses coûts, l'entreprise doit utiliser la quantité minimale de main-d'œuvre. La *fonction de production* est la relation entre cette quantité minimale et la quantité produite correspondante. On fait ici l'hypothèse qu'une unité de travail supplémentaire permet de produire davantage (on parle de *productivité marginale du travail* positive) mais que cette augmentation faiblit à mesure que l'entreprise grandit. La productivité marginale est donc décroissante (à tout le moins au-delà d'un seuil). Si l'entreprise se comporte également de manière concurrentielle sur le marché du bien (le prix de vente s'impose à elle), elle a intérêt à occuper du personnel jusqu'à ce que sa productivité marginale égale son salaire réel (le salaire w divisé par le prix de vente). Cela définit implicitement une relation décroissante entre la quantité de main-d'œuvre demandée et le salaire réel, que l'on appelle la *demande de travail*. Cette analyse s'étend aisément au cas d'une firme qui dispose d'un pouvoir sur le marché des biens (voir Cahuc et Zylberberg, 2001, p. 93).

Étendons ce qui précède à deux facteurs de production. Il s'agira des flux de service du travail de deux types de main-d'œuvre ou encore des flux de service du travail et du capital. Pour fixer les idées, retenons l'exemple de deux types de main-d'œuvre utilisés en quantité E_i et payés au salaire (réel) w_i (i valant 1 ou 2). Admettons aussi que les deux facteurs de production soient des substituts. Étant donné un niveau de production Υ, il existe alors un ensemble de combinaisons des quantités E_1 et E_2 permettant de produire efficacement Υ. Tenant compte de

ces possibilités technologiques, l'entreprise minimise ses coûts. Elle exprime ainsi une demande de travail *conditionnelle* (à Υ) qui dépend des salaires w_i et bien entendu du niveau Υ. Compte tenu d'hypothèses standards à propos de la technologie de production, la demande de travail de type i baisse avec son prix w_i. L'hypothèse de substituabilité entraîne que la demande E_i croît en fonction du prix de l'autre facteur. L'*élasticité de substitution* mesure la facilité avec laquelle un facteur peut être substitué à l'autre quand le prix relatif des facteurs change. La demande E_i est d'autant plus sensible (*élastique*) aux salaires que l'élasticité de substitution est élevée.

La quantité produite Υ ne s'impose généralement pas à une entreprise. Selon le contexte concurrentiel de l'entreprise, le canal par lequel l'entreprise détermine Υ diffère (choix de la quantité offerte ou choix du prix de vente). Quoi qu'il en soit, les prix des facteurs influencent la quantité Υ qui maximise le profit et celle-ci, on l'a vu, a un impact sur la demande de facteurs. On parle d'*effet volume*. Quand celui-ci est pris en compte, la demande de facteurs devient *inconditionnelle*. L'élasticité inconditionnelle par rapport au prix du facteur correspondant est en valeur absolue plus élevée que l'élasticité conditionnelle. Autrement dit, l'effet volume vient renforcer l'effet de substitution.

Ces résultats doivent se comprendre à prix des facteurs inchangés dans les autres entreprises. Si le salaire d'un type de main-d'œuvre augmente dans toute l'économie, les effets sont plus complexes. Si les entreprises produisent des biens substituts, la hausse du salaire chez les concurrents va influencer leurs comportements d'offre ou de fixation de prix. Cela aura en retour une influence sur le volume des ventes de l'entreprise considérée. En outre, le salaire et l'emploi déterminent le revenu des ménages et donc leur demande de consommation. Les effets induits d'une telle hausse de salaire vont en sens divers. La mesure de l'effet macroéconomique net d'une hausse de salaire sur l'emploi est donc une tâche de nature empirique. Pour de petites économies ouvertes au commerce international, le signe estimé de cet effet est généralement négatif.

La prise en compte de l'information imparfaite

Le modèle de base ci-dessus suppose que l'entreprise peut occuper sans délai autant de travailleurs qu'elle le souhaite au salaire du marché. En réalité, l'entreprise crée un poste de travail, déclare une vacance d'emploi, puis cherche un candidat adéquat. En information imparfaite à propos des travailleurs disponibles, un poste vacant a une certaine probabilité d'être pourvu au cours d'une période donnée. Le processus d'*appariement* entre le poste vacant et les offreurs de travail prend du temps. Si le poste est pourvu, l'entreprise dégage un profit obtenu en soustrayant le coût du travail de la productivité du travailleur. Dans ce contexte marqué par des *frictions* informationnelles, l'entreprise ouvre des postes de travail et les déclare vacants tant que le coût associé est inférieur au bénéfice escompté. Ce dernier est le produit de deux termes : la probabilité de trouver un candidat approprié (fonction de l'état du marché du travail) et le profit réalisé lorsque le travailleur contribue à la production de biens. Dans ce cadre d'analyse frictionnel, on retrouve la propriété qu'une hausse du coût salarial diminue la demande de travail de l'entreprise (à savoir le nombre d'emplois déclarés vacants).

Voir le Document 3 (À propos de la demande de travail…) de l'Annexe 1 en ligne.

ÉQUILIBRES SUR LE MARCHÉ DU TRAVAIL

Considérons un marché du travail pour une certaine profession dans un environnement donné (unité géographique, conditions dans lesquelles la profession est exercée…). Supposons que l'information soit parfaite (absence de frictions) et que la *concurrence* soit *parfaite* sur ce marché du travail (aucun agent économique n'a de pouvoir de marché : tous considèrent donc que le salaire est donné ; la mobilité

est parfaite). L'addition des demandes de travail des entreprises concernées fournit une demande de travail pour cette profession. La quantité demandée de travailleurs (ou d'heures de travail) D décroît avec le niveau du salaire w. La somme des offres individuelles de travail engendre une offre de travail S sur ce marché (mesurée en effectifs ou en heures). Admettons que l'effet de substitution l'emporte sur l'effet revenu, si bien que la relation $S(w)$ est croissante.

Le salaire d'équilibre w^* est ici celui qui égalise l'offre et la demande de travail. Il y correspond une quantité de travail d'équilibre E^* (voir figure 2). Il s'agit d'un *équilibre partiel* dans la mesure où nous avons considéré un marché particulier conditionnellement à l'emploi et au salaire sur tous les autres marchés du travail et aux prix des biens et services. La propriété essentielle de l'équilibre E^* se comprend aisément si l'on fait un instant l'hypothèse qu'il n'y a qu'un seul type de main-d'œuvre dans l'économie. Alors en E^*, le taux marginal de substitution entre la consommation et le loisir est égal au salaire et à la productivité marginale du travail. Autrement dit, à l'équilibre, le marché concurrentiel recourt au travail jusqu'au point où le supplément de consommation requis par les offreurs pour travailler une unité de plus équivaut à la contribution de cette unité supplémentaire à la production et *in fine* à la consommation. On dit alors que *l'allocation des ressources est efficace*. Si le salaire était supérieur à w^*, par exemple du fait de la présence d'un salaire minimum w_{min}, la quantité de travail offerte dépasserait la quantité demandée. Cela engendrerait du *chômage involontaire*. Par cela, on entend que des individus sont désireux de travailler au salaire en vigueur mais ne trouvent pas d'emploi.

Définir l'équilibre concurrentiel est une chose, concevoir les mécanismes qui y conduisent en est une autre. Bien souvent, on recourt, explicitement ou non, à l'idée qu'un déséquilibre sur le marché du travail engendre une pression sur le salaire. Si en l'absence de salaire minimum, le salaire se situait au niveau w_{min}, le chômage entraînerait une compétition accrue entre les offreurs de travail. Le salaire baisserait jusqu'à atteindre le niveau d'équilibre w^*.

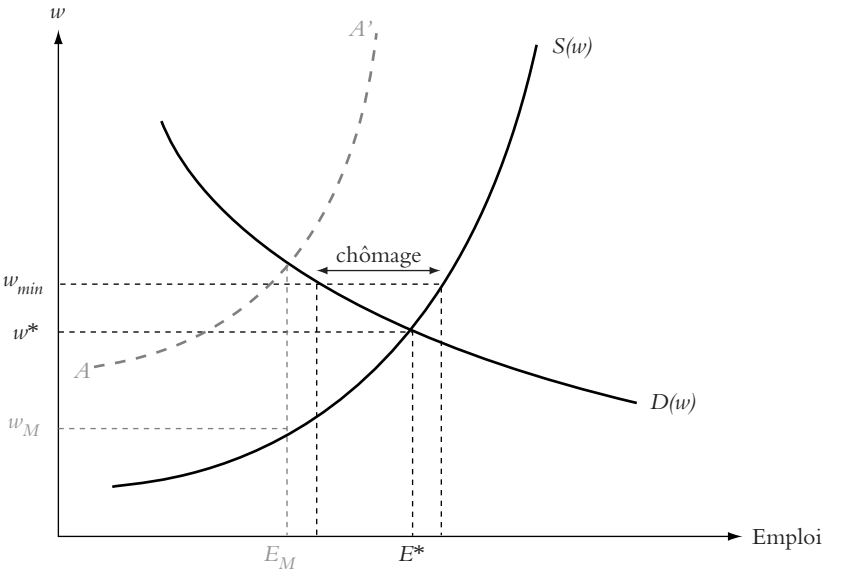


Figure 2 – L'équilibre concurrentiel sur le marché du travail

Dans ce cadre d'analyse, il y a donc un et un seul salaire d'équilibre pour le travail exercé dans le cadre d'une profession et d'un environnement donnés. Il n'est pas aisé de tester une telle prédiction car les notions de profession et d'environnement sont difficiles à cerner. Selon Mortensen (2003) notamment, cette propriété d'unicité n'est pas confirmée par les observations. Les entreprises ont des politiques salariales et donc un pouvoir de marché.

Le cadre théorique ci-dessus suppose implicitement une mobilité parfaite des travailleurs occupés par une entreprise quelconque. Si celle-ci payait moins que le salaire qui s'impose à elle, elle perdrait toute sa main-d'œuvre au profit des entreprises qui, elles, octroient ce niveau de salaire. En pratique, la mobilité est coûteuse pour un travailleur et ce, même si l'on fait abstraction de déplacements géographiques. L'information imparfaite est un facteur explicatif central. Changer

d'employeur implique une *prospection d'emploi* selon des mécanismes fort proches de ceux décrits plus haut dans le cas du chômeur. Cette mobilité imparfaite procure à l'employeur un pouvoir sur le marché du travail. L'analyse de l'équilibre sort des limites de ce chapitre lorsque les entreprises se concurrencent et la main-d'œuvre est imparfaitement mobile (voir Manning, 2003, 2006). Contentons-nous de considérer *une* entreprise quelconque qui, du fait de la mobilité limitée de la main-d'œuvre, fait face à une fonction d'offre de travail élastique, mais non infiniment, au salaire. Des barrières à l'entrée empêchent, en outre, l'entrée de concurrents. On parle de *monopsone*. Dans le cadre de la figure 2, réinterprétons *S(w)* comme l'offre de travail de l'entreprise. Elle utilise uniquement le facteur travail et choisit le salaire identique pour tous ses travailleurs. Un travailleur supplémentaire lui procure un surcroît de revenu égal à la productivité marginale du travailleur (si le prix de vente est normalisé à l'unité). Réinterprétons donc *D(w)* comme la demande de travail de cette entreprise. Pour attirer ce travailleur supplémentaire, l'entreprise devra augmenter son salaire pour l'ensemble de sa main-d'œuvre (sans quoi, il y a de la discrimination). Le coût marginal d'un travailleur supplémentaire est, dès lors, non seulement croissant, il se situe au-dessus de la courbe *S(w)*, et a une pente plus forte (c'est la courbe AA' de la figure 2). La firme maximise son profit en égalisant ce coût marginal à la productivité marginale du travailleur. Elle occupe donc E_M travailleurs au salaire w_M. Dans le même contexte, mais en présence d'une entreprise au comportement concurrentiel, l'offre et la demande s'égaliseraient à l'équilibre. L'emploi et le salaire seraient supérieurs. L'introduction d'un salaire minimum compris entre w_M et w^* relèverait à présent tant l'emploi que le salaire.

Le débat entre les défenseurs du salaire minimum et ses détracteurs a été ravivé par des travaux empiriques récents aux États-Unis et au Royaume-Uni, qui concluent à un effet faiblement négatif, voire inexistant, du salaire minimum sur l'emploi (Card et Krueger, 2000, Metcalf, 2007). Ils contrastent avec les résultats obtenus en Europe continentale (voir en particulier Kramarz et Philippon, 2001). Ce contraste

provient, pour partie, de niveaux différents du salaire minimum : au 1er janvier 2007, le salaire minimum légal national à temps complet s'élevait en France à 1 254 € contre 676 € aux États-Unis et 1 361 € au Royaume-Uni. Il s'explique aussi par des différences institutionnelles (en particulier, le degré de couverture des conventions collectives) qui procurent plus ou moins de pouvoir de marché aux entreprises.

Il y a de nombreux obstacles à l'émergence de l'équilibre concurrentiel et à son efficacité. L'information asymétrique entre l'employeur et ses travailleurs et les comportements discriminatoires en sont des exemples.

Voir l'Annexe 2 en ligne (l'éducation, la formation et la discrimination).

Les coûts d'embauche et de licenciement, les frictions et les coalitions de travailleurs ou d'employeurs engendrent de la concurrence imparfaite, conduisant parfois à des rigidités salariales. La rigidité des salaires peut aussi s'expliquer par une préoccupation d'assurance des travailleurs contre le risque de chômage (Drèze et Gollier, 1993). On admet de plus en plus que l'agent économique se soucie de la distribution de la richesse ou de celle des salaires, et pas seulement de son niveau de consommation et de loisir. Cela engendre des *externalités* qui minent l'efficacité de l'équilibre concurrentiel (Cahuc et Zylberberg, 2001, Fehr et Schmidt, 2006).

L'économie du travail s'intéresse également aux comportements d'éducation et de formation.

LA NÉGOCIATION COLLECTIVE

L'objet de la négociation collective varie fortement d'un système institutionnel à l'autre. La négociation salariale est la facette la plus évidente et la plus étudiée. D'autres thèmes de négociation sont le temps de travail, les conditions de sécurité dans l'entreprise, l'investissement

en formation continue des travailleurs, etc. Dans plusieurs pays, les partenaires sociaux participent à la gestion de la Sécurité sociale, à l'élaboration du droit du travail, à divers organismes consultatifs.

Le cadre légal de la négociation collective varie également beaucoup. Ici, les syndicats et organisations patronales reconnus par les autorités publiques ont le monopole de la négociation ; là, les syndicats ne sont reconnus que si un vote majoritaire des travailleurs en décide ainsi. Ici, une convention collective signée par les partenaires sociaux a force de loi pour tous les travailleurs et toutes les entreprises ; là, seuls les membres du syndicat sont concernés par la convention, ou bien les employeurs individuels ont l'option de ne pas adhérer à la convention. Face à une telle diversité, on retiendra qu'une analyse économique pertinente de la négociation collective requiert une bonne connaissance du cadre institutionnel.

Limitons-nous ici à la formation des salaires. Si le syndicat (de travailleurs) dispose d'un pouvoir de négociation et que le point de référence est la concurrence parfaite, on doit s'attendre à ce que les conséquences de l'intervention syndicale apparaissent généralement dommageables. N'a-t-on pas en effet conclu que l'équilibre concurrentiel était efficace sous un ensemble d'hypothèses ? Une analyse « ouverte » du rôle des syndicats requiert donc qu'on lève certaines de ces hypothèses.

Au préalable, la question de l'objectif du syndicat se pose avec acuité. Partant des préférences individuelles, on peut postuler que le syndicat est une organisation démocratique qui soumet ses décisions au vote. À l'autre extrême, certains auteurs considèrent que les responsables syndicaux ont une large autonomie entre deux élections professionnelles. Le plus souvent cependant, on postule que le syndicat poursuit un objectif qui augmente avec le niveau de salaire (réel) et le volume de l'emploi des travailleurs représentés (membres formels ou groupe plus vaste). Salaire et emploi sont, par hypothèse, des substituts. Les courbes d'indifférence du syndicat sont donc convexes dans un espace où l'emploi figure en abscisse et le salaire en ordonnée (voir la figure 3).

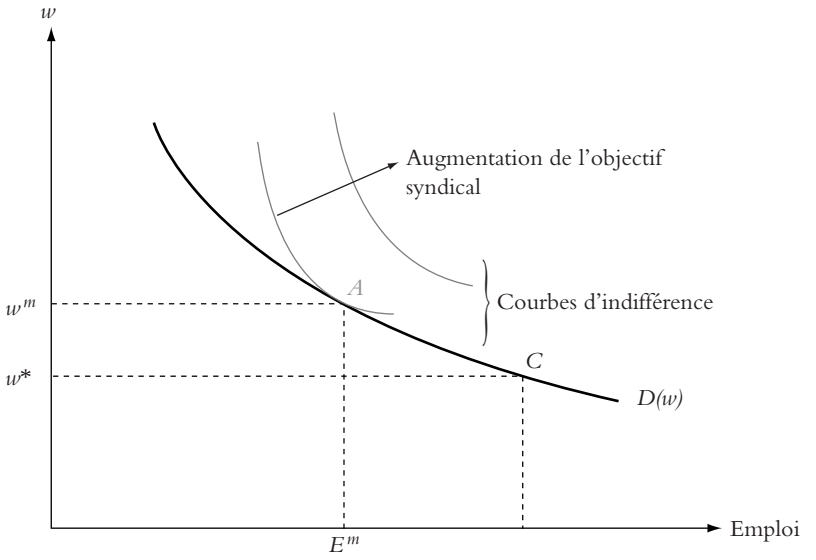


Figure 3 – Issues de la négociation salariale

Considérons donc une firme dont la demande de travail est notée *D(w)*. Comme les conventions collectives stipulent les niveaux de salaires mais rarement ceux de l'emploi, on supposera que le salaire est négocié et qu'ensuite l'employeur choisit le volume d'effectifs qui maximise son profit à ce salaire négocié. Lors de la négociation salariale, les partenaires sociaux prennent donc en compte l'impact de leur négociation sur la quantité de travail *D(w)* demandée par l'entreprise.

Commençons par supposer qu'à côté de cette entreprise, il existe un marché où prévaut la concurrence parfaite. Notons w^* le salaire d'équilibre sur ce marché concurrentiel. Tout travailleur qui n'a pas d'emploi dans l'entreprise considérée peut être embauché à ce salaire w^*. Celui-ci est donc le salaire de réservation des travailleurs lors de la négociation salariale. De son côté, l'entreprise peut fermer ses portes si ses profits sont insuffisants. Si, à l'équilibre concurrentiel, les profits sont nuls, il n'y a rien à négocier : tout salaire supérieur à w^* provoque

la disparition de l'entreprise. La négociation salariale est donc intimement liée à l'existence de profits au salaire w^* (par l'existence de barrières à l'entrée de concurrents, par exemple). Supposons dorénavant que de tels profits existent.

Le cas limite est celui où le syndicat décide seul du salaire. Dans ce cas, à la figure 3, son choix se portera sur le point où sa courbe d'indifférence est la plus haute tout en se situant sur la courbe *D(w)*. Il s'agit du point de tangence A. En A, il se peut que le profit de la firme soit trop faible (celle-ci pourrait délocaliser sa production). Si tel est le cas, le point A n'est pas atteignable. Pour la clarté de l'exposé et du graphique, nous ignorons cette possibilité. À mesure que l'entreprise acquiert du pouvoir de négociation (et que le syndicat en perd), l'issue de celle-ci va s'éloigner du point A et se rapprocher du point C. Autrement dit, le segment AC de la courbe *D(w)* représente l'ensemble des issues possibles de la négociation selon le pouvoir de négociation du syndicat.

Une propriété émerge sans surprise : l'emploi est moindre qu'en concurrence parfaite en raison du fait syndical. D'autres conclusions sont possibles lorsque la négociation collective porte sur le salaire et l'emploi (voir Cahuc et Zylberberg 2001 aux pages 326 et suivantes).

Abandonnons à présent cette hypothèse d'environnement concurrentiel. Dans un souci de simplicité, rapportons-nous à nouveau à la figure 2, où *S(w)* désigne l'offre de travail qui s'adresse à l'entreprise. En l'absence de syndicat, la solution est celle du monopsone (E_M,w_M). Grâce à un syndicat doté d'un pouvoir approprié de négociation, l'emploi et les salaires peuvent augmenter jusqu'à atteindre le niveau concurrentiel. Entre E_M et E^* en effet, l'entreprise a intérêt à rester le long de la courbe d'offre de travail car la productivité marginale du travail (qu'on lit en ordonnée par l'intermédiaire de la relation *D(w)*) est supérieure au salaire. L'emploi est alors déterminé par l'offre de travail. Si le pouvoir de négociation dépasse un seuil (celui qui conduit au salaire négocié w^*), il n'est plus profitable de demeurer sur la

courbe d'offre de travail (le salaire dépasse en effet la productivité marginale). Dès lors, l'emploi est fixé par la demande de travail où le salaire égale la productivité marginale du travail. Dans ce cadre d'analyse, la relation entre l'emploi et le pouvoir de négociation syndicale a donc l'allure d'un U inversé (voir Falch et Strom, 2007).

Voir le Document 4 (L'analyse des conventions collectives) de l'Annexe 1 en ligne.

LE CHÔMAGE

Cette partie présente tout d'abord quelques notions statistiques et appellations usuelles et s'attache brièvement aux faits observés depuis 1960. Ensuite, nous traiterons de la modélisation macroéconomique et de ses apports à la compréhension des causes du chômage.

La *population active* est la somme de la population en emploi et de celle en chômage. Par convention internationale du Bureau international du travail, pour être *chômeur*, il faut cumuler trois caractéristiques : être sans emploi (ne pas avoir travaillé, pas même une heure, au cours d'une semaine de référence), disponible (dans un délai de deux semaines au plus) et à la recherche d'un emploi (avoir mené des démarches spécifiques à cet effet au cours des quatre dernières semaines). On remarque que bénéficier d'une indemnité ne fait pas partie des critères. Dès qu'une des trois caractéristiques fait défaut, la personne est en emploi ou, en l'absence d'emploi, elle est déclarée « inactive » ou « hors de la population active ». Dans un certain nombre de cas, les frontières entre emploi, chômage et inactivité sont ténues. Par conséquent, le *taux de chômage* (la part de la population active en chômage ; nombre compris entre 0 et 1, souvent exprimé en pourcentage) n'est pas l'unique indicateur à considérer. On s'intéresse de plus en plus à la notion de *taux d'emploi*. Celle-ci rapporte l'emploi à la *population en âge*

de travailler (par convention, la population dont l'âge se situe entre 15 et 64 ans). Si on définit le *taux d'activité* comme la part de la population active dans la population en âge de travailler, le taux d'emploi n'est autre que le produit du taux d'activité par l'expression (1 – le taux de chômage). On voit ainsi que le taux d'emploi mêle une information sur le comportement de participation au marché du travail (le taux d'activité) et une autre relative à la fréquence du risque de chômage au sein de la population.

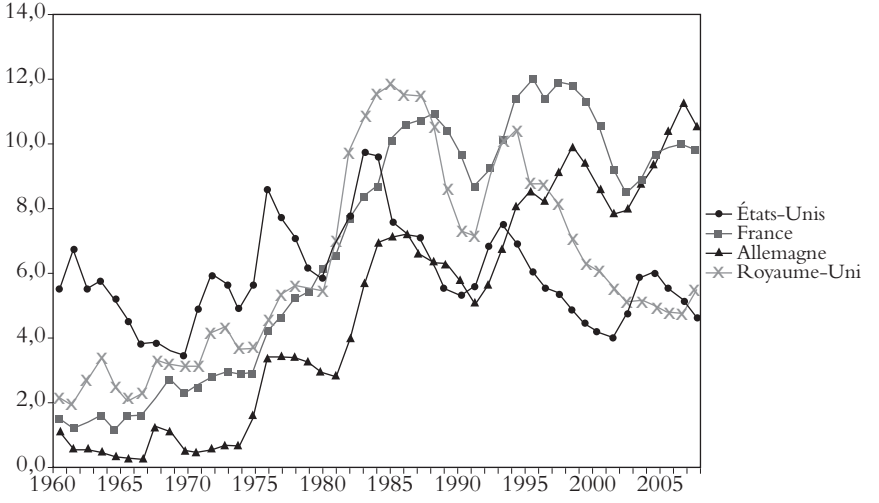


Figure 4 –Taux de chômage calculés par le Bureau of Labor Statistics (calculs fondés sur la population civile) : 1960-2006

La figure 4 présente le taux de chômage d'une sélection de pays de 1960 à 2006. L'établissement de statistiques comparables sur une longue période est un réel défi en raison notamment de ruptures statistiques et de spécificités nationales (la réunification allemande, par exemple). Quoi qu'il en soit, on constate que le taux de chômage présente d'amples fluctuations autour d'une tendance. Des années 1970 jusqu'au début ou au milieu des années 1980, la hausse tendancielle est générale (la Suède

fait exception). Ensuite, on assiste à une divergence : aux États-Unis et au Royaume-Uni, le taux de chômage a tendance à baisser, tandis qu'en Allemagne et en France la tendance demeure à la hausse.

Pour mieux identifier les causes du chômage, les économistes du travail et les macroéconomistes élaborent des modèles explicitant tant la demande de travail que les comportements de formation des salaires. Ces modèles ne peuvent expliciter tous les comportements ni toutes les interactions. Ils considèrent que certains facteurs sont *exogènes* (c'est-à-dire non expliqués par le modèle). Il s'agira, par exemple, des caractéristiques technologiques ou du pouvoir de négociation des syndicats. Supposons que les facteurs exogènes restent stables au cours du temps (ou varient à un taux constant). Partant d'une situation initiale, l'économie modélisée suit une trajectoire qui, à terme, converge vers un état stable (ou « d'équilibre ») généralement unique. Dans cet état, si du chômage s'observe, on le qualifiera de *chômage d'équilibre*. Les comportements modélisés permettent alors de mettre en évidence les déterminants du chômage d'équilibre. Ces derniers font intervenir les facteurs exogènes *via* les comportements et interactions modélisés. Seule une action sur ces déterminants exogènes peut permettre de faire varier le chômage d'équilibre. Ce chômage d'équilibre est compatible avec une inflation stable. On parle alors de chômage non accélérateur d'inflation (plus communément appelé *NAIRU : Non Accelerating Inflation Rate of Unemployment*).

Intéressons-nous à présent à quelques modèles macroéconomiques. L'approche dite « classique » suppose des prix parfaitement flexibles qui équilibrent en permanence les marchés. Selon cette approche, il n'y a pas, à proprement parler, de chômeurs mais bien des personnes qui choisissent d'être inactives au salaire en vigueur. Les fluctuations d'emploi sont dues à des chocs dits *réels* (affectant des grandeurs telles que la productivité du travail). Seule une forte élasticité de l'offre globale de travail permet de rendre compte de la réalité. Or, les études empiriques ne confirment pas une telle hypothèse. Dès lors, pour garder un marché du travail équilibré, il faut de grandes fluctuations de salaires réels face aux chocs réels. Cela ne s'observe guère, en particulier sur le continent européen.

Selon la vision keynésienne, le salaire *nominal* est rigide à court terme. Le marché du travail ne s'équilibre donc plus à chaque instant. La vision traditionnelle consiste à introduire une relation (sans fondement théorique) entre le taux de variation du salaire nominal d'une part, le taux de chômage, la croissance de la productivité et le taux de variation du niveau général des prix d'autre part. Cette relation porte le nom de *courbe de Phillips (augmentée)*, suite au travail empirique de ce dernier publié en 1958. Moins la croissance des salaires nominaux est sensible au niveau du chômage, plus les *rigidités réelles* sont fortes. Plus lent est l'ajustement de la croissance salariale nominale aux variations du niveau des prix, plus les *rigidités nominales* sont fortes. Dans ce cadre d'analyse, la hausse quasi générale du taux de chômage observée durant les années 1970 résulte de chocs défavorables (hausse des prix du pétrole et baisse du rythme de croissance de la productivité). Plus les rigidités réelles étaient fortes, plus le chômage d'équilibre s'est élevé suite à ces chocs. Là où les rigidités nominales étaient fortes, des politiques macroéconomiques (par exemple, monétaires) accommodantes ont pu limiter temporairement la hausse du chômage observé, au prix d'une croissance de l'inflation.

La persistance de taux de chômage élevés au début des années 1980 s'explique par la mise en place de politiques restrictives visant à réduire l'inflation. Lorsque l'inflation se stabilisa, le maintien de taux de chômage élevés dans de nombreux pays a conduit à se pencher sur les mécanismes de persistance de ce dernier. Parmi ceux-ci, le développement du chômage de longue durée a attiré l'attention. Divers arguments ont été avancés pour expliquer ce développement et ses conséquences sur le taux de chômage d'équilibre. Citons l'insuffisante prise en compte des chômeurs par la négociation collective, le manque de soutien public efficace à la réinsertion des chômeurs, la perte de savoir-faire à mesure que le chômage s'allonge. Selon l'état des connaissances actuelles, de tels mécanismes peuvent expliquer que le taux de chômage ne rejoint que lentement son niveau d'équilibre suite à des chocs économiques temporaires.

Ces travaux ont marqué le début d'un grand intérêt pour le rôle des institutions du marché du travail dans la détermination du niveau de chômage. Par institutions, on entend en particulier les conditions d'indemnisation du chômage, les politiques dites « actives » du marché du travail (formation des chômeurs, conseil, guidance et contrôle de ceux-ci, etc.), la protection de l'emploi et les mécanismes de négociation salariale. Les études comparatives internationales ne convergent pas en tout point (voir, par exemple, le chapitre 18 de Carlin et Soskice, 2006). Les auteurs concluent *souvent* qu'il y a une corrélation positive entre les indicateurs de « générosité » de l'assurance-chômage, de pouvoir syndical, de protection de l'emploi et de pression fiscale d'une part, et le niveau du taux de chômage d'autre part. En revanche, plus la coordination des acteurs de la négociation collective est forte, plus faible est le taux de chômage. Une plus grande compétition sur les marchés des biens permet en outre, généralement, d'abaisser le niveau du chômage. Ces études s'appuient cependant sur des indicateurs synthétiques assez grossiers. À titre d'exemples, l'architecture précise de la protection de l'emploi et l'interaction entre l'indemnisation du chômage, les prélèvements obligatoires et les politiques dites « actives » ne sont perçus qu'imparfaitement par ces indicateurs. L'étude de pays particuliers peut, à cet égard, être utile (Carlin et Soskice, 2006, chapitre 18). Il en va de même des évaluations microéconomiques de politiques du marché du travail qui exploitent des trajectoires individuelles de travailleurs et de chômeurs.

Le manque de fondement théorique clair à la courbe de Phillips préoccupe de longue date les macroéconomistes. Lorsqu'on exploite cette relation empirique pour prédire l'impact de changements de politiques, rien ne garantit en effet que ce changement ne modifie pas cette relation fondée sur des corrélations historiques (cette critique est associée au prix Nobel R. Lucas). Pour éviter cet écueil, il faut des fondements microéconomiques clairs aux relations macroéconomiques. Dans cette perspective, on partira, par exemple, de l'équation de formation des salaires découlant d'un modèle microéconomique de la

négociation collective puis, d'une manière ou d'une autre, on étendra cette équation à un cadre macroéconomique (d'une région ou d'un pays). Cela conduit à une relation de long terme entre le (logarithme du) *niveau* du salaire et le taux de chômage. Cette relation appelée la « courbe de salaire » *(the wage curve)* a fait l'objet de nombreux travaux empiriques sur données macroéconomiques et microéconomiques (Blanchflower et Oswald, 1994, 2005). Une nouvelle relation empirique de long terme émerge de ces travaux. Selon celle-ci, à un doublement du taux de chômage serait associée une baisse d'environ 10 % du niveau du salaire réel. En dépit de différences institutionnelles, parfois fortes, cette relation s'observe, curieusement, dans beaucoup de pays. Dans la même perspective méthodologique, d'autres travaux permettent, de nos jours, de fonder les traits essentiels de la courbe de Phillips (voir Blanchard et Gali, 2005). Au-delà de leur apport méthodologique, ces travaux mettent en évidence des déterminants plus complexes du taux de chômage d'équilibre (voir Cahuc et Zylberberg, 2001, p. 406-414).

Signalons enfin le développement d'une analyse macroéconomique du chômage fondée sur la présence de frictions sur le marché du travail (Pissarides, 2000). À tout moment, de nombreux emplois sont détruits et d'autres se créent, entraînant d'importants mouvements de personnes sur le marché du travail. Les principales explications de la montée et de la persistance du chômage sont, selon cette approche, la présence de chocs macroéconomiques sur la demande de travail et une perte d'efficacité du processus d'appariement entre les postes vacants et les demandeurs d'emploi.

Voir le Document 5 (L'analyse du chômage) de l'Annexe 1 en ligne.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 8

Croissance économique et développement

FRANÇOIS BOURGUIGNON

Directeur d'études à l'École des hautes études en sciences sociales, directeur de l'École d'économie de Paris

À l'aune de l'histoire, il y a quelque chose de magique dans la croissance et le développement économique qui marquent l'époque contemporaine. Selon les estimations du grand historien économiste Angus Maddison, le niveau de vie des habitants de la terre n'a pratiquement pas varié durant le premier millénaire de notre ère, et la population du globe n'a elle-même crû que très lentement, d'environ 15 % en mille ans. Une croissance positive, mais très lente, s'installe durant les siècles suivants. De l'an mille à l'aube de la révolution industrielle, soit en 1820, la population mondiale est multipliée par à peu près 4 pour atteindre l'effectif de 1 milliard tandis que le produit intérieur brut (PIB) réel par tête (la mesure la plus usuelle du niveau de vie) augmente d'environ 50 %. La croissance économique fait donc un peu plus que compenser l'augmentation de la population. Puis, c'est le coup de théâtre ! De 1820 à nos jours, soit en moins de deux siècles, la population mondiale est multipliée par 6, le niveau de vie moyen par 8 et la production mondiale par 50 ! Au cours des 19e et 20e siècles, le revenu mondial par tête, corrigé de l'inflation des prix, a augmenté au taux annuel de 1,2 %, contre environ 0,05 % durant les siècles précédents. Ce taux a même eu tendance à augmenter au cours du temps. Légèrement inférieur à 1 % jusqu'à la Première Guerre mondiale, il a

atteint le double sur le demi-siècle qui a suivi la Seconde Guerre mondiale et s'est encore accéléré depuis le tournant du millénaire. On estime aujourd'hui que le potentiel de croissance du niveau de vie moyen des habitants du globe pourrait être de l'ordre de 2 % sur la très longue période.

La croissance du bien-être matériel s'est accompagnée de progrès tout aussi considérables en matière de santé ou d'éducation. On peut déduire de certains recensements effectués à l'ère romaine que l'espérance de vie à la naissance pouvait être de l'ordre de 25 ans dans les premiers siècles de notre ère. Il est probable que ce chiffre a peu progressé avant le milieu du deuxième millénaire. À l'aube de la révolution industrielle, l'espérance de vie était encore inférieure à 40 ans en Angleterre. Dans les pays européens et leurs extensions dans le nouveau monde (États-Unis, Canada, Australie…), on vivait en moyenne 50 ans au tournant du 20e siècle, plus de 65 ans au lendemain de la Seconde Guerre mondiale, et plus de 75 ans à l'aube du troisième millénaire. En 1870, le nombre moyen d'années de scolarité de la population mondiale était d'une seule année. Il était de sept années en l'an 2000. Dans ces deux domaines, santé et éducation, le progrès tend aussi à s'accélérer.

Bien entendu, cette fantastique accélération du développement économique n'a pas été uniforme. Tant s'en faut. Derrière les moyennes que l'on vient d'évoquer se cachent des disparités régionales ou nationales colossales. Toujours selon les chiffres de Maddison, le niveau de vie moyen de l'Europe occidentale était le double de celui de l'Asie à l'aube de la révolution industrielle. Le rapport était de 5 un siècle plus tard, et de 10 en 1970. Il a depuis diminué sous l'effet de la croissance rapide de l'Asie du Sud-Est, et surtout de la Chine à partir de 1980. Si la comparaison est effectuée avec l'Afrique, l'évolution est encore plus dramatique. Le niveau de vie africain était probablement dans un rapport de 1 à 3 comparativement à l'Europe au début de la colonisation. Il était de 1 à 13 à la fin du deuxième millénaire. *Grosso modo*, le niveau de vie moyen observé actuellement en

Afrique subsaharienne était celui de l'Europe occidentale et ses extensions du Nouveau Monde vers le milieu du 19e siècle. L'Asie actuelle serait, quant à elle, au niveau des pays aujourd'hui développés au début du 20e siècle.

Il faut se méfier de ces comparaisons trop éloignées dans l'espace et le temps. Mais les écarts suggérés par les indicateurs de niveaux de vie et tous ceux qui s'y rattachent sont sans nul doute considérables. Ce sont eux, et le fait qu'ils continuent parfois d'augmenter, qui expliquent l'importance croissante accordée au développement économique par les économistes, les spécialistes des autres disciplines des sciences sociales et les politiciens.

Cette notion de développement économique est complexe. Elle englobe en fait tous les phénomènes qui participent à l'évolution d'une société, et qui trouvent leur origine dans la sphère économique ou qui conditionnent l'évolution de celle-ci. Mais la croissance, c'est-à-dire l'augmentation du volume des richesses dont dispose une société aux fins de consommation, d'investissement ou de production de biens publics (comme l'éducation, la défense ou la culture), peut être considérée comme le principal moteur de ce développement. Sans augmentation continue de ses ressources, ou de leur productivité – c'est-à-dire la quantité moyenne de biens qu'elles permettent de générer –, toute société serait condamnée à la stagnation et très probablement, après un laps de temps, au déclin. Grâce à la croissance, au contraire, une société peut non seulement améliorer le niveau de vie de ses membres, mais aussi leur éducation, leur santé, leur sécurité ou leur accès à des services sociaux. Ces modifications peuvent à leur tour conduire au bout d'un certain temps à des transformations plus profondes de la société, qu'il s'agisse des structures de décision collective (gouvernement démocratique ou non, centralisé ou décentralisé, etc.), de la famille ou des relations sociales, ces transformations étant elles-mêmes susceptibles de modifier les conditions de la croissance.

C'est sur la croissance économique en tant que telle, et non pas sur ces autres dimensions, plus générales, du développement que porte ce chapitre. La question fondamentale à laquelle on s'intéresse est celle des déterminants de la croissance.

Qu'est-ce qui fait qu'un pays croît plus rapidement qu'un autre durant une période donnée ? Y a-t-il une mécanique immuable qu'il suffit de mettre en marche pour que la richesse d'un pays s'accroisse de façon régulière ? Est-ce une question de chance, qui tient à des conditions particulières dans un environnement mondial donné ? Ou la croissance exige-t-elle avant tout la mise en œuvre de politiques adaptées à ces conditions particulières ? Si tel est le cas, qu'est-ce qui explique alors que certains pays ne parviennent pas à identifier ou appliquer ces politiques ? Telles sont les questions successives que se sont posé les économistes intéressés par la croissance. On se propose ici de résumer brièvement les réponses apportées et l'état actuel de leur réflexion. On se concentrera plus particulièrement sur le cas des pays en développement plutôt que sur celui des pays développés pour lesquels les problèmes actuels seraient peut-être plus les conséquences négatives de la croissance économique sur l'environnement que la croissance elle-même.

EXISTE-T-IL UNE MÉCANIQUE DE LA CROISSANCE ?

La croissance économique était vue par les économistes classiques du 18e et 19e siècles, de Smith à Ricardo et Marx, comme résultant avant tout d'un mécanisme d'« accumulation » de moyens de production. Selon cette vision de la croissance, une société croît en augmentant les équipements qu'elle peut mettre à disposition de sa main-d'œuvre pour la rendre plus productive. Ces équipements peuvent être de natures très diverses : des machines, des moyens de transport, des routes, des réseaux de lignes électriques, des tracteurs agricoles, etc. Dans le bilan des entreprises, ils correspondent aux « immobilisations » ou encore à leur « capital ».

Une analogie avec l'agriculture permet de saisir le mécanisme de base de l'accumulation de capital. La récolte de blé d'un agriculteur dépend de la quantité de blé qu'il a semée. En une année donnée, cette quantité est le « capital » de l'agriculteur. Ce qui limite ce capital, c'est le besoin de consommation de l'agriculteur et de sa famille. Si l'on suppose que l'agriculteur met de côté un pourcentage s de sa récolte pour l'ensemencement de l'année suivante, et que chaque grain de semence produit a grains au moment de la récolte. Alors, on peut voir que la production de l'agriculteur en une année donnée est $s \times a$ fois celle de l'année précédente. Si $s \times a$ est plus grand que l'unité, il y a croissance de la production, de la consommation de l'agriculteur et de sa famille, et du stock de semence au même taux $s \times a - 1$. L'« épargne » de l'agriculteur, c'est-à-dire la quantité non consommée de sa production, entretient une croissance continue de la quantité de grains disponible pour l'ensemencement.

Cet exemple est instructif à deux points de vue. D'une part par ce qu'il laisse dans l'ombre : comment l'agriculteur peut-il étendre sans limites sa récolte si la quantité de terre qu'il peut exploiter est limitée, ou si la main-d'œuvre qu'il peut engager pour labourer, ensemencer puis moissonner, est aussi limitée ? Dans ce dernier cas, il peut toujours augmenter les salaires qu'il offre à ses employés, mais alors son propre revenu et donc son épargne et sa capacité d'accumulation en seront réduits. D'autre part, par le rôle joué par le « comportement » de l'agriculteur en matière d'arbitrage entre consommation et épargne.

Sur le premier point, et en élargissant à la totalité des entreprises plutôt qu'un agriculteur isolé, il existe à l'évidence des limites au produit de l'activité économique. Le nombre de machines, de bâtiments disponibles pour y loger de l'industrie, la main-d'œuvre les faisant fonctionner, selon son degré de qualification, la terre, les infrastructures de transport, d'énergie, de communication, sont autant d'éléments en quantité fixe en un point du temps qui limitent le produit total de l'économie. On appelle tous ces éléments, qui préexistent à l'activité économique durant une période donnée, des « facteurs de produc-

tion ». Pour que la production croisse d'une période à l'autre, il faut que les quantités disponibles de ces facteurs augmentent. Il en sera ainsi si la société dans son ensemble procède pour chacun d'entre eux à l'« accumulation » décrite dans le cas de l'agriculteur, c'est-à-dire si elle consacre une partie de la production permise par ces facteurs à l'augmentation de leur volume plutôt qu'à la consommation.

Dans cette perspective purement productive de l'activité économique, la croissance est complètement déterminée par l'accumulation de facteurs de production ou par la croissance exogène de la quantité disponible de ces facteurs. Avec le temps, les économistes ont eu tendance à se concentrer sur trois facteurs considérés comme les plus limitatifs à la production : le capital physique, c'est-à-dire les machines et les bâtiments ; le capital humain, c'est-à-dire le savoir-faire ou la qualification de la main-d'œuvre souvent approximés par son niveau moyen d'éducation ; l'importance de cette main-d'œuvre elle-même, c'est-à-dire le facteur « travail ». Selon la perspective productive précédente, il doit donc exister une relation entre la croissance de la production et la croissance de ces trois facteurs. En toute logique, elle doit dépendre du type de production et doit donc être spécifique à chaque pays. En même temps, si les connaissances technologiques et si les quantités de facteurs non pris en compte sont les mêmes, cette relation devrait être très similaire d'un pays à un autre.

L'argument précédent n'est complètement convaincant que si la relation physique qui lie la production aux facteurs de production est essentiellement invariante au cours du temps. Or, les connaissances technologiques, l'organisation de la production ou l'intensité des efforts effectués par la main-d'œuvre peuvent évoluer, permettant ainsi au volume de production de croître même si les quantités de facteur restent inchangées. Dans le cas de notre agriculteur, le progrès technique peut permettre une meilleure conservation des semences et augmenter le coefficient *a*, ou rendre disponibles des semences plus productives (par exemple en évoluant vers les OGM). L'augmentation du prix du blé peut aussi justifier un travail plus intensif tout au long

du cycle de production. On dira dans ces différents cas que la productivité totale des facteurs (PTF) a augmenté. Clairement, il s'agit là d'une nouvelle source de croissance différente de la simple accumulation de facteurs.

Cette vision simple de la mécanique de la croissance a eu une influence considérable sur la façon d'envisager le développement. Négligeant d'une certaine façon la PTF, ou en supposant qu'elle évoluait au cours du temps de façon largement exogène, c'est-à-dire indépendamment de la politique économique, plusieurs générations de décideurs et de planificateurs du développement ont concentré leur attention sur l'accumulation de capital. À la différence de la terre ou de la force de travail dont la quantité est fixe ou évolue de façon exogène, il s'agissait en effet du principal levier que pouvaient actionner les décideurs à travers leurs politiques d'investissement. Dans l'URSS des années 1930, l'Inde des années 1950 et dans la plupart des pays en développement – et certains pays développés comme la France – dans les années 1960 et 1970, une grande partie de l'attention s'est ainsi concentrée sur la stratégie optimale d'investissement conduisant au coefficient *a* le plus élevé et sur le financement de ces programmes, soit interne à travers l'épargne nationale, soit externe à travers l'aide au développement.

Les résultats obtenus n'ont pas toujours été à la hauteur des ambitions. Il est vrai que la croissance a été rapide dans un grand nombre de pays dans les 20 ou 30 années qui ont suivi la Seconde Guerre mondiale. Mais beaucoup d'autres phénomènes tels que la reconstruction ou la décolonisation, pour n'en citer que quelques-uns, poussaient les économies dans cette direction favorable. Passée cette période, baptisée en France « les Trente Glorieuses », la croissance a semblé tomber en panne dans des pays qui avaient eu jusque-là de bons résultats. Il est vrai que le premier coup d'arrêt a été donné par la crise mondiale issue du boom pétrolier des années 1970. Mais, même par la suite, de nombreux pays ne sont pas parvenus à renouer avec le rythme des Trente Glorieuses, comme si le modèle précédent n'était plus valide.

En revenant aux mécanismes de base esquissés plus haut, l'analyse que l'on peut faire de cette situation est assez simple. Deux points sont à souligner. D'une part, l'accumulation n'est que l'un des mécanismes de la croissance et, d'autre part, il n'est pas sûr que même ce mécanisme simple soit complètement sous le contrôle des décideurs.

On peut bien chercher à accumuler de la façon la plus rapide possible, l'impact de cet effort sur la croissance sera faible si la PTF est faible et, pis encore, si elle diminue. Les études empiriques réalisées ces dernières années pour essayer d'expliquer les disparités de croissance entre pays révèlent que l'évolution de la PTF joue un rôle majeur pour expliquer ces différences. On a trouvé que ce rôle pouvait être plus important que les rythmes d'accumulation, à tel point que certains économistes en sont même venus à mettre en doute le fait que l'accumulation soit un facteur significatif dans l'explication de la croissance (voir W. Easterly et R. Levine, « It's not accumulation : stylized facts and growth models », *World Bank Economic Review*, 2001). On savait depuis longtemps que la PTF était importante dans les pays développés. En 1955, l'économiste Robert Solow, qui devait recevoir plus tard le prix Nobel en partie pour ses travaux sur la croissance, avait montré que la PTF avait contribué à environ un tiers de la croissance totale du produit intérieur américain sur la très longue période. Des travaux plus récents montrent que son rôle pourrait être beaucoup plus important dans des économies en développement. De fait, les exemples d'économies où les efforts d'investissement n'ont pas débouché sur une accélération de la croissance sont nombreux. On peut penser, par exemple, à des pays pétroliers comme l'Algérie ou le Venezuela qui, malgré des investissements gigantesques après le boom des années 1970, ne sont pas parvenus à faire décoller leur secteur non pétrolier. Sur une période beaucoup plus longue, l'Union soviétique et ses satellites offrent l'exemple d'économies où les investissements ont effectivement été massifs, mais où la PTF est restée faible et a même diminué à partir d'un certain moment, faute d'inci-

tation économique adéquate. Dans l'autre sens, il y a aussi des exemples de pays qui ont opéré de véritables décollages économiques sans un effort initial considérable d'investissement. La Chine du début des années 1980 est probablement le plus frappant. Les réformes entreprises par Deng Xiaoping à partir de 1978 avaient en partie pour ambition de recréer des incitations individuelles à la production. C'est ainsi que les paysans reçurent la permission de vendre à leur profit les récoltes en excédent de ce qui leur était imposé par les services de planification. L'effet a été presque immédiat. Pendant quelques années, la production agricole a crû à plus de 10 % par an, avec des conséquences importantes pour le reste de l'économie, sans qu'aucun investissement d'envergure n'ait été nécessaire (l'investissement était déjà élevé à environ 28 % du PIB, niveau moyen observé durant la décennie précédente. Ce n'est que dans les années 1990 que la part de l'investissement dans le PIB a augmenté pour atteindre les 40 % et plus observés dans les années les plus récentes). Bel exemple de croissance, presque spontanée, de la PTF !

Le second point qui explique la variabilité des effets de l'accumulation entre pays, et entre périodes pour un même pays, est le comportement d'épargne et d'entreprise des agents économiques individuels. Même si un gouvernement veut se lancer dans une politique ambitieuse d'investissement, il ne peut contrôler directement que ce qui relève du secteur public, généralement certaines infrastructures essentielles comme les routes, les ports ou les écoles. Évidemment, un gouvernement peut toujours augmenter le périmètre du secteur public en créant des entreprises publiques dans les domaines dont il veut avoir le contrôle. Le cas limite est celui d'une économie où tous les moyens de production sont propriété publique, comme l'Union soviétique avant la perestroïka ou la Chine avant les réformes de 1978. Un cas intermédiaire est celui de ces pays, notamment africains, qui ont fait porter leur effort d'investissement par des entreprises publiques créées pour la circonstance dans les domaines les plus divers, de la production et distribution électrique à la production

d'engrais ou au raffinage de produits pétroliers. Mais le contrôle direct de l'État par l'extension du secteur public s'accompagne le plus souvent d'une certaine inefficacité relative et d'une PTF plus basse. Dans une économie où la plus grande partie de la production provient du secteur privé, l'accumulation et la croissance ne peuvent résulter que du dynamisme de ce secteur, de sa volonté d'augmenter ses profits en élargissant sa production et améliorant sa productivité.

L'expérience internationale des 30 années qui ont suivi les Trente Glorieuses suggère donc que si l'accumulation est sans nul doute un mécanisme important de la croissance, il n'est pas le seul et il n'est pas nécessairement sous le contrôle direct des gouvernements. Cela ne revient pas à mettre en doute le modèle de départ où la croissance de la richesse d'un pays dépend de la croissance des facteurs de production à sa disposition, mais à le réinterpréter. D'une part, en tenant compte de l'évolution de la productivité de ces facteurs dont on a vu qu'elle pouvait être très variable d'un pays ou d'une période à l'autre. D'autre part, en réalisant qu'il n'est possible d'influencer l'évolution de cette productivité ou le rythme d'accumulation dans le secteur privé que de façon indirecte, à travers les politiques économiques mises en place, y compris, bien évidemment, certaines dépenses d'infrastructures sous le contrôle des autorités nationales ou locales. D'une vision « mécanique » de la croissance, on passe donc à une approche en termes de politiques économiques.

Pour répondre aux premières questions posées au début de ce chapitre, on peut donc conclure qu'il n'y a pas de mécanismes immuables qui assurent la croissance sous certaines conditions physiques ou historiques propres à un pays, mais qu'il existe des politiques susceptibles de la déclencher ou tout au moins de la favoriser.

Quelles sont-elles ?

LES POLITIQUES ÉCONOMIQUES COMME DÉTERMINANTS DE LA CROISSANCE

Le raisonnement qui précède suggère que les politiques favorisant la croissance doivent être celles qui offrent au secteur privé un environnement dans lequel peut s'exprimer pleinement l'esprit d'entreprise, et à l'ensemble de l'économie un mode de fonctionnement efficace. Plusieurs objectifs viennent alors à l'esprit : minimiser les risques de tous ordres qui peuvent affecter la bonne marche d'une entreprise, assurer un environnement qui réduit autant que possible les coûts de transaction, qu'ils soient physiques ou administratifs, disposer d'un système de régulation qui permet d'entretenir le dynamisme du secteur privé, en particulier en empêchant l'apparition de situations de monopole conduisant à l'immobilisme.

Ces objectifs simples suggèrent à leur tour toute une série de politiques que l'on se contentera ici d'esquisser. S'agissant de la minimisation du risque, on peut d'abord penser au risque strictement économique. De ce point de vue, la stabilité macroéconomique qui découle de politiques budgétaires et monétaires rigoureuses paraît être une condition nécessaire à la croissance. Il en va de même de la stabilité politique. Des conflits sociaux, ethniques ou frontaliers, ne sont guère propices à l'investissement et la création d'entreprises. Ils entraînent plutôt la fuite des capitaux et des entrepreneurs. Dans le même ordre d'idée, on peut penser que la sécurité physique est importante et qu'il y a peu de chances que se crée un tissu d'affaires florissant au milieu d'activités criminelles.

S'agissant d'un environnement minimisant les coûts de transaction de l'entreprise, on peut de nouveau imaginer un grand nombre de composantes, par exemple la disponibilité d'équipements collectifs adéquats tels que des routes, une alimentation fiable en énergie ou encore des services efficaces de télécommunication. En l'absence de tels équipements, il reviendrait à l'entrepreneur de financer lui-même des substituts, ce qui ne peut qu'alourdir ses coûts. Il en va de même de la

disponibilité de main-d'œuvre qualifiée évitant à l'entrepreneur d'avoir à assurer la formation de son personnel. La qualité de l'infrastructure administrative est également importante. Par exemple, une administration ou une justice corrompues augmentent les coûts d'une entreprise, non seulement à travers les pots de vin payés mais aussi par les coûts de négociation et de renégociation avec ceux qui exercent ce pouvoir de contrôle.

Une entreprise en situation de monopole tend à vivre sur sa rente plutôt que d'innover et de chercher à ouvrir de nouveaux marchés. Pour que le secteur privé soit dynamique, il importe donc de le réguler pour empêcher de telles situations. Pour la production de biens et services non échangeables avec le reste du monde, comme la construction ou les services aux particuliers ou aux entreprises, la régulation doit entretenir un régime de concurrence en favorisant, en particulier, l'entrée de nouvelles entreprises sur le marché. Pour les biens échangeables avec le reste du monde – qu'ils soient importés ou exportés –, l'ouverture commerciale assure automatiquement cette concurrence. Des entreprises nationales non performantes ne seront pas capables de vendre leur production sur les marchés étrangers ou de résister aux importations.

On pourrait continuer l'énumération de ces politiques destinées à créer un « climat d'affaires » propice à la croissance économique. Mais on peut percevoir à ce stade le danger d'aller trop loin dans cet exercice. D'une part, certaines des recommandations précédentes ne sont favorables à l'accumulation et les gains de productivité que sous certaines conditions. Libéraliser brutalement les importations d'un bien en baissant les tarifs douaniers crée sans aucun doute de la concurrence, mais peut empêcher une entreprise nationale d'atteindre la taille de marché qui la rendra efficace. Il faut probablement libéraliser, mais de façon progressive. D'autre part, ces recommandations ne sont pas nécessairement indépendantes les unes des autres. Par exemple, le degré de corruption d'une administration peut dépendre de l'intensité de l'activité de régulation qu'elle affiche officiellement.

L'analyse statistique consistant à expliquer les différences de croissance entre pays par des variables décrivant ces diverses politiques révèle qu'elles jouent *en moyenne* un rôle significatif. Mais elles n'expliquent qu'une part minime des disparités nationales de croissance et sont donc relativement impuissantes à expliquer les performances de croissance d'un pays donné. Il est vrai aussi que les techniques statistiques utilisées pour cette analyse font intervenir ces politiques de façon additive dans la détermination de la croissance alors que, comme on l'a suggéré plus haut, il est vraisemblable que ce sont des combinaisons de ces politiques qui importent, et que leur importance dépend elle-même des caractéristiques d'un pays particulier. Malheureusement, l'observation d'un nombre relativement faible de périodes de croissance dans les différents pays du monde ne permet pas de disposer de suffisamment de données pour pouvoir mettre en évidence une relation statistique aussi complexe.

La façon dont ces diverses politiques peuvent affecter la croissance à travers le comportement des acteurs économiques privés est donc complexe. Il faut donc se garder de les transformer en recettes qui seraient applicables à tout pays et à tout moment. Certaines organisations internationales comme le Fonds monétaire international et la Banque mondiale n'ont pas eu cette prudence. S'agissant des économies en voie de développement, s'est établie dans le courant des années 1980 une sorte de doctrine sur les politiques qu'elles devaient suivre pour accélérer leur croissance. Cette doctrine d'inspiration ultra-libérale, à laquelle est resté attaché le nom de « consensus de Washington » du fait de la localisation des organisations qui la promouvaient, mettait en avant certains des principes listés plus haut, tels que la stabilité macroéconomique ou l'ouverture au commerce international comme remèdes à une croissance trop lente. Mais elle en ajoutait d'autres, notamment la privatisation nécessaire d'entreprises publiques inefficaces ou la suppression d'impôts et subventions entraînant une distorsion du système de prix. Ce qui est plus grave, est que ces organisations ont eu tendance à imposer ces politiques aux

pays qu'elles aidaient sans beaucoup de discernement, et en rendant leurs prêts ou leur assistance conditionnels à la mise en œuvre de ces principes.

Les résultats n'ont pas été à la hauteur des attentes. De façon générale, la croissance n'a pas accéléré suite à la mise en œuvre des politiques du consensus de Washington, même si certaines recommandations comme des politiques monétaires et budgétaires garantissant la stabilité macroéconomique semblent avoir atteint leur but. D'autres politiques, notamment celles qui concernent la privatisation, ont soulevé beaucoup de mécontentement sans toujours produire de gain d'efficacité. Qui plus est, les résultats les plus spectaculaires de croissance jamais obtenus, ceux de la Chine depuis 1978, se sont appuyés sur des politiques qui, dans une large mesure, allaient à l'encontre des recommandations du consensus de Washington.

En résumé, il est clair que dans les économies mixtes d'aujourd'hui, le secteur privé est l'artisan majeur de la croissance et ne peut faire son office qu'avec les politiques adéquates. Ce que suggère l'expérience de ces 60 dernières années est qu'il n'existe pas de recette unique, à l'opposé de ce que prône le consensus de Washington. La Chine est la championne de la croissance avec une économie qui est encore très contrôlée et un secteur d'entreprises d'État encore pléthoriques. L'Inde a aussi crû très vite au cours des 20 dernières années malgré un déficit budgétaire chronique important. Les tigres asiatiques se sont développés grâce à l'expansion de leur commerce international mais initialement avec une solide protection douanière. Ces expériences, comme l'analyse statistique approfondie des relations entre croissance et politiques économiques, suggèrent qu'il y a quelque chose de spécifique dans la combinaison de politiques à employer pour déclencher la croissance. Il faut seulement pouvoir identifier cette combinaison puis pouvoir la mettre en œuvre.

Ceci conduit à la dernière question que l'on veut poser dans ce court aperçu de l'économie de la croissance. Quels acteurs doivent prendre la responsabilité de l'identification des politiques les plus appropriées à accélérer la croissance et se charger de leur application ?

LE RÔLE DES INSTITUTIONS

Cette dernière question découvre une nouvelle perspective dans l'analyse de la croissance, celle des institutions qui gouvernent la décision publique. L'analyse des politiques propres à déclencher le développement dans un pays donné n'est pas nécessairement d'une difficulté insurmontable. En revanche, mettre en œuvre ces politiques peut se révéler impossible du fait de la structure de décision publique dans le pays considéré parce qu'elles peuvent aller à l'encontre des intérêts de ceux qui disposent du pouvoir politique.

Un phénomène fréquemment observé dans les pays qui n'ont pas amorcé ou sont en cours de transition démocratique est celui de la *captation* du pouvoir politique, et souvent économique, par certains groupes de la population ou encore certaines « élites ». Le développement de l'ensemble de la population peut ne pas être l'objectif principal de ces élites, mais bien plutôt leur enrichissement personnel à travers les rentes qu'elles peuvent tirer de leur situation de dominance. Or, on l'a vu, c'est justement ce type de pratique qui décourage l'initiative privée susceptible de dynamiser l'économie. Les réformes à effectuer pour que s'accélère la croissance peuvent être connues par tous, elles ne seront pourtant pas entreprises tant que les élites n'y verront qu'une source de perte pour elles. Le problème est essentiellement distributif. Les réformes peuvent conduire à une augmentation du revenu total de l'économie, mais certains groupes y perdront et, dans certains cas, c'est justement eux qui détiennent le pouvoir. Certes, on peut également penser à les compenser pour cette perte, mais comme cette compensation ne peut s'opérer que progressivement dans le temps, ceci exige des engagements réciproques de groupes sociaux ou politiques qui ne seront tenus qu'en présence d'institutions adéquates (sur cette question du rôle des institutions dans la croissance, voir le livre important de Acemoglu et Robinson, *Economic origins of dictatorships and democracy*, Cambridge University Press).

Les institutions qui président à la façon dont les décisions publiques sont prises peuvent donc en fin de compte se révéler les principaux

obstacles au développement. L'analyse historique permet de documenter de telles situations. Elle permet également de comprendre comment, dans certains cas, les institutions se sont modifiées d'elles-mêmes, conduisant alors à un changement de régime dans la croissance économique. Les réformes de 1978 qui ont finalement « réveillé » la Chine, pour reprendre le titre du livre prémonitoire d'Alain Peyrefitte, reposent sur un changement dans l'équilibre des pouvoirs au sein du parti communiste chinois. La perestroïka lancée par Gorbatchev est, elle aussi, une tentative de réforme qui va à l'encontre des intérêts d'une partie des dirigeants soviétiques. Elle échouera, mais en même temps elle précipitera le changement institutionnel, peut-être le plus radical de ce dernier demi-siècle. Dans ce domaine, encore moins qu'ailleurs, il ne peut cependant exister de règle ou de recette. On sait que, en dernier recours, une grande part des performances de développement d'un pays tient aux institutions qui sont en place et à la force de résistance de ceux qui peuvent en profiter. Mais on sait que leur renversement, quand il a lieu, est de nature essentiellement politique et tient souvent à des causes largement exogènes.

CONCLUSION

Ces considérations sur le rôle des institutions marquent le terme de notre tour d'horizon rapide des théories de la croissance appliquées aux pays en voie de développement. À ce stade, le lecteur sera peut-être tenté de conclure que, finalement, l'on ne sait pas grand-chose sur les déterminants de la croissance et que celle-ci reste d'un certain point de vue un peu magique. La Chine ou la Corée auraient eu la « chance » d'occuper une niche dynamique dans l'économie mondiale et l'auraient saisie, tandis que le Malawi ou le Niger ne l'auraient pas eue.

Une telle interprétation serait tout à fait erronée. En fait, notre connaissance des mécanismes de la croissance économique a fait des progrès considérables depuis les économistes classiques, ou même plus

récemment depuis Solow. On a également compris le principe fondamental qu'il n'y a pas à proprement parler de « recette » pour croître et se développer. Il y a toute une panoplie d'instruments et de politiques dont la pertinence dépend essentiellement des conditions spécifiques d'un pays en un point donné du temps.

Reste le problème d'inciter les dirigeants politiques ou économiques d'un pays à choisir les bonnes politiques lorsque celles-ci vont le plus vraisemblablement à l'encontre de leurs intérêts et que la société de dispose pas des institutions permettant l'engagement à une compensation équitable dans le futur. Les institutions politiques jouent là un rôle majeur et, de ce point de vue, on peut avoir le sentiment qu'un début de transition démocratique est nécessaire pour que les bonnes réformes soient entreprises dans ces pays.

Mais on peut être plus optimiste en comptant également sur l'effet de démonstration des expériences réussies de développement sur les pays voisins. Après tout, c'est en examinant l'expérience de Singapour que Deng Xiaoping a lancé les réformes en Chine, et c'est en comparant l'URSS avec les États-Unis que Gorbatchev s'est convaincu que des réformes étaient nécessaires pour rendre l'économie soviétique plus efficace. De même, l'expérience japonaise a beaucoup inspiré le développement de la Corée du Sud et de Taiwan. Dans ces conditions, on peut penser que s'il se confirme, le succès des quelques pays africains, comme le Ghana ou la Tanzanie, qui ont adopté durant ces deux dernières décennies des politiques économiques rigoureuses destinées à favoriser la croissance, pourra avoir un effet majeur sur le reste de la région malgré la vision pessimiste qu'ont certains de l'avenir de ce continent.

Il y a donc certaines raisons d'être optimiste, à moins que, évidemment, les conséquences négatives d'une croissance mondiale trop rapide et trop dommageable pour l'environnement n'introduisent à terme de nouvelles contraintes.

Chapitre 9

Commerce international et mondialisation

JEAN MARC SIROEN

Professeur à l'université Paris IX Dauphine

Le commerce international a fortement augmenté depuis la Seconde Guerre mondiale grâce à la libéralisation des échanges préconisée par les organisations internationales comme le GATT et l'OMC. Cette mondialisation commerciale n'est pourtant pas dépourvue d'ambiguïtés. Le débat doctrinal n'a pas cessé. La théorie n'explique que partiellement les causes et les effets de la mondialisation. Les grands principes multilatéraux des organisations internationales souffrent de nombreuses exceptions et n'empêchent pas la prolifération d'accords bilatéraux. L'entrée des pays émergents ou en développement dans le processus de mondialisation pourrait en influencer le cours, tout comme les tensions plus fortes sur les ressources et la montée des préoccupations environnementales.

Dans un sondage réalisé en juillet 2007, le *Financial Times* constatait que, dans des pays comme la France, les États-Unis et le Royaume-Uni, moins de 20 % des personnes interrogées considéraient avec certitude que la mondialisation avait des effets positifs pour leurs pays. Cette information révélait ainsi que, malgré plusieurs siècles de plaidoyers en faveur du libre-échange, les opinions publiques restaient sceptiques sur les vertus de l'ouverture au monde.

La pensée a souvent été fluctuante. Les périodes de libéralisme enthousiaste – le milieu du 19e siècle, le dernier tiers du 20e – ont alterné avec des phases de remises en cause – la fin du 19e siècle, les années 1930. À chaque épisode, des doctrines souvent anciennes ont été adaptées à une situation nouvelle.

Après la Seconde Guerre mondiale, les États-Unis étaient parvenus à imposer un ordre économique qui synthétisait des doctrines *a priori* inconciliables en délimitant ce qui relevait de la souveraineté des États, d'une part, et de celle des marchés, d'autre part. Cet ordre reposait sur des règles internationales incarnées par les institutions de Bretton Woods – Fonds monétaire international (FMI) et Banque mondiale – et, en matière commerciale, par le GATT (General Agreement on Tariffs and Trade).

Dès les années 1970, l'épuisement de ce compromis a conduit les dirigeants à élargir le territoire du marché, accélérant ainsi le processus de mondialisation. La libéralisation financière a néanmoins été plus spectaculaire que la libéralisation commerciale, même si les années 1980 et 1990 voient le ralliement raisonné au libre-échange des pays en développement les plus protectionnistes comme l'Inde, la Chine ou le Brésil.

Pourtant, après un quart de siècle de libéralisme économique, ne sommes-nous pas aujourd'hui à un nouveau tournant ? L'affirmation des pays émergents, les tensions sur les marchés de matières premières, de nouvelles préoccupations en matière, notamment, d'environnement ou de sécurité, une certaine déception à l'égard des bienfaits de la mondialisation et la remise en cause des organisations internationales ne conduisent-elles pas à « repenser » la mondialisation ?

LES GRANDS COURANTS DOCTRINAUX

Les doctrines ont été nombreuses. Raymond Aron les limitait à quatre : le mercantilisme, le système national intégré, le socialisme et le marxisme. Ces doctrines issues de la philosophie politique se distinguent

plutôt par la nature de l'ordre social que portent respectivement les États et le marché. L'adhésion au libre-échange est un critère commode, mais imparfait, pour les départager. En réalité, deux sont pérennes et s'affrontent depuis maintenant trois siècles et répondent différemment à la question : l'échange international attise-t-il ou, au contraire, apaise-t-il les conflits ?

La mondialisation conflictuelle du mercantilisme

La vision sans doute la plus ancienne et la plus populaire, voire la plus populiste, est celle d'un monde conflictuel par nature. Les intérêts des États sont opposés et incompatibles. Tout doit être mis en œuvre pour renforcer le pouvoir du Prince et lui permettre de s'accaparer des ressources limitées. L'échange international n'est acceptable que pour cette fin. Il doit, par exemple, conduire à accumuler les métaux précieux qui enrichiront le pays et financeront les armées. La mondialisation est une guerre économique, une lutte pour l'excédent de la balance des paiements et l'augmentation des réserves d'or ou d'argent. Ce mercantilisme, théorisé dès le 15e siècle, ne rejette donc pas la mondialisation. Ainsi, le 16e siècle des grandes découvertes fut à la fois un siècle de mondialisation et de triomphe mercantiliste.

Cette « obsession » de l'excédent commercial explique la nature conflictuelle des relations internationales. Les exportations des uns étant les importations des autres, dégager un excédent chez soi, c'est nécessairement imposer un déficit ailleurs. Comme l'exprimait le mercantiliste Antoine de Montchrestien (1575-1621), « *nous faisons autant de perte que l'étranger fait de gain* ». Le jeu est à somme nulle.

Le protectionnisme préconisé est très sélectif. Pour des pays comme la France de Louis XIV et de Colbert, dégager un excédent, c'est transformer des produits de base acquis au prix le plus bas possible pour exporter des biens de luxe à forte valeur ajoutée. Les spécialisations ne sont donc pas équivalentes et toutes ne sont pas « gagnantes ». Le

secteur préféré sera fréquemment l'industrie « lourde » (comme la sidérurgie), nécessaire à la fois au développement économique et à la pleine souveraineté des nations, qu'elle soit économique ou politique. Les canons ne sont-ils pas faits de fonte ? À des degrés et dans des contextes différents, la doctrine de la protection pour les « industries naissantes » (Hamilton, List) s'inscrit dans cette option industrialiste, comme d'ailleurs les stratégies de développement « staliniennes » qui exacerbent les vertus de l'industrie lourde, les politiques de substitution aux importations – doctrine dominante de l'après-guerre dans les pays du tiers-monde –, ou les théories de la dépendance longtemps très populaires en Amérique latine.

Le succès fut parfois au rendez-vous : les États-Unis et l'Allemagne au 19e siècle, l'Asie de l'Est (Japon, Corée du Sud) au 20e. Mais cette approche a aussi parfois conduit à des déconvenues plus ou moins catastrophiques dans les pays socialistes, en Algérie, en Inde, en Amérique latine.

La mondialisation harmonieuse du libéralisme

Dans *De l'esprit des lois*, Montesquieu (1689-1755) constatait que « *c'est presque une règle générale que partout où il y a des mœurs douces, il y a du commerce ; et que partout où il y a du commerce, il y a des mœurs douces* ».

Ce constat de double causalité se fonde sur la philosophie politique du 18e siècle, qui remet en cause l'État-Léviathan (Hobbes) et qui attaque cette obsession belliqueuse de l'excédent de la balance des paiements. Le mercantilisme reposerait, en effet, sur l'idée fausse que la richesse d'une nation se confond avec sa contrepartie monétaire. Mais la richesse ne trouve son origine que dans le travail productif.

L'excédent n'est d'ailleurs pas soutenable. David Hume (1711-1776) démontre ainsi que tout déséquilibre se traduit par une hausse des prix dans les pays excédentaires et, symétriquement, par une baisse dans le pays déficitaire, ce qui tend à rétablir « automatiquement » l'équilibre

des balances des paiements. Une cause irréductible de conflit entre les nations se diluerait donc dans la théorie économique. La « main invisible » d'Adam Smith (1723-1790) nous assure ainsi que :

> *« [même si] chaque individu… ne pense qu'à son propre gain…, il est conduit par une main invisible à remplir une fin qui n'entre nullement dans ses intentions… Tout en ne cherchant que son intérêt personnel, il travaille souvent d'une manière bien plus efficace pour l'intérêt de la société, que s'il avait réellement pour but d'y travailler. »*

La théorie du commerce international, telle qu'elle se développe à partir d'Adam Smith et de David Ricardo (1772-1823), appuie cette vision optimiste des relations économiques internationales. La théorie des avantages comparatifs démontre ainsi que, non seulement tous les pays pourront gagner à l'échange, mais encore que tous pourront y participer dès lors que les plus « inefficaces » d'entre eux (en termes de productivité du travail) le sont relativement moins dans certains secteurs dans lesquels ils devront alors se spécialiser.

Toutefois, si les gains mutuels sont assurés, son caractère égalitaire ne l'est pas. L'approche apaisante de la théorie du commerce international repose donc sur un postulat plus philosophique qu'économique : les individus – ou les États – rationnels acceptent de s'ouvrir à l'échange dès que le gain est positif, quel que soit le gain du partenaire. Comme l'écrit le père de l'utilitarisme, Jérémy Bentham (1748-1832) : « *Tout commerce est avantageux par essence, même pour celui pour lequel il l'est moins. Toute guerre est par essence ruineuse.* »

Les leçons de l'économie réelle

Aujourd'hui, l'économie expérimentale nous fait douter de la réalité de cette posture utilitariste. Les politistes « néo-réalistes » la récusent. Comme l'écrit un de leurs chefs de file, Kenneth Waltz (né en 1924) :

> *« Quand les États, confrontés à une possibilité de coopération avec des gains mutuels, se demandent s'ils vont coopérer, la question qu'ils se posent est de savoir comment le gain se partagera. Ils sont contraints de se demander non pas : "Est-ce que nous allons tous les deux gagner ?" mais : "Qui va gagner le plus ?" »*

Certes, l'économie académique dominante, issue du courant de la mondialisation harmonieuse, a apporté une multitude de nuances à cette vision idyllique, mais elle les a souvent reléguées aux marges de la théorie. Les « protections optimales » où le grand pays abuserait de sa position de monopole ou de monopsone (nombre limité d'acheteurs) sur les marchés mondiaux resteraient une curiosité intellectuelle. Elles risqueraient, d'ailleurs, de déclencher des guerres commerciales mutuellement ruineuses (Johnson). De même, il faudrait recourir au sempiternel exemple de l'effondrement du cours du café brésilien pour trouver un exemple de croissance appauvrissante (Bhagwati), qui risque de survenir lorsque le surcroît du volume d'exportation est plus que compensé par la baisse des cours mondiaux qu'il provoque. Quant aux industries naissantes, le système financier devrait être en mesure de financer durablement les secteurs concernés. L'État a-t-il d'ailleurs la compétence pour les identifier et donc les protéger ?

Dans cette conception, la persistance du protectionnisme ne peut donc pas être due à la volonté des gouvernants d'améliorer le bien-être de leurs citoyens, mais à la tendance naturelle de l'*homo œconomicus* d'utiliser tous les moyens disponibles pour maximiser ses propres intérêts. L'« économie politique de la protection » a ainsi développé une multitude de modèles qui ne considèrent plus la politique commerciale comme une variable exogène « tombée du ciel » et, encore moins, comme susceptible d'améliorer le bien-être. La persistance du protectionnisme dérive d'une demande de protection qui émane des groupes de pression et à laquelle répond une offre des responsables politiques. Les démocraties ne seraient d'ailleurs pas les mieux protégées de cette

déviation en laissant libre cours aux groupes de pression, ce qui, par ailleurs, contredit Montesquieu si on range la liberté politique dans les « mœurs douces ».

Mais le plus grand défi vient paradoxalement du modèle dominant, dit néoclassique, qui localise l'avantage comparatif des pays dans leurs dotations relatives en facteurs de production (modèle HOS pour Heckscher, Ohlin et Samuelson ; et modèle Stolper et Samuelson). Si ces auteurs ne remettent pas en cause l'essentiel, le gain mutuel à l'échange, ils montrent que, puisque les pays se spécialisent dans les biens qui utilisent le plus intensément les facteurs qu'ils ont en (relative) abondance, les facteurs de production initialement les plus rares le deviennent de moins en moins avec l'ouverture et voient donc leur rémunération diminuer. Ainsi, la situation des travailleurs non qualifiés devrait se dégrader dans les pays avancés, alors que, symétriquement, la situation des « capitalistes » et des travailleurs qualifiés devrait se détériorer dans les pays en développement.

Pourtant, la mondialisation de l'après-guerre a surtout concerné l'échange entre pays industriels aux dotations relatives en facteurs similaires. En conséquence, si les gains de l'échange étaient plus limités, l'ouverture avait aussi moins d'impact sur les inégalités. La « nouvelle théorie du commerce international », proposée notamment par Paul Krugman (né en 1953) montre la possibilité d'échanges entre pays similaires et donc dépourvus d'avantages comparatifs, dans des secteurs en situation de concurrence imparfaite où les produits sont différenciés et les coûts unitaires sensibles au volume de la production (économies d'échelle). D'une manière plus générale, la structure imparfaite des marchés « suffit » à créer de l'échange. Les gains viennent de l'accès à une plus grande diversité des produits. La diminution des prix vient, le cas échéant, des économies d'échelle et de l'érosion des pouvoirs de monopole des firmes nationales.

Cette approche, qui ne remet pas en cause le principe des avantages mutuels, a pourtant ébranlé, dans les années 1980, certaines convictions libre-échangistes. L'État n'est-il pas en mesure de manipuler les

marchés ? Subventionner Airbus, par exemple, ne vise-t-il pas à évincer les autres producteurs et donc à permettre au constructeur européen de conquérir un monopole mondial générateur de rentes ? Cette théorie de la « politique commerciale stratégique » réintroduisait le conflit dans les relations économiques internationales. Mais le manque de robustesse de ces théories aux hypothèses, les risques de représailles, l'incapacité des États à identifier les secteurs concernés ont conduit à un revirement de ses prosélytes, Paul Krugman notamment, ralliés à une mondialisation « innocente » des maux qui lui sont attribués.

LES INSTITUTIONS DU COMMERCE INTERNATIONAL

Depuis deux siècles, le commerce mondial croît plus vite que la production. Si la vague protectionniste de la fin du 19e siècle n'a pas ralenti ce mouvement – au contraire si l'on en croit Paul Bairoch – il n'en fut pourtant pas de même entre les deux guerres mondiales.

À la suite de la crise de 1929, la valeur du commerce mondial est divisée par trois. La récession et la déflation (baisse des prix) en sont la cause principale, mais le recours massif au protectionnisme a amplifié cet effondrement (loi Smoot-Hawley de 1930 aux États-Unis). Les pays ont été collectivement victimes d'un jeu non coopératif « perdant-perdant » où chacun aggravait la crise mondiale en espérant en atténuer les effets pour soi-même. Ce traumatisme a incité les vainqueurs de la Seconde Guerre mondiale à se doter d'institutions qui empêcheraient ces politiques d'« égoïsme sacré ». Les politiques commerciales nationales seraient encadrées par des règles internationales multilatérales.

Les institutions internationales

La charte de La Havane, signée en 1948, ne tranchait pas entre les doctrines conflictuelles et harmonieuses. Plus exactement, dans la lignée de Keynes (1883-1946), négociateur britannique des accords de

Bretton Woods (1944), elle validait la stratégie mercantiliste des économies menacées de sous-emploi, à condition toutefois que des règles du jeu suffisamment contraignantes effacent la fatalité du conflit.

Si les pays cherchent légitimement à « tirer » leur économie par les exportations, cette stratégie n'est cohérente au niveau international que si les mêmes pays acceptent d'importer davantage en ouvrant leur économie. Le libre-échange, tel qu'il est prôné, ne vise donc pas à faire jouer les avantages comparatifs pour exploiter plus efficacement des ressources limitées, mais à initier un processus de croissance où la locomotive des exportations permettrait de se rapprocher du plein-emploi et de favoriser l'augmentation des niveaux de vie. Les importations apparaissent comme le prix à payer pour accéder aux marchés d'exportations.

La très (trop ?) ambitieuse charte de La Havane n'a pourtant pas été ratifiée. Un accord transitoire, le GATT, signé l'année précédente entre 23 pays, va s'y substituer. D'un champ plus limité, il en reprend néanmoins la doctrine. Le principe fondamental est celui de la non-discrimination entre les « parties contractantes ». La « clause de la nation la plus favorisée » leur assure qu'ils bénéficieront du régime commercial le plus favorable pratiqué par leur partenaire ; chaque pays appliquera le droit de douane le plus bas à l'ensemble des autres. Le principe du traitement national garantit qu'une fois dédouané, un produit importé ne sera pas désavantagé par rapport aux produits nationaux.

Le multilatéralisme s'affirme également par les cycles de négociation (*rounds*) qui réunissent l'ensemble des parties. Sur la base de la réciprocité, ils visent à engager un nouveau processus d'ouverture des marchés. Ces cycles ont d'abord permis d'abaisser les droits de douane élevés hérités des années 1930. À partir des années 1960, ils s'élargissent aux barrières non tarifaires et à certaines règles du GATT. À l'issue du cycle d'Uruguay (1986-1993), l'accord de Marrakech (1994) crée une institution bénéficiant du statut d'organisation internationale, que n'avait pas le GATT : l'Organisation mondiale du commerce

(OMC). Celle-ci administrera le GATT mais également les nouveaux accords inclus dans le traité de Marrakech comme l'accord sur les services (AGCS) ou celui sur la propriété intellectuelle (ADPIC).

Plus, peut-être, que la fondation de l'OMC et l'extension de son champ, l'innovation institutionnelle la plus importante est la réforme de la procédure de règlement des différends. Pour être crédible, la coopération implique en effet un mécanisme de sanction. Contrairement à l'ancien GATT, la nouvelle procédure ne peut plus être interrompue à la seule discrétion d'un pays et notamment du pays mis en cause. Depuis 1995, date d'installation de l'OMC, les plaintes se sont en effet multipliées. Elles peuvent aboutir à des sanctions lorsque le pays visé ne suit pas les rapports des groupes spéciaux ou de l'organe d'appel. Si les grandes puissances commerciales – États-Unis, Union européenne – ont souvent porté plainte, elles ont fréquemment été mises en cause et elles ont le plus souvent perdu.

Les grands principes du multilatéralisme ne doivent pourtant pas faire illusion. La clause de la nation la plus favorisée n'empêche pas les traités de libre-échange ou d'union douanière (article 24) qui dérogent à la clause de la nation la plus favorisée. Les pays en développement bénéficient d'un traitement spécial et différencié. Si les cycles de négociation sont formellement « multilatéraux », les arbitrages se réalisent entre un nombre limité de pays, en général, les États-Unis et l'Union européenne (qui parle d'une seule voix à l'OMC). La réciprocité dans les concessions est, par ailleurs, difficile à vérifier et à appliquer, ce qui laisse la place à des comportements de passagers clandestins (*free-riding*) de pays qui espèrent bénéficier de l'ouverture des autres sans consentir eux-mêmes à l'ouverture de leurs marchés. Enfin, si la fonction de l'OMC est de promouvoir la libéralisation du commerce, ses fondements mercantilistes s'expriment par les possibilités de protections « non négociées », qu'elle laisse le soin aux pays membres d'administrer, quitte à les conditionner à des règles souvent imprécises : droits antidumping, droits compensateurs (antisubventions), clauses de sauvegarde, protection de la balance des paiements.

Les difficultés du cycle de Doha, ouvert en 2001, illustrent les limites de la libéralisation multilatérale. Soixante ans de libéralisation conduisent aujourd'hui les négociateurs à se heurter au « noyau dur » du protectionnisme : l'agriculture dans certains pays industriels (les États-Unis, l'Union européenne, mais aussi la Suisse ou la Corée), les industries ou les services plus ou moins naissants dans les pays émergents. L'élargissement de l'organisation à plus de 150 pays, le maintien de la règle du consensus, l'absence de date limite (2004 initialement), associés à des doutes croissants sur les vertus du libre-échange et la distribution équitable de ses gains, ont contribué à retarder l'achèvement du cycle. Quelques pays émergents – le Brésil, l'Inde, l'Afrique du Sud – seront parvenus à s'imposer dans le club très restreint des leaders de la négociation.

Les politiques commerciales nationales

Les règles multilatérales auraient dû réduire la marge de manœuvre des politiques commerciales nationales. En réalité, les pays conservent des tarifs moyens différents : 0 % à Hong Kong mais 30 % aux Bahamas (moyenne simple du tarif NPF – nation la plus favorisée – appliqué en 2006), avec des pics tarifaires qui dépassent parfois trois chiffres et qui révèlent les priorités de chaque pays.

Les droits « consolidés », que les pays de l'OMC s'engagent à ne pas dépasser, sont fixés à des niveaux tellement élevés que les tarifs appliqués se situent, en général, à des niveaux beaucoup plus bas, ce qui préserve une marge de flexibilité importante. C'est par exemple le cas des importations agricoles de l'Union européenne, où les droits de douane peuvent suivre les cours mondiaux pour préserver le revenu des agriculteurs.

La protection « conditionnelle », qui doit respecter les règles plutôt conciliantes de l'OMC, reste également un instrument favori, administré à la discrétion des gouvernements. Il permet de sanctionner les entreprises exportatrices étrangères qui vendent sur le marché

intérieur à un prix « anormalement » bas. Très controversé, il n'est pourtant plus l'apanage des seuls pays industriels. C'est ainsi l'Inde, avant les États-Unis, qui en fait le plus grand usage.

Les contraintes de l'OMC n'empêchent ni l'opacité, ni les contournements. Les barrières non tarifaires sont difficiles à détecter et à quantifier. Leur caractère protectionniste n'est d'ailleurs pas toujours acquis : quels critères permettraient de distinguer les normes sanitaires légitimes de celles qui relèvent d'un protectionnisme déguisé ?

Le GATT a d'abord été dévoué au seul commerce de marchandises. Avec l'accord de Marrakech, l'OMC couvre aussi les services et la propriété intellectuelle. Pour beaucoup de commentateurs, charger encore la barque en intégrant des « sujets » comme la politique de la concurrence, les investissements, les normes de travail, risquerait de la faire couler. Pourtant, développer la voilure permettrait aussi d'accélérer la navigation. On peut craindre que, si l'OMC ne régulait que les politiques commerciales *stricto sensu*, elle laisserait la voie libre aux contournements. Les commandes publiques peuvent privilégier les fournisseurs locaux, les firmes nationales peuvent imposer des barrières « privées » aux échanges lorsqu'elles disposent d'un pouvoir monopolistique sur le marché intérieur. Si le niveau des droits de douane est contraint par les accords, différentes formes de *dumping* (terme certes abusif) peuvent aboutir aux mêmes effets que les protections prohibées sans pour autant violer les règles de l'OMC : « dumping monétaire » par la sous-évaluation du taux de change, « dumping fiscal et réglementaire » notamment dans les zones franches, « dumping environnemental », « dumping social ».

Toutes les tentatives faites pour inclure ces sujets se sont heurtées à l'absence de consensus. Les pays en développement et émergents craignent, en effet, que cet élargissement autorise les pays développés à lever contre eux des mesures protectionnistes de rétorsion. Si le danger existe, il n'en reste pas moins qu'un des effets pervers sous-estimé de la libéralisation multilatérale des échanges est d'inciter les pays à substituer aux instruments traditionnels du protectionnisme des

instruments indirects encore plus inefficaces. Peut-être faut-il trouver dans cet inachèvement du multilatéralisme la cause principale de la prolifération des accords commerciaux.

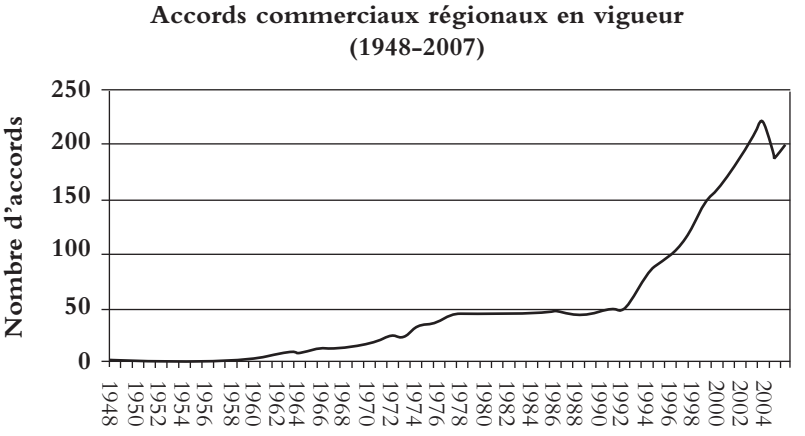
Les accords de préférence commerciale

Au 19e siècle, la libéralisation des échanges reposait sur la dynamique de traités bilatéraux, comme celui signé entre la France et la Grande-Bretagne en 1860. Ils préparaient néanmoins le multilatéralisme en insérant la clause de la nation la plus favorisée. Dans les années 1930, le repli des grandes puissances sur les empires coloniaux, éventuellement agrandis par la guerre (Allemagne, Italie, Japon) et sur des zones fédérées par des traités commerciaux, avait rendu soutenable un protectionnisme très dur. Néanmoins, malgré les réticences des États-Unis et après de longues discussions, le texte du GATT finira par inclure (article 24) une dérogation à la clause de la nation la plus favorisée en faveur des zones de libre-échange (élimination des droits de douane dans la zone) et des unions douanières (politique commerciale commune vis-à-vis du reste du monde).

Ces « zones de préférence commerciale » restent néanmoins contestées. Depuis Jacob Viner (1892-1970), on sait que, si ces traités « créent » du commerce au sein de la zone, ils en « détruisent » aussi à l'encontre des autres pays qui disposent d'avantages comparatifs.

Pendant longtemps, l'article 24 a été peu utilisé. Le traité de Rome (1957) suivi de la création de l'Association européenne de libre-échange (1960) furent donc des exceptions notables. Mais, au fur et à mesure que le GATT puis l'OMC s'élargissaient, le nombre de traités commerciaux augmentait avec une accélération dans les années 1990.

Beaucoup de ces accords regroupent des petits pays. L'élargissement aux services de la dérogation de l'article 24 (article V du GATT) a mécaniquement augmenté le nombre de notifications. Les événements géopolitiques, et notamment l'éclatement du bloc socialiste, comptent également dans cette prolifération.



Source : OMC (www.wto.org)

Figure 1 – Accords régionaux en vigueur (1948-2007)

Cette évolution ne doit pourtant pas être sous-estimée car de nombreux accords récents marquent une évolution radicale de leur nature. Jusqu'à la fin des années 1990, les accords réunissaient des pays proches géographiquement et économiquement. Ils concernaient essentiellement les droits de douane, même si l'Union européenne entrait dans un processus d'« intégration profonde » allant bien au-delà du libre-échange. Depuis, de nombreux accords concernent des pays géographiquement éloignés (par exemple : Japon-Chili), au niveau de développement très éloigné (par exemple : États-Unis-Maroc), mais, surtout, ils ne se limitent pas à la baisse des droits de douane et incluent aussi les sujets exclus de l'OMC : normes de travail, environnement, concurrence, investissement, marchés publics. Ces nouvelles règles, parfois assorties d'une procédure spécifique de règlement des différends, reproduisent les mécanismes de l'OMC, d'où les termes d'« OMC plus » ou de « minilatéralisme ».

Paradoxalement, le refus par les pays en développement de traiter ces sujets au niveau multilatéral conduit les pays industriels à les traiter au niveau bilatéral qui leur est beaucoup moins favorable… Le Maroc est

moins isolé dans une grande négociation multilatérale que dans un bras de fer avec les États-Unis ou l'Union européenne.

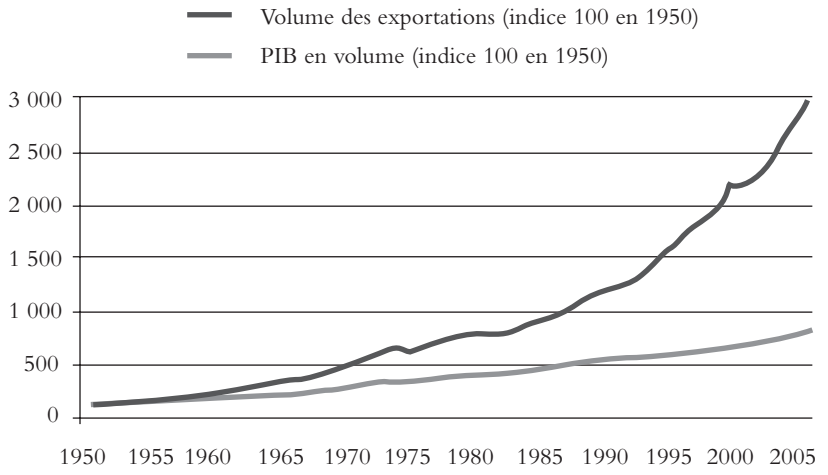
La prolifération de ce type d'accords a pour effet positif de contourner les carences de l'OMC et d'« aller plus vite ». Elle a aussi pour effet d'affaiblir le multilatéralisme qui, malgré ses insuffisances, atténue les asymétries entre les pays et uniformise les conditions du commerce international.

LA MONDIALISATION AUJOURD'HUI

Les historiens et les économistes aiment dire que la mondialisation n'est pas un phénomène nouveau. L'époque contemporaine aurait connu deux vagues de mondialisation assez similaires à un siècle d'intervalle. La première aurait commencé au milieu du 19e siècle après l'adoption du libre-échange par la puissance « leader », la Grande-Bretagne. La seconde aurait été engagée après la Seconde Guerre mondiale sous l'impulsion de la puissance américaine et se serait accélérée à partir des années 1980 avec la libéralisation des mouvements financiers, le ralliement au libre-échange des pays émergents (Inde, Brésil), ex-socialistes (Europe de l'Est), ou communistes (Chine, Vietnam). Les analogies doivent pourtant rester raisonnées. Sur de nombreux points, les deux épisodes de mondialisation s'opposent.

L'ampleur du processus de mondialisation

À première vue, l'analogie la plus évidente concerne l'ampleur de la mondialisation commerciale. Entre 1950 et 2006, le volume des exportations a été multiplié par 30 alors que celui du PIB ne l'a été que par 8. En conséquence, le taux d'ouverture de l'économie mondiale (commerce/PIB) n'a cessé d'augmenter, même s'il a fallu attendre les années 1980 pour qu'il rejoigne celui qui avait été atteint à la veille de la Première Guerre mondiale.



Source : d'après OMC (www.wto.org)

Figure 2 – Évolution des exportations et du PIB (1950-2006)

Ces indicateurs, pourtant très employés, sont trompeurs. Ils indiquent que davantage de produits franchissent les frontières ; ils ne signifient pas que les économies soient aujourd'hui plus dépendantes du commerce extérieur. Un taux d'ouverture (exportations/PIB) de 50 % n'implique pas que la moitié des emplois soit dédiée à l'exportation. Des pays comme Hong Kong et le Luxembourg dépassent d'ailleurs les 100 % ! L'explication est simple : le commerce, les exportations comme les importations, est comptabilisé à son prix de vente, alors que le PIB additionne des valeur ajoutées. Les réacteurs d'Airbus achetés aux États-Unis entrent dans la valeur des avions exportés par la France, mais pas dans son PIB puisqu'ils sont produits ailleurs.

Lorsque le contenu en importations des biens exportés s'accroît, le taux d'ouverture augmente mécaniquement puisque le dénominateur – la création de valeur ajoutée – diminue. Malheureusement, aucune statistique ne permet de mesurer la valeur ajoutée nationale des biens exportés.

De plus, lorsqu'on mesure les exportations mondiales, on comptabilise le même bien, incorporé souvent à d'autres biens, à chaque fois qu'il franchit la frontière. Avant de se retrouver dans le tableau de bord d'un avion, le morceau de silice sera ainsi comptabilisé dix, vingt fois ou plus sous la forme de silicium, de composant électronique, etc.

La croissance du commerce extérieur ne signifie donc pas nécessairement que la valeur ajoutée mondiale consacrée au commerce international augmente dans la même proportion, mais que les processus de production se sont allongés et internationalisés. Les pays ne se spécialisent plus dans des produits finis, mais dans certaines phases du processus de production. La France n'est pas spécialisée dans la construction aéronautique, mais dans l'assemblage de composants, par ailleurs importés. La Roumanie n'est pas spécialisée dans le textile-habillement, mais dans la couture de petits crocodiles sur les T-shirts importés de France (et rapidement réexportés).

Les formes de la mondialisation

Cette fragmentation de la production est sans doute la grande spécificité du processus actuel de mondialisation commerciale. Le ralliement des pays à bas salaires au libre-échange, la multinationalisation des firmes, la multiplication des zones franches (*maquiladoras* au Mexique, zones économiques spéciales en Chine, trafic de perfectionnement passif dans les pays de l'Est) auront donc eu pour effet d'approfondir la division « verticale » du travail.

Apparemment, les exportations industrielles ont augmenté plus vite que les exportations de produits primaires. Mais, pour reprendre l'exemple précédent, notre silice ne sera comptabilisée qu'une fois dans les exportations de biens primaires, mais très vite incorporée dans des produits transformés, donc industriels, la silice entrera dans la valeur des biens plusieurs fois exportés. La part des produits industriels dans le commerce mondial ne se confond donc pas avec leur part dans la valeur ajoutée mondiale, qui a d'ailleurs tendance à régresser.

Il en est de même des services. Certes, leur part dans le commerce mondial stagne, voire régresse (19 % en 1996 et 18,6 % en 2006 d'après l'OMC), en même temps qu'elle s'accroît dans les PIB, mais, contrairement aux produits industriels, la valeur ajoutée des services exportés n'est, en général, comptabilisée qu'une fois dans cette catégorie, ce qui sous-estime leur part dans la valeur ajoutée mondiale. L'internationalisation des services, facilitée par une circulation plus aisée des personnes et les immenses progrès dans les technologies de la communication, est sans doute un des phénomènes majeurs de la mondialisation actuelle que les indicateurs usuels ne permettent pas de mettre en évidence.

Les défis de la mondialisation

La poursuite du processus de mondialisation commerciale se heurte aujourd'hui à plusieurs défis.

Si, depuis une vingtaine d'années, la mondialisation s'est accompagnée d'une réduction de la pauvreté dans les pays en développement, elle n'a pas démontré sa capacité à réduire les inégalités. Inégalités entre pays d'abord, même si certains comme la Chine, le Vietnam ou l'Inde ont su s'immerger dans le processus de mondialisation. Si la théorie économique s'attendait à l'accroissement des inégalités dans les anciens pays industriels, elle prévoyait une réduction dans les pays émergents ou en développement – ce qui n'a pas toujours été constaté. Comme au Mexique ou en Chine, les inégalités interrégionales se sont parfois greffées sur les inégalités interindividuelles. Si les économistes préconisent des politiques de redistribution pour indemniser les perdants, ils se sont aussi attachés aux effets distorsifs des transferts. La mondialisation durcit d'ailleurs une contrainte d'efficacité qui fragilise ces politiques de redistribution.

Quelques années après l'écroulement financier des pays émergents, de 1996 en Asie à 2001 en Argentine (en passant par la Russie, la Turquie et le Brésil), la croissance de l'économie mondiale s'est accélérée. Ses

caractéristiques pourraient remettre en cause notre perspective de la mondialisation. La croissance a en effet provoqué une forte augmentation de la demande de biens primaires : pétrole, métaux ferreux, cuivre, blé, etc. Les tensions sur les prix ont redistribué les cartes au profit des pays en développement (y compris l'Afrique jusque-là délaissée) ou émergents (Chine, Russie…). Elles réhabilitent des choix sectoriels en faveur des matières premières. En relançant sa politique d'incitation à la production de biocarburants, le Brésil joue sur deux tableaux : exploiter son avantage comparatif dans ce type de production et pousser à la hausse le prix de ses matières premières exportées.

La mondialisation début de siècle renoue d'ailleurs avec certains aspects du mercantilisme : accumulation de réserves de change dans les pays émergents qui alimentent les « fonds souverains », dont la Chine ; stratégie « sauvage » de croissance par les exportations ; rivalités croissantes pour la captation des matières premières. Comme, en Europe ou aux États-Unis, les politiques agricoles pourraient s'inverser – des taxes à l'exportation se substituant à des subventions à l'exportation dans les productions les plus sensibles –, les pays importateurs de biens alimentaires risquent d'être les plus affectés, ce qui pourrait remettre en cause la tendance observée à la réduction de la pauvreté.

Les fonds souverains

Fonds d'États visant à placer plus efficacement les réserves que les placements traditionnels, comme les bons du Trésor. Ils ont notamment pris des participations importantes dans des banques occidentales fragilisées par la crise des *subprimes*.

Cette évolution se soucie assez peu des contraintes environnementales, même si elle peut s'en prévaloir comme la défense, par les États-Unis ou le Brésil, des biocarburants dont le bilan écologique est très controversé. L'écart croissant entre l'offre et la demande de biens primaires permettra de développer des techniques alternatives, mais elle favorisera aussi des agricultures plus intensives, donc plus polluantes,

ou l'exploitation de gisements miniers dangereux (on pense au charbon en Chine). L'approfondissement de la division verticale du travail conduit, par ailleurs, à multiplier les échanges, donc les transports polluants.

D'une manière plus générale, si la libéralisation des échanges a « ouvert » les frontières, d'autres évolutions ont contribué à les refermer. La crainte du terrorisme ou des épidémies a ainsi conduit à augmenter le coût de franchissement des frontières, dont les études empiriques récentes révélaient d'ailleurs l'importance (voir Anderson et Wincoop).

Conclusion

La mondialisation commerciale n'est pas un phénomène nouveau. Depuis le 18e siècle, les termes du débat ont pourtant peu évolué. Chaque période a marqué sa spécificité. Une phase de ce processus est en train de s'achever : elle aura vu la remise en cause des États-nations, l'approfondissement de la division « verticale » du travail agencée par les firmes multinationales, la montée de nouvelles puissances commerciales parmi les pays émergents.

Aujourd'hui, la réaffirmation des États, les réticences de l'opinion publique à l'égard de la mondialisation, la montée des craintes environnementales ou sécuritaires, le risque de pénurie dans certains produits de base, l'affirmation de nouveaux leaders, la fragilisation des organisations internationales laissent penser que les évolutions passées ne se poursuivront pas à l'identique.

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 10

Crises et cycles

BRUNO AMABLE

Professeur à l'université Paris 1 Panthéon-Sorbonne

Ce chapitre présente les principales théories des crises économiques et du cycle. On distingue le cycle des affaires, d'une périodicité de l'ordre de quelques années, du cycle long, relié aux innovations majeures. L'efficacité de la politique de stabilisation macroéconomique, inspirée de la théorie keynésienne et destinée à amoindrir les effets du cycle économique, a pu donner l'impression vers la fin des années 1960 que le cycle économique avait disparu. La crise des années 1970 a mis un terme à ces espoirs. La théorie française de la régulation a analysé les crises et les fluctuations économiques comme les conséquences endogènes du fonctionnement de l'économie. D'autres théories se sont tournées vers des conceptions prékeynésiennes du cycle : l'adaptation des économies aux chocs externes qu'elles subissent. Enfin, certaines approches lient politique et cycles. Ces derniers seraient la conséquence des modifications de politique économique effectuées sous l'influence des élections ou de l'alternance politique.

L'observation des séries statistiques sur longue période révèle un constat banal, celui de l'absence de développement uniforme de l'activité économique. C'est particulièrement le cas lorsqu'une économie connaît une période de rupture brutale dans la conjoncture économique, caractérisée par la montée du chômage, les faillites d'entre-

prises ou de banques ou encore la baisse brutale des prix, notamment pour certains actifs comme les titres financiers ou l'immobilier. On utilise le terme de *crise* pour désigner ces épisodes, terme emprunté au vocabulaire médical désignant un moment décisif dans l'évolution d'une maladie. Bernard Rosier, dans *Les Théories des crises économiques*, dénombre huit crises économiques pour le seul 19e siècle.

Le développement économique peut donc se concevoir comme une succession de phases de croissance rapide et de phases de repli ponctuées par des crises, ce qui conduit à parler de cycles, c'est-à-dire de périodes à l'issue desquelles les phénomènes se répètent dans le même ordre que précédemment. Comme l'écrit Clément Juglar en 1862, les « crises se renouvellent avec une telle constance, une telle régularité, qu'il faut bien en prendre son parti ».

Ce chapitre présente les principales conceptions du cycle développées depuis le 19e siècle. On peut distinguer deux types de cycles : le cycle des affaires, d'une périodicité de l'ordre de quelques années, et le cycle long. La crise des années 1930 a mis en avant l'intérêt des politiques de stabilisation pour contrer les effets du cycle, mais la crise des années 1970 a remis en question l'efficacité de ces politiques et favorisé un nouvel essor des conceptions classiques du cycle. L'article se conclut par la présentation de quelques contributions contemporaines à l'analyse du cycle.

LES FONDEMENTS DES THÉORIES DES CYCLES

Le cycle des affaires

La récurrence des crises économiques au cours du 19e siècle a conduit à s'interroger sur les causes et conséquences des fluctuations économiques. L'ouvrage de Clément Juglar paru en 1862, *Des crises commerciales et de leur retour périodique en France, en Angleterre et aux États-Unis*, est communément considéré comme la première contribution à la

théorie économique du cycle, bien que ce dernier mot ne soit pas employé par l'auteur lui-même. L'ouvrage avait pour but de répondre à une question posée par l'Académie des sciences morales et politiques : « Rechercher les causes et signaler les effets des crises commerciales survenues en Europe et dans l'Amérique du Nord au cours du 19e siècle. »

La réponse de Juglar tient dans une formule célèbre, « La seule cause de la dépression est la prospérité », qui contient l'idée que le « mouvement des affaires » est par nature cyclique. Pour Juglar, les crises peuvent être prévues, « adoucies », et il est possible de « s'en préserver jusqu'à un certain point », ce qui indique qu'une intervention de politique économique peut avoir une certaine utilité. Il observe toutefois qu'il n'a été donné à personne de supprimer les crises commerciales, ce qui suggère qu'il serait illusoire de chercher à abolir le mouvement cyclique de l'économie. Ce ne serait pas nécessairement désirable non plus car les crises posséderaient certaines vertus. Elles corrigeraient les « excès » de la période de prospérité en « rétablissant l'équilibre », faisant baisser des prix d'actifs surévalués, éliminant les entreprises les plus faibles dans la compétition économique, rendant par là possible l'émergence d'une nouvelle phase de prospérité. Ces différents thèmes, l'utilité d'une politique économique pour affaiblir le cycle, l'inévitabilité du caractère cyclique de l'économie et l'éventuel effet « épurateur » des récessions, vont se retrouver dans les débats économiques jusqu'à nos jours.

Le cycle identifié par Juglar, parfois appelé « cycle court » ou « cycle des affaires », a une périodicité moyenne de l'ordre de huit années. Il comporte trois phases : l'expansion (ou prospérité), la crise et la liquidation. Les auteurs ultérieurs considèrent communément une quatrième phase, la reprise.

L'expansion correspond à une élévation de l'ensemble des prix, l'augmentation de la production ainsi que l'accroissement des crédits et la hausse des revenus. C'est une période où le climat des affaires est à

l'optimisme. La crise est relativement brève, de quelques jours à quelques semaines, et correspond à une crise des financements : il devient difficile ou impossible de se voir accorder un crédit supplémentaire, ce qui contraint très strictement l'activité économique et entraîne un retournement des anticipations. À l'optimisme de la période de prospérité succèdent des anticipations plus pessimistes sur l'avenir. La dépression qui suit se traduit par une baisse des prix, des crédits et de l'activité. Les anticipations sur le climat des affaires, notamment celles portant sur la profitabilité des investissements, vont jouer un rôle dans les théories du cycle développées ultérieurement par l'économiste suédois K. Wicksell et encore plus tard chez J.M. Keynes.

Le dernier quart du 19e siècle et le début du 20e voient le développement des théories marginalistes et de l'équilibre général, avec les travaux de Menger, Jevons et Walras. Ces théories sont centrées sur la recherche d'un équilibre économique et semblent donc *a priori* peu compatibles avec la conception de l'économie comme cyclique et sujette à des crises. Toutefois, l'idée selon laquelle les crises de surproduction sont inévitables est largement répandue, tant il est vrai que la réapparition de ces crises est fréquente.

Dans la pensée de K. Marx, les crises sont inévitables en régime capitaliste en raison de la contradiction entre le caractère social de la production et l'appropriation privée, par les propriétaires des moyens de production, des produits du travail socialisé. L'« anarchie de la production » exclut tout développement harmonieux de l'économie. Les marchandises ne sont pas fabriquées pour satisfaire les besoins de ceux qui les ont produites, mais pour être vendues, contre de la monnaie. Il n'y a pas de coordination effective entre les décisions d'investissement et de production prises par les entreprises individuelles. C'est une fois les marchandises produites et mises sur le marché que l'on constatera si les décisions de produire et d'investir étaient fondées ou pas. Comme l'accumulation du capital conduit à la chute du taux de profit, les capitalistes tentent d'élargir la production pour augmenter la masse des profits. Cet accroissement de la production se heurtant à une

insuffisance de la demande du côté des travailleurs, les crises sont inévitables comme « solutions momentanées et violentes des contradictions existantes ».

Les mécanismes des crises de surproduction sont précisés par d'autres auteurs. Albert Aftalion étudie ainsi, dans deux articles de 1909 et 1911, les conséquences du décalage entre le temps nécessaire à la production des biens de production et celui nécessaire à la production des biens de consommation, ce dernier étant plus court. Quand la demande fluctue, les temps de réaction des deux secteurs diffèrent, ce qui explique en partie les crises de surproduction. La production de moyens de production ne peut être en permanence exactement proportionnée pour satisfaire les besoins des industries de consommation. La considération de l'aspect temporel de l'investissement et de la production se retrouvera plus tard dans la théorie du cycle de F. Hayek.

J. Kitchin propose en 1923 des cycles courts, de trois à quatre ans, fondés sur les variations de stocks. En période d'expansion, la constitution de stock renforce la croissance de la production nécessaire à la satisfaction d'une forte demande. Ce mécanisme s'inverse en période de dépression, le déstockage effectué par les entreprises venant accentuer le ralentissement de l'activité.

Le cycle long

Le cycle des affaires, d'une périodicité de quelques années, n'est pas le seul cycle pris en compte dans la théorie économique. La possibilité de cycles de périodicité plus longue a été envisagée dès le début du 20e siècle. Le concept d'un cycle de longue durée est attaché au nom de l'économiste russe Nicolaï Kondratiev, bien qu'on puisse trouver des précurseurs tels que l'économiste russe Parvus et les Hollandais van Gelderen et de Wolff.

Dans sa thèse de 1922 et un article de 1926, Kondratiev remarque que les prix de gros suivent des phases de croissance (phase A pour reprendre l'expression de F. Simiand) et de régression ou de croissance

ralentie (phase B) qui durent environ une vingtaine d'années chacune. Son travail statistique ne se limite pas à l'observation des séries de prix, il fait aussi appel à des indicateurs de taux d'intérêt, de volume de crédit, d'importations et d'exportations, ainsi qu'à des grandeurs réelles telles que la production de charbon ou de fer pour mieux caractériser les phases des cycles longs. Il identifie ainsi trois cycles : (i) 1790-1815 (phase A), 1815-1845 (phase B) ; (ii) 1845-1870 (phase A), 1870-1895 (phase B) ; (iii) 1895-1920 (phase A), 1920- (phase B). On peut noter qu'il identifie une phase déclinante dans les années 1920, décennie pourtant considérée par les contemporains comme une période de prospérité, jusqu'en 1929 tout au moins.

Pour Kondratiev, chaque cycle possède des caractéristiques historiques propres et ne constitue pas la répétition du cycle précédent. Il est toutefois selon lui possible de distinguer quatre caractéristiques communes :

- des changements importants dans les méthodes de production se produisent au début de la phase ascendante. Les inventions réalisées pendant la phase descendante ne sont exploitées qu'à la toute fin de cette dernière et au début de la phase ascendante ;
- la phase ascendante du cycle s'accompagne de bouleversements sociaux qui s'exacerbent à la fin de cette phase ;
- la phase descendante se caractérise par une longue dépression agricole ;
- le même processus dynamique lie cycles courts et cycles longs ; dans les phases ascendantes d'un cycle long, les phases d'expansion des cycles courts durent plus longtemps, et c'est l'inverse dans les phases de dépression.

La phase ascendante du cycle s'accompagne d'une hausse des investissements et de l'activité ainsi que d'une hausse des prix et des taux d'intérêt. Les opportunités d'investissement finissent par se raréfier, et les hausses de prix vont décourager l'expansion. Il s'ensuit alors un déclin de l'activité économique (phase B) qui provoque alors une

baisse des prix et des taux d'intérêt. Les opportunités d'investissement inexploitées pendant la phase B servent de socle à la nouvelle phase ascendante.

Comme on le voit, les inventions et innovations jouent un rôle dans la dynamique longue de l'économie. L'économiste autrichien Joseph Schumpeter va, dans sa théorie des cycles, systématiser le lien entre innovation et cycles longs en effectuant une distinction entre les innovations qu'on pourrait qualifier de « radicales » et celles qui sont plus « incrémentales ». Les premières sont des bouleversements conséquents des méthodes de production et peuvent donner naissance à de nouveaux secteurs d'activité et augmenter significativement la demande. Les innovations radicales peuvent déclencher une vague d'imitations comme des innovations de nature plus incrémentales, visant à améliorer ou à adapter à des contextes particuliers les innovations les plus importantes. Certaines innovations radicales peuvent impulser une phase ascendante d'un cycle long. L'épuisement de la phase ascendante correspondra à celui des opportunités d'investissement liées à ces innovations.

Il n'est pas nécessaire que ces innovations surviennent plus ou moins simultanément, « en grappes ». Il suffit que leur nature radicale soit suffisamment prononcée. Aucune innovation n'est susceptible à elle seule de déclencher une vague d'investissement suffisamment forte pour amorcer un cycle long d'une cinquantaine d'années, mais la nature radicale de l'innovation se reconnaît à l'ampleur de la vague d'innovations incrémentales qui va suivre et conduire à une augmentation significative de l'investissement et l'activité.

Schumpeter mentionne ainsi le chemin de fer, l'électrification et le moteur à explosion comme des innovations radicales qui ont donné naissance à des cycles longs. Chacune de ces innovations est évidemment susceptible de connaître des améliorations très importantes ; il n'est qu'à penser aux nombreuses différences existant entre la première locomotive et les actuels trains à grande vitesse pour en être persuadé.

Mais pour être à l'origine d'un cycle long, ces innovations doivent surtout être capables d'exercer de puissants effets d'entraînement sur d'autres secteurs en suscitant des innovations complémentaires dont le développement suscitera une vague d'investissement. Comme le dit Schumpeter, le véhicule automobile n'aurait jamais pris l'importance qu'il a eue et n'aurait jamais autant bouleversé l'économie et la vie quotidienne s'il n'avait pas progressé du point de vue technique depuis son invention et s'il n'avait pas réussi à modifier son environnement, notamment en termes d'infrastructures routières et de réseau de stations-service.

Il faut noter que Schumpeter n'est guère explicite sur les mécanismes qui peuvent relier le développement d'innovations radicales à celui des innovations incrémentales. Il mentionne le terme de « révolution industrielle » à propos des premier et troisième cycles de Kondratiev sans développer le concept beaucoup plus avant. Il reviendra, dans les années 1980, aux auteurs « post-schumpetériens » (C. Freeman, G. Mensch, C. Perez, G. Dosi…) la tâche de proposer une théorie plus explicite des liens entre innovations et diffusion dans le cadre d'une théorie des cycles longs. L'accent sera mis sur les interrelations entre les techniques donnant lieu à de véritables systèmes ou « paradigmes » technologiques. Le cycle long sera alors le résultat de l'apparition d'innovations liées entre elles pour des raisons techniques ou organisationnelles, plutôt que sur les innovations isolées en elles-mêmes.

Schumpeter quant à lui était plus préoccupé par le côté financier des cycles que par leur aspect technologique ; il était avant tout soucieux de développer une théorie de l'innovateur sur la base de laquelle il devait avancer que la bureaucratisation de l'innovation devrait conduire à la mort de l'entrepreneuriat et même du capitalisme, tout en reconnaissant à la firme monopoliste un avantage comparatif dans l'innovation.

LA CRISE DE 1929 ET LA THÉORIE DE KEYNES

La crise des années 1930, caractérisée par son ampleur et la gravité de ses conséquences, modifie l'appréhension des crises économiques et des politiques à mettre en œuvre. Pour les économistes d'inspiration marxiste, il s'agit plus que d'une banale crise de surproduction ; c'est une manifestation des contradictions entre le développement des forces productives et les rapports sociaux. Les théoriciens des cycles longs y voient la manifestation brutale d'une conjonction de phase de dépression d'un cycle court et de phase descendante d'un cycle long. Les théoriciens libéraux l'interprètent comme une crise classique dont la gravité ne provient que des mécanismes qui empêcheraient le libre ajustement de l'offre et de la demande : la rigidité des salaires à la baisse en particulier expliquerait la persistance d'un chômage élevé. Une politique d'équilibre budgétaire et le libre jeu des mécanismes de l'offre et de la demande devraient suffire à faire retourner l'économie vers un équilibre de plein-emploi.

Dans ce contexte, Keynes avance qu'il convient d'abandonner l'idée que les équilibres économiques se rétablissent en laissant fonctionner le mécanisme de l'offre et de la demande. La baisse du niveau de salaire ne peut être la réponse à la crise. Pour lui, l'économie peut rester durablement à un équilibre de sous-emploi, c'est-à-dire sans diminution du niveau de chômage, même si les mécanismes de l'offre et de la demande jouent librement.

Pour assurer le plein-emploi, il convient d'utiliser les moyens offerts par la politique économique. L'État peut financer des grands travaux au prix d'un déficit budgétaire. Par ses commandes, l'État augmente l'investissement, ce qui se traduit par une augmentation de la demande effective, par un effet multiplicateur de l'investissement. Cet effet multiplicateur sera d'autant plus élevé que la propension à épargner des ménages est faible.

Il existe donc un rôle pour la politique économique dans la détermination du niveau d'activité. La politique budgétaire permet d'augmenter

le niveau de la demande en augmentant les dépenses de l'État ou les revenus disponibles des agents privés (ménages ou entreprises) par une baisse des impôts ou une hausse des prestations sociales. La politique monétaire consiste à modifier l'offre de monnaie en diminuant ou relevant le niveau des taux d'intérêt. Ces changements affectent la demande de crédits par les ménages et les entreprises, et conduisent à des modifications de la demande des ménages et de l'investissement des entreprises.

L'État peut aussi agir de façon plus directe en accroissant les revenus d'une fraction de la population. C'est par exemple ce qui se produit quand est décidée une augmentation du salaire minimum ou des prestations sociales. Une politique ciblée sur les populations les plus pauvres a, dans l'optique keynésienne, une efficacité *a priori* plus grande. Comme ces populations consacrent une plus grande partie de leurs revenus à la consommation que les ménages riches, tout supplément de revenu dont elles bénéficient conduit à un accroissement proportionnellement plus grand de la demande finale et par conséquent à une efficacité plus grande de l'effet multiplicateur mentionné précédemment.

Le gouvernement dispose donc d'outils de politique économique qui devraient lui permettre d'atténuer les effets les plus déplaisants du cycle économique, notamment en termes de chômage ou de faillites, en stimulant ou en ralentissant l'activité selon la phase du cycle dans laquelle se trouve l'économie. En phase de récession, où le chômage augmente et l'activité baisse, il conviendra d'augmenter les dépenses publiques, de diminuer les impôts ou le niveau des taux d'intérêt pour atténuer la récession ou stimuler la reprise de l'activité. En phase d'expansion et en situation de plein-emploi, on peut s'attendre à ce que le niveau des prix augmente ; une diminution du déficit budgétaire ou une hausse des taux d'intérêt ralentira l'économie et freinera l'inflation.

L'impact de la théorie keynésienne sur la conduite de la politique économique fut tout à fait considérable sur toute la période de l'après-guerre. La réinterprétation des idées de Keynes par des économistes

plus formalisateurs tels que R. Harrod, J. Meade, J. Hicks, F. Modigliani, P. Samuelson ou R. Solow, a conduit à la définition d'un modèle macroéconomique effectuant une synthèse entre la théorie keynésienne et la théorie « classique » à laquelle Keynes adressait ses critiques dans son ouvrage de 1936, *La Théorie générale de l'emploi, de l'intérêt et de la monnaie*. La théorie classique était jugée valable dans le long terme, les ajustements de prix permettant un retour à l'équilibre de plein-emploi. En revanche, dans le court terme, la théorie keynésienne était considérée plus pertinente, l'élimination du sous-emploi revenant à la politique économique et non au jeu de l'offre et de la demande.

Cette conception de l'économie effectue donc une séparation nette entre le cycle et la croissance. Le cycle est conçu comme un phénomène de court terme susceptible d'être influencé par la politique économique. La croissance est en revanche conceptualisée comme un phénomène de long terme, sur lequel la politique budgétaire ou monétaire n'a aucune influence.

La politique d'inspiration keynésienne consistant à intervenir pour contrer les effets du cycle économique a connu un relatif succès, au moins jusque dans les années 1970. L'observation de la série du produit intérieur brut (PIB) des États-Unis sur la période de l'après-guerre, par exemple, révèle des cycles beaucoup moins accentués que ceux que l'on peut observer sur la période précédant la Seconde Guerre mondiale. La « volatilité » du PIB a ainsi diminué au moins en partie grâce à la conduite de politiques destinées à contrer le cycle et ses effets. La confiance dans la possibilité pour la politique économique d'affaiblir très significativement le comportement cyclique de l'économie se résume dans le titre d'un ouvrage de M. Bronfenbrenner paru en 1969 : *Le cycle des affaires est-il devenu obsolète ?* Cette confiance a toutefois été sérieusement ébranlée par la crise des années 1970.

LA CRISE DES ANNÉES 1970 ET LA THÉORIE DE LA RÉGULATION

La crise des années 1970 et l'impossibilité d'y apporter une solution avec les recommandations habituelles de politiques budgétaire et monétaire ont conduit à redonner au cycle et à la crise une place centrale dans l'analyse macroéconomique. Certaines théories ont ainsi pris un point de départ résolument prékeynésien, en concevant le cycle comme résultant de chocs exogènes sur un système intrinsèquement en équilibre. Ces théories sont abordées dans la prochaine section.

D'autres approches ont revisité les théories des cycles longs et interprété la crise des années 1970 comme l'amorce d'une phase B d'un cycle de Kondratiev. La sortie de la crise se ferait alors lorsqu'une ou plusieurs innovations radicales pourraient initier un retournement de phase. La recherche de telles innovations s'est faite en direction des technologies de l'information et de la communication, tenues pour capables de constituer un nouveau « paradigme technologique » par les auteurs « post-schumpetériens » tels que G. Dosi. Les auteurs contemporains préfèrent utiliser le terme d'« onde » longue plutôt que celui de cycle, qui suppose une trop grande régularité dans les fluctuations.

La théorie de la régulation

Dans ce contexte, la théorie française de la régulation, à laquelle sont associés les noms de Michel Aglietta ou de Robert Boyer, se distingue par une interprétation originale des périodes de croissance et de crise. Plutôt que de penser la crise comme une perturbation survenue dans une économie qui devrait être stable, ils retournent la question en se demandant comment il est possible d'expliquer la période de forte croissance stable qui a précédé la crise. La crise n'est alors pas conceptualisée comme une anomalie, comme la conséquence d'erreurs de politique économique ou comme le résultat de l'imperfection des marchés qui ne peuvent s'ajuster librement, mais comme l'expression de la dynamique d'un *mode de régulation* et d'un *régime d'accumulation*.

Plus précisément, la théorie de la régulation s'appuie sur l'analyse de cinq *formes institutionnelles* :

- les formes de la concurrence ;
- la monnaie ;
- l'État ;
- le rapport salarial ;
- l'insertion dans l'économie mondiale.

Dans une perspective historique, la théorie de la régulation caractérise le capitalisme comme une succession de *modes de régulation* définis comme des ensembles de procédures et comportements permettant de reproduire les rapports sociaux fondamentaux à travers la conjonction de formes institutionnelles historiquement déterminées. Le *régime d'accumulation* est défini comme l'ensemble des régularités qui assurent une progression générale et cohérente de l'accumulation du capital. Le mode de régulation « soutient » et « pilote » le régime d'accumulation. Il est alors distingué différents modes de régulation :

- le mode de régulation à l'ancienne, jusqu'à la fin du 18e siècle, où la dynamique économique est impulsée par les aléas qui affectent l'agriculture ;
- le mode de régulation concurrentielle du 19e siècle, où la dynamique économique est impulsée par l'industrie manufacturière ; les ajustements sont fondés sur les mécanismes de marché et la concurrence ;
- un mode de régulation de transition dans l'entre-deux-guerres ;
- un mode de régulation fordiste pour la période de l'après-Seconde Guerre mondiale, où la dynamique économique repose sur une distribution des gains de productivité aux salariés permettant d'alimenter la croissance de la demande et l'investissement ;
- on peut ajouter l'existence d'un mode de régulation financiarisée caractérisant la période contemporaine.

Dans cette conception du capitalisme, les évolutions cycliques de l'économie résultent des caractéristiques des formes institutionnelles, en particulier la concurrence et le rapport salarial. La croissance stable sera le résultat de la cohérence d'un ensemble de formes institutionnelles. De même que, pour l'historien français Ernest Labrousse, toute société possède la conjoncture de sa structure, tout mode de régulation aura les crises de sa structure.

Ceci conduit la théorie de la régulation à distinguer cinq types de crises selon Robert Boyer.

Les cinq types de crise

Le premier type est la crise comme perturbation externe. Les difficultés économiques sont alors liées à des pénuries, d'origine climatique par exemple, ou aux conséquences de perturbations extérieures telles que les guerres. Ce type de crise correspond en partie à la conception qu'on retrouve chez certains auteurs inspirés des classiques : des chocs externes sur le système économique viennent perturber sa dynamique. Ces crises ne remettent pas en cause le mode de régulation ou le régime d'accumulation.

Le deuxième type est en revanche une crise *endogène*. C'est le fonctionnement même de l'économie qui donne naissance à un comportement cyclique, non les chocs provenant de l'extérieur du système. Les crises ne sont alors que l'expression du mode de régulation en vigueur, et on peut leur associer le même type de propriétés que celles qu'observaient déjà Juglar : elles résorbent les déséquilibres accumulés pendant les phases d'expansion, notamment en matière d'investissement excessif. Ces crises sont ainsi partie intégrante du fonctionnement de l'économie et surviennent donc sans entraîner de modifications profondes ou rapides du mode de régulation.

En revanche, le troisième type est une crise du mode de régulation lui-même. Les mécanismes associés à celui-ci ne permettent plus de rétablir une conjoncture favorable. Il s'agit de crise dont l'économie

ne peut sortir sans une altération profonde des formes institutionnelles et de leur articulation, mais le régime d'accumulation demeure viable. C'est ainsi que la théorie de la régulation interprète la période allant de 1929 à 1932 aux États-Unis. La chute de l'activité ne permet pas une augmentation des profits mais accélère au contraire leur chute. Le mode de régulation concurrentielle ne permet pas le retour à une phase de croissance stable et régulière mais entraîne au contraire l'économie dans une spirale dépressionniste.

Le quatrième type est la crise du régime d'accumulation. Cette crise est l'arrivée aux limites et l'exacerbation des contradictions au sein des formes institutionnelles les plus essentielles conditionnant le régime d'accumulation. C'est ce qui se produit avec la crise du fordisme dans les années 1970 ou avec la grande crise des années 1930.

À terme, ces crises mènent à la crise du mode de développement de l'économie, qui constitue le cinquième type de crise. La crise ne conduit alors plus seulement à la remise en cause des formes institutionnelles mais à celle des rapports sociaux fondamentaux dont elles sont l'expression. L'effondrement des économies planifiées telles que l'Union soviétique est un exemple d'une telle crise.

QUELQUES APPROCHES CONTEMPORAINES

Les théories du cycle réel

La crise des années 1970 a permis le retour en vogue de conceptions prékeynésiennes du cycle à la faveur de ce qui était perçu comme un échec des politiques de stabilisation et de relance, échec supposé traduire les erreurs fondamentales de la théorie keynésienne. À la suite des travaux de R. Lucas puis de ceux de F. Kydland et E. Prescott, s'est développé un courant dit des « cycles réels ».

L'idée fondamentale est que les cycles ne sont pas des dysfonctionnements de l'économie, mais la conséquence des décisions rationnelles des agents économiques face à des chocs que subit l'économie conçue comme un système intrinsèquement stable et autorégulé. Ces chocs ont d'abord été conçus comme d'origine monétaire, par exemple des modifications des politiques monétaires mal anticipées par les agents économiques. Il leur a été ensuite préféré des chocs d'origine *réelle* comme des modifications dans les préférences des agents, ce qui a des conséquences sur le temps que les agents souhaitent consacrer au travail plutôt qu'au temps libre ; des chocs de productivité, qui modifient l'efficacité du processus de production et vont donc avoir des conséquences sur la décision de produire et d'investir ; ou tout autre choc non monétaire.

Pour être capable de rendre compte des fluctuations que connaissent les économies, les chocs doivent avoir des effets suffisamment persistants, faute de quoi l'économie perturbée reviendrait rapidement à l'équilibre. C'est pourquoi les modèles de cycles réels font appel à des mécanismes se fondant sur le comportement intertemporel des agents économiques dans leur choix de travailler ou de profiter du temps libre. Si un choc réel amène les agents à modifier leur offre de travail, en diminuant par exemple le temps qu'ils consacrent au loisir, il est supposé que l'offre future de travail dépendra du choix présent que font les agents quant au temps qu'ils souhaitent consacrer au loisir. Une perturbation donnée prolongera donc ses effets dans le temps. Mais la « solution » la plus radicale, et la plus utilisée, est encore de supposer que les chocs sont eux-mêmes persistants, ce qui garantit que l'économie qui les subit aura elle-même un comportement persistant.

Dans cette perspective théorique, les cycles économiques ne sont pas une inefficience à laquelle la politique économique devrait remédier. Le comportement cyclique n'est que la conséquence « naturelle » de l'adaptation de l'économie aux chocs qu'elle subit. Il en résulte que la politique de stabilisation de la conjoncture préconisée par les keyné-

siens n'a plus lieu d'être. Pire, en empêchant les marchés de s'équilibrer « naturellement », elle risque d'imprimer à l'économie des fluctuations qui n'auraient pas lieu d'être et qui seraient, elles, inefficientes.

L'absence de rôle donné aux perturbations monétaires par la théorie du cycle réel a suscité des interrogations, compte tenu des observations empiriques. Il semblerait bien que les modifications de la politique monétaire aient des effets *réels*, sur le niveau de production par exemple, et pas seulement des effets monétaires, sur le niveau des prix. Il est possible de prendre en compte de tels effets si on suppose l'existence d'« imperfections » diverses sur les marchés, qui empêchent les ajustements de prix à leur niveau « naturel », par exemple une réglementation sur les salaires ou l'impossibilité de baisser les prix. En combinant l'appareillage théorique de la théorie des cycles réels avec de telles imperfections, il est possible d'obtenir une théorie où les chocs réels et monétaires (et donc la politique monétaire) jouent un rôle dans le comportement cyclique de l'économie.

Le cycle politico-économique

Certaines théories font, elles, jouer à la politique économique et plus profondément à la politique tout court un rôle de premier plan dans les évolutions cycliques de l'économie. Il s'agit des théories dites du « cycle politico-économique ». L'expression elle-même remonte à M. Kalecki dans son article de 1943, « Political aspects of full employment ». Pour Kalecki, la politique du gouvernement est déterminée par la pression de groupes sociaux porteurs d'intérêts conflictuels, le patronat, les rentiers et les salariés. Selon la conjoncture macroéconomique, des alliances diverses entre ces groupes sont susceptibles de se former pour exercer une pression sur le gouvernement pour lui faire adopter telle ou telle politique.

Dans les phases de récession, la pression populaire incite le gouvernement à des mesures de politique expansionniste pour lutter contre le chômage de masse. Le patronat peut lui-même soutenir ces revendications car cela améliorera les débouchés des entreprises.

En revanche, en période d'expansion, le maintien d'une politique de relance risque de conduire à l'inflation. Les intérêts du patronat et des rentiers sont alors de faire pression pour que le gouvernement mène une politique « orthodoxe » d'équilibre des comptes et de réduction des dépenses publiques car l'inflation détériore la profitabilité de leurs investissements.

La première formalisation d'un cycle politico-économique est due à Nordhaus en 1975. Les éléments pris en compte sont différents de ceux de Kalecki. Le gouvernement ne réagit pas à la pression de groupes sociaux aux intérêts divergents, mais à une fonction de vote qui relie le soutien dont bénéficie le gouvernement à la situation économique d'ensemble, en particulier le taux de chômage et le taux d'inflation. Le gouvernement trouve son intérêt dans le soutien électoral, et son comportement engendre un cycle que Nordhaus définit comme « opportuniste ».

Dans la période qui précède immédiatement une élection, le gouvernement stimule l'économie par une politique monétaire expansionniste. Le chômage diminue en raison de l'absence d'anticipation de cette politique par les agents économiques. Si les agents avaient anticipé le changement de politique monétaire, l'effet de cette politique se serait retrouvé dans le niveau des prix, sans conséquence importante sur le niveau d'activité ou le chômage. Le gouvernement obtient donc un faible niveau de chômage sans augmentation significative du niveau des prix, ce qui favorise sa réélection.

Dans la période après l'élection, l'inflation augmente, conséquence de la politique monétaire menée avant l'élection. Le gouvernement modifie alors sa politique et met en œuvre une récession *via* une politique monétaire restrictive pour ramener à la baisse les anticipations d'inflation. Le gouvernement garde l'activité économique faible pour brider les anticipations d'inflation jusqu'à la période qui précède immédiatement l'élection, de façon à ce qu'une expansion économique puisse être obtenue à un relativement faible coût en termes d'inflation.

Le cycle économique, alternance de périodes d'expansion et de périodes de récession, se trouve alors rythmé par le calendrier électoral.

Dans le modèle de Nordhaus, les politiciens sont « opportunistes » au sens où ils ne se préoccupent que de leur réélection et n'ont pas de préférence particulière vis-à-vis de la politique qu'ils sont conduits à mener. D'autres approches prennent en compte les préférences partisanes des politiciens, qui désirent alors, une fois au pouvoir, mener des politiques proches de celles que souhaite leur base sociale. Le cycle politico-économique résulte alors de l'alternance politique. Pour D. Hibbs, les partis de gauche mènent des politiques de lutte contre le chômage tandis que les partis de droite mènent des politiques de lutte contre l'inflation. Le cycle économique sera alors une alternance de phases expansionnistes et inflationnistes correspondant à des gouvernements de gauche, et de phases récessionnistes correspondant à des gouvernements de droite.

Chapitre 11

Qu'est-ce que l'économie ?

BERNARD WALLISER

Professeur à l'ENPC, directeur d'études à l'EHESS

L'économie a longtemps admis une ontologie fondée sur le modèle rationnel de décision et la notion d'équilibre entre actions individuelles. Cette vision se décline à un niveau de spécification élevé dans la théorie des jeux et à un niveau plus spécifique dans des modèles d'échanges de biens. Ces deux piliers tendent à s'affaiblir, permettant l'intégration d'une plus grande variété de concepts et une articulation avec les autres sciences sociales.

L'économie a longtemps admis une épistémologie fondée sur l'usage de modèles formels qui ne sont qu'indirectement confrontés aux données historiques. Ces modèles sont idéaux et sont destinés à mettre en scène les conditions possibles d'un phénomène plus que les conditions effectives de ces phénomènes. Cependant, le développement de l'économie expérimentale conduit à des modèles plus proches des faits et plus simples à instrumentaliser.

L'économie est souvent décrite comme la plus molle des sciences dures ou la plus dure des sciences molles. Elle partage avec les autres sciences sociales une ontologie par laquelle des acteurs agissent intentionnellement pour donner naissance à des phénomènes collectifs. Elle partage avec les sciences dures une épistémologie fondée sur des modèles formalisés qui permettent une simulation aisée des phénomènes. Elle a ainsi longtemps cultivé des caractéristiques spécifiques qui tendent peu à peu à s'estomper.

ONTOLOGIE DE L'ÉCONOMIE

L'ontologie des disciplines scientifiques est fondée sur des principes généraux, qui s'appliquent parfaitement à l'économie. L'univers est structuré en niveaux emboîtés d'organisation, chacun caractérisé par des entités spécifiques en interaction dans un environnement commun. Le fonctionnement des entités d'un niveau (qualifié de macroscopique) est expliqué par les entités de niveau inférieur (qualifié de microscopique). Les différences entre économie et autres sciences sociales reposent alors sur la mise en œuvre de schémas d'explication spécifiques.

Niveaux d'organisation

La discipline économique distingue trois niveaux d'organisation essentiels, chacun caractérisé par des entités propres. Le niveau intermédiaire, le « niveau actorial », est formé par des acteurs économiques qui ont des comportements autonomes. En fait, l'économie a pris l'habitude de considérer comme des acteurs autonomes non seulement des individus, mais des entités collectives (entreprises, syndicats). Le niveau supérieur est le « niveau social », qui est le siège de phénomènes socio-économiques divers (distribution des revenus, relations privilégiées, prix, phénomènes de croissance). Il comprend en particulier des institutions, conçues comme des entités favorisant la coordination entre les acteurs (normes, monnaie, marchés). Le niveau inférieur est le « niveau mental », formé des états mentaux attribués aux acteurs (individuels ou collectifs). Il s'agit plus spécifiquement des croyances et des préférences des acteurs (et de leurs opportunités). Plus récemment, un niveau encore inférieur a été considéré, le « niveau neuronal », formé des neurones du cerveau.

Chaque niveau est décrit par les propriétés attribuées aux entités, par des liaisons entre ces entités et par l'environnement global. Les liaisons peuvent définir un réseau préalable entre les acteurs ou être tissées par

les acteurs eux-mêmes. L'environnement est lui-même formé d'un contexte matériel et des éventuelles entités du niveau supérieur. Au niveau neuronal, les neurones ont un comportement très simple et communiquent dans des processus beaucoup plus complexes. Au niveau mental, les états mentaux de base sont souvent traités comme relativement indépendants entre eux, mais sont fondamentalement conditionnés par le contexte. Au niveau actorial, les acteurs posent des actions autonomes ou interagissent avec leurs congénères par des interactions variées, dans un contexte matériel et institutionnel donné. Au niveau social, les phénomènes socio-économiques se conditionnent les uns les autres, dans un contexte socio-juridique commun.

Le cadre ontologique est très bien décrit par la théorie des jeux, qui décrit les rapports stratégiques entre acteurs à un niveau élevé de généralité. Annexée par l'économie, elle est définie comme la discipline qui étudie les choix interdépendants entre des acteurs autonomes. Elle considère des joueurs qui élaborent des actions de type quelconque dans un face-à-face sans médiation institutionnelle. Les joueurs ont des préférences qui dépendent tant de leurs actions que de celles des autres, dans la mesure où elles s'exercent sur des conséquences communes. Ils ont des opportunités indépendantes qui délimitent le domaine des actions possibles, ces actions pouvant se dérouler dans le temps. Ils ont enfin des croyances qui portent sur leur environnement commun, mais aussi sur les actions et les états mentaux des autres, ce qui engendre des croyances croisées. Finalement, les joueurs peuvent jouer un jeu (statique ou dynamique) unique ou un jeu répété sur un horizon fini ou infini.

Des propriétés plus spécifiques sont introduites dans un contexte particulier, ainsi des échanges de biens. L'économie apparaît alors sous sa forme traditionnelle d'étude des phénomènes de production, d'échange et de consommation de biens. Les biens sont simplement des entités suffisamment divisibles et manipulables pour faire l'objet de transactions indépendantes. Les agents, caractérisés par leurs rôles dans le processus économique (producteurs, consommateurs), expriment des offres et des demandes de biens en fonction de signaux particuliers

que sont les prix. Les préférences des agents dépendent des prix et des biens produits ou consommés, et les croyances des agents portent sur les protocoles d'échanges (et leurs intervenants). Les institutions sont des procédures d'échanges (marchés multilatéraux, procédures d'enchères), encadrées par des institutions plus générales (droits de propriété, confiance).

Types d'explication

Certaines explications proposées en économie relient exclusivement les propriétés des entités d'un niveau d'organisation donné. Ainsi, au niveau social, les grandeurs macroéconomiques sont supposées se déterminer directement les unes les autres. Cependant, en règle générale, les entités d'un niveau sont expliquées par celles du niveau inférieur, avec un retour possible du niveau supérieur sur le niveau inférieur. Au niveau social, les phénomènes collectifs (y compris les institutions) sont expliqués par des interactions individuelles entre acteurs placés dans un environnement institutionnel. Mais ces effets collectifs peuvent agir en retour sur le comportement des agents ou leurs états mentaux (conditionnement social). Au niveau actorial, les actions des agents sont expliquées par leurs états mentaux, dans un contexte matériel. Là encore, les actions des agents peuvent venir influencer en retour leurs états mentaux (préférences endogènes). Au niveau mental, les états mentaux sont supposés être expliqués par les opérations neuronales, l'existence d'une rétroaction restant un problème ouvert.

Dans le passage du niveau mental au niveau actorial, les explications sont généralement de nature intentionnelle. C'est dire que les actions résultent d'un processus de délibération transformant certains états mentaux en d'autres. Le schéma rationnel de décision affirme que les acteurs combinent leurs croyances et leurs préférences dans un processus de délibération mentale pour aboutir à des intentions d'action (ensuite traduites en action). En fait, deux formes distinctes de rationa-

lité sont souvent distinguées. La rationalité instrumentale traduit l'adéquation entre moyens disponibles et objectifs poursuivis. La rationalité cognitive traduit l'adéquation entre informations disponibles et croyances adoptées. La rationalité instrumentale forte est traduite par un schéma optimisateur pour lequel l'acteur maximise une utilité (symbolisant ses préférences) sous contraintes (symbolisant ses opportunités), pour des croyances données. Mais une rationalité limitée, engendrée par les capacités limitées de traitement d'information, est de plus en plus souvent considérée.

Dans le passage du niveau individuel au niveau social, les explications sont généralement de type causal. C'est dire que les phénomènes sociaux sont la résultante mécanique des actions individuelles posées par les acteurs. La notion d'équilibre joue un rôle primordial et traduit simplement l'idée que les acteurs sont coordonnés entre eux. Dans un jeu, l'équilibre traduit l'idée que les actions des joueurs sont coordonnées les unes aux autres, sans que les processus qui mènent à l'équilibre soient forcément décrits. Dans une économie d'échanges, l'offre et la demande de biens sont équilibrées par un « commissaire-priseur », qui peut ou non avoir une traduction concrète (Bourse, criée). Toutefois, l'idée d'équilibre tend à se dissoudre dans un courant de pensée contemporain, l'« économie évolutionniste » (voir encadré). Ayant une rationalité limitée, les acteurs se contentent de réagir à leurs observations passées et s'engagent dans un processus d'apprentissage conjoint. On obtient un processus global qui connaît des états transitoires et des états asymptotiques, ces derniers pouvant être comparés à des équilibres.

Analogies physiques et biologiques

L'économie s'est inspirée, tout au long de son développement, d'analogies puisées dans les sciences de la nature mais qu'elle a dûment domestiquées. Ainsi, le concept d'équilibre est issu de la mécanique et traduit l'existence d'une situation stable d'un système en l'absence de perturbations extérieures. De même, le concept de division du travail

est issu de la biologie et traduit l'apparition d'une différenciation fonctionnelle entre diverses activités au sein d'un système. Plus récemment, l'économie a importé des modèles analogiques, qui sont d'abord transférés dans leur forme, mais conservent aussi une partie de leur interprétation. Ainsi, le modèle d'Ising, issu de la physique des milieux désordonnés, rend compte de l'équilibre entre les actions d'un grand nombre d'agents similaires, soumis à une influence sociale. De même, le modèle de Darwin, issu de la biologie évolutionnaire, introduit les mécanismes de mutation et de sélection, appliqués aux agents eux-mêmes ou à certaines de leurs caractéristiques (actions, croyances).

Les explications sont d'un type plus incertain quand il s'agit de rendre compte de la transformation entre états physiques et psychiques. Dans le sens direct, il en est ainsi dans le passage des états neuronaux aux états mentaux, où des opérations du cerveau donnent naissance à des manifestations mentales plus ou moins conscientes. Dans le sens inverse, il en est ainsi dans le passage du niveau mental au niveau actorial, où les intentions d'action des acteurs sont transformées en actions effectives. Par ailleurs, des explications plus précises que des explications causales ou intentionnelles restent parfois mobilisées en économie, même si elles tendent à être réduites. Au niveau social, une explication fonctionnaliste rend compte de certaines institutions par les fonctions qu'elles remplissent dans le système social global. De même, au niveau actorial, une explication behaviouriste rend compte des actions comme le simple effet de *stimuli* physiques.

Économie et autres sciences sociales

L'économie se distingue peu des autres sciences sociales (psychologie, sociologie, politologie) quant à ses caractéristiques ontologiques générales. Les mêmes niveaux d'organisation sont invoqués dans ces sciences, même si le niveau mental y est moins nettement singularisé. Les entités de base (acteurs, liaisons, institutions) sont structurellement identiques, même si les institutions y jouent un plus grand rôle. En revanche, les

phénomènes particuliers auxquels s'intéresse l'économie au regard des autres sciences sociales sont différents. Les acteurs sont plus diversifiés dans ces sciences, en fonction de caractéristiques variées (âge, sexe, catégorie socioprofessionnelle, appartenances diverses). Les liaisons entre acteurs sont de nature plus variée, qu'elles soient organiques (liens de parenté) ou symboliques (statut social). Les institutions (règles de mariage, conventions de langage) y sont omniprésentes et souvent déterminantes.

Le mode d'explication global des phénomènes sociaux, à savoir le mode intentionnel, est également le même qu'en économie. Cependant, les acteurs sont supposés avoir une rationalité moins forte dans ces sciences que celle postulée en économie. Non seulement ils sont limités dans leurs capacités de traitement de l'information, mais ils subissent des contraintes sociales très fortes. De même, la notion d'équilibre n'est quasiment pas mobilisée, même si elle est parfois implicite (oligopole social). L'articulation entre les actions des acteurs s'exerce à travers divers modes de régulation toujours imparfaits, mais essentiellement dynamiques. Enfin, la rétroaction du social est mise en avant plus vigoureusement, en mettant en cause le postulat d'individualisme méthodologique adopté en économie (le social est exclusivement déterminé par les actions individuelles). L'acteur est fortement conditionné par le social et il est même considéré comme une entité directement sociale, non pensable en isolation.

De fait, l'économie tend à étendre son champ vers les autres sciences sociales, d'abord à un niveau spécifique, en étendant la notion de bien à des entités de plus en plus hétérogènes. Initialement, elle ne considérait comme biens que des entités physiques divisibles et échangeables comme des casseroles ou des automobiles. Elle a ensuite étendu cette notion à des services moins facilement divisibles comme la coiffure ou la prostitution. Elle a fini par traiter de la même façon des biens de nature plus complexe comme le travail ou la monnaie. Elle a enfin investi des activités certes transactionnelles, mais sans réel échange matériel, comme le crime, le mariage ou le culte. Ce faisant, elle empiétait peu à peu sur le terrain traditionnel d'autres sciences sociales. Une évo-

lution parallèle a porté sur la notion de capital, appliquée d'abord au capital matériel (équipements), puis immatériel (savoir-faire) et enfin relationnel (réseau de connaissances) ou symbolique (réputation).

L'économie tend également à étendre son champ en se plaçant d'emblée à un niveau de généralité élevé. Elle fait de la théorie des jeux une sorte de langage général des sciences sociales qui ne demande qu'à être spécifié dans tel ou tel contexte. Déjà, le système économique apparaît comme un gigantesque jeu entre des producteurs et des consommateurs. Mais la théorie des jeux est aussi appliquée en sciences politiques pour traduire les rapports politiques entre des États ou les rapports d'hostilité entre des belligérants. Elle est appliquée en sociologie pour traduire les rapports entre groupes sociaux aux intérêts plus ou moins convergents. Elle est appliquée en psychologie pour traduire les rapports entre les « moi » multiples d'un même individu. De fait, la théorie des jeux continue à se développer de façon à pouvoir prendre en compte un maximum de phénomènes sociaux, en particulier en intégrant l'incertitude et la dynamique.

ÉPISTÉMOLOGIE DE L'ÉCONOMIE

L'épistémologie des sciences repose là encore sur certains principes généraux parfaitement applicables à l'économie. Trois types d'objets conceptuels maniés par les scientifiques sont traditionnellement distingués : les données, les modèles et les théories. Ils sont articulés les uns aux autres selon des démarches diverses, en particulier des méthodes statistiques d'ajustement des modèles aux données. Ces objets conceptuels ne sont pas limités à la description de systèmes réels, mais ont des usages bien plus diversifiés.

Niveaux de description

L'économie met en œuvre trois types essentiels d'objets conceptuels, conçus comme des « quanta » de connaissance autonomes. Au centre

figurent les modèles, qui se présentent comme des ensembles cohérents d'hypothèses, desquelles on peut dériver des conséquences. Ils peuvent être ordonnés en familles cohérentes, un modèle de référence admettant diverses variantes. À la base figurent des données qui ont été recueillies par des protocoles d'observation précis (recueil systématique, enquêtes). Recueillies par les agents eux-mêmes et traitées par des organismes spécialisés, elles sont structurées en bases de données. Au niveau supérieur, des théories expriment des principes généraux dûment articulés entre eux. Peu nombreuses, elles ont pour rôle d'exprimer l'ontologie retenue du système économique. Plus haut encore, on considère parfois les programmes de recherches, qui énoncent les principes doctrinaux d'une école de pensée. Là encore, il existe un programme de recherches dominant, le programme néoclassique, et des hétérodoxies moins bien conceptualisées.

Depuis un demi-siècle, l'économie exprime ses objets conceptuels dans un langage formel. Les données sont énoncées sur des échelles de mesure en général cardinales, mais ne sont guère accompagnées d'une estimation des erreurs de mesure. Ainsi, les quantités de biens comme les prix apparaissent comme des grandeurs directement quantitatives. Les modèles sont énoncés dans un cadre analytique précis, en général le calcul différentiel. Ainsi, les modèles se présentent comme des systèmes de relations entre des variables datées. Ils sont généralement stochastiques, les aléas introduits exprimant des indéterminations intrinsèques (comportements inexplicables, technologies imprévisibles) ou des défauts de spécification des relations (variables omises, relations analytiquement fausses). Les théories sont exprimées dans un cadre axiomatique, et leurs conséquences sont déduites logiquement. En particulier, la théorie des jeux est entièrement formalisée et ses notions d'équilibre reçoivent des justifications tant épistémiques qu'évolutionnistes.

Les données sont obtenues par des expérimentations de trois types :

- l'« expérimentation historique » concerne le déroulement normal de la vie économique. Certains régimes historiques (ère Thatcher,

expérience soviétique) sont interprétés comme autant d'expériences particulières ;

- l'« expérimentation de terrain » consiste à modifier volontairement certaines variables de commande d'un secteur par rapport à un secteur témoin. Elle est appliquée à des réformes fiscales (TVA, impôt sur le revenu) ou juridiques (durée du travail) ;
- l'« expérimentation en laboratoire » consiste à fabriquer un système artificiel (mais avec des acteurs réels) en l'isolant de son contexte et en manipulant certaines influences externes. Elle a porté sur des modèles généraux (théorie de la décision, théorie des jeux) ou plus spécifiques (marchés fictifs).

Si le premier type reste le plus commun et conduit à la constitution de grandes bases de données internationales, le troisième type se développe rapidement (voir encadré).

Économie expérimentale

Depuis une vingtaine d'années, l'économie teste ses modèles en procédant à des expériences en laboratoire. Dans un environnement partiellement fixé et partiellement simulé, des acteurs procèdent à des choix soumis à des protocoles d'expérience. Ceux-ci conditionnent les états mentaux génériques des acteurs par l'influence de facteurs spécifiques (gains monétaires). Les domaines explorés sont divers, en théorie des jeux (choix en incertitude, jeux de coopération, choix social, vote) comme en théorie des échanges (biens publics, mécanismes d'enchères, transactions financières, procédures de négociation). Les observations sont de deux natures et déterminent deux étapes du développement de l'économie expérimentale. En psycho-économie, on observe les choix réels (ou intentionnels) qui ont été réalisés par les acteurs et les états mentaux associés affichés par ces acteurs. En neuro-économie, on enregistre les images cérébrales des processus de délibération effectuées par les acteurs. Les résultats de ces expériences, dûment répétées, agissent en retour sur les modèles qui sont proposés.

Pour les modèles, le langage mathématique joue un rôle non seulement de description mais de déduction. Des outils mathématiques nouveaux sont constamment importés pour mieux traduire les phénomènes (logique épistémique pour les croyances). Des outils de simulation sont également introduits pour faciliter la simulation numérique des modèles lorsque la résolution analytique est trop complexe (« systèmes multi-agents »). Les modèles déterminent des classes générales, qui correspondant à des structures mathématiques particulières et sont transférables d'un champ à un autre. Ainsi, les modèles de choix individuels en fonction de critères multiples ont été transférés à des choix collectifs. Les mécanismes de marché entre des biens instantanés (marchés au comptant) ont été appliqués à des biens datés (marchés à terme). Certaines de ces classes peuvent d'ailleurs provenir d'autres disciplines. Ainsi, la biologie évolutionnaire a fourni des mécanismes généraux (mutation, sélection) qui s'appliquent aux rapports entre des joueurs. De même, la physique des milieux désordonnés a fourni des formalismes qui s'appliquent à des marchés sous influence sociale.

Types d'articulation

La principale articulation consiste à ajuster les modèles et les données, deux démarches étant généralement mises en œuvre :

- la « démarche projective » suppose qu'il existe un modèle préexistant qui est ensuite confronté aux donnés ;
- la « démarche inductive » part des données et s'efforce d'inférer un modèle qui puisse en rendre compte.

La démarche projective est dominante en économie où des modèles d'origine variée sont ensuite testés au regard soit de données historiques, soit de données de laboratoire. En fait, une démarche inductive préalable est effectuée pour transformer les données brutes en faits stylisés, qui traduisent des régularités diverses (trends, corrélations lourdes, structures persistantes). Les modèles sont alors directement confrontés

aux faits stylisés plutôt qu'aux données brutes. En sens inverse, une démarche projective est toujours partiellement nécessaire si un modèle contient des grandeurs non observables (utilités) ou des opérateurs complexes (maximisation). Il s'agit alors de les éliminer et d'en dériver des conséquences testables qui seules peuvent être confrontées aux données.

L'articulation consistant à projeter les modèles sur les données se fait ensuite selon deux approches elles-mêmes distinctes :

- l'« approche confirmationniste » cherche à montrer que les faits connus sont en accord avec les conséquences du modèle ;
- l'« approche réfutationniste » consiste à rechercher des faits susceptibles d'infirmer les modèles.

Là encore, l'économie adopte une approche délibérément confirmationniste. Elle se contente souvent de montrer que des conséquences privilégiées des modèles sont en accord avec les faits. En effet, la réfutation d'un modèle économique est délicate, qu'elle soit trop facile ou trop difficile. Dans le cas d'un modèle déterministe, il est toujours réfuté avec suffisamment d'acharnement concernant les conséquences ultimes du modèle et la multiplication des données. Dans le cas d'un modèle stochastique, la réfutation n'est jamais possible, dans la mesure où il est toujours compatible avec les données. En pratique, la réfutation se fait alors à l'aide de méthodes statistiques qui introduisent des seuils conventionnels de rejet. Ici, les mathématiques remplissent un rôle nouveau en fournissant des méthodes d'adaptation des modèles aux données.

L'adaptation entre les modèles et les théories s'avère bien plus simple car les théories fournissent des principes généraux que les modèles déclinent de façon spécifique. Cette articulation est nécessaire car les théories demeurent souvent non directement testables si on ne leur associe pas des hypothèses supplémentaires plus précises. Ainsi, la théorie des jeux est fondée sur des principes généraux qui semblent s'imposer *a priori* ou apparaissent au moins comme des principes métho-

dologiques jugés sur leur fécondité. Le principe de rationalité n'est pas testable dans sa formulation générale, mais le devient dans des contextes précis (choix en incertitude, comportement de consommation). Le principe d'équilibre est lui-même admis *a priori*, toute situation réelle étant assimilée à une forme ou une autre d'équilibre. En pratique, les hypothèses théoriques sont remises en cause par vagues en fonction de deux critères. D'une part, des hypothèses alternatives formalisées doivent être disponibles. D'autre part, des conséquences originales doivent résulter des hypothèses alternatives. Ainsi, la rationalité limitée n'a été introduite que fort tard, lorsque ces conditions ont été satisfaites.

Les modèles apparaissent eux-mêmes comme des objets conceptuels idéaux qui ne traduisent les systèmes réels que sous certains aspects et dans certaines limites. Ils sont contrefactuels dans la mesure où ils expriment la possibilité de certains phénomènes plus que leur réalité. Ainsi, le modèle de l'équilibre général a pour objet de montrer qu'une coordination des agents par les seuls signaux-prix est possible. Il ne prétend pas que tel ou tel marché fonctionne effectivement selon ce modèle, même s'il est possible de montrer que certains marchés en sont plus proches que d'autres. Ainsi, le marché financier s'en approche beaucoup, le marché des voitures un peu moins, et le marché du travail encore moins. Dans l'évolution de la discipline, un modèle séminal est souvent proposé, qui rend compte d'un phénomène avec des hypothèses minimales. Ce modèle est ensuite enrichi pour être plus réaliste sans pour autant prétendre prendre en compte l'intégralité du phénomène.

Économie positive et économie normative

Les modèles servent d'abord à faire des prévisions opérationnelles sur les phénomènes économiques qui vont s'établir dans le futur. Ces prévisions concernent en particulier les effets induits par des mesures nouvelles de politique économique. De ce point de vue, l'économie suggère elle-même des instruments de politique économique, qui s'exercent sur les agents (impôt sur le revenu) et leurs relations (TVA) ou agissent

directement au niveau macroéconomique (politique monétaire). Les prévisions sont souvent situées dans un certain éventail, depuis les scénarios gris jusqu'aux scénarios roses. En fait, les modèles étant idéaux, il est nécessaire de réintégrer, au niveau des prévisions affichées, des considérations hors modèles. Ainsi, les prévisions vont faire apparaître une grande variabilité selon les modèles (pourtant souvent proches) et les modélisateurs. Cependant, on observe souvent des évaluations groupées qui vont s'avérer assez loin de la réalité. Par un effet Panurge, les prévisionnistes préfèrent, par sécurité, se tromper ensemble plutôt que d'avoir raison (ou de se tromper) tout seul.

Les modèles servent ensuite d'aide à la décision en contribuant à préconiser telle ou telle action possible. Ces décisions concernent les mesures de politique économique dont ils simulent et évaluent les effets. Là encore, l'économie propose ses propres critères de choix pour évaluer les états sociaux résultant de ces mesures. Différents types de critères sont proposés, mais ils agrègent généralement les préférences individuelles sur les états sociaux (voir encadré). Des divergences très fortes peuvent apparaître entre les préconisations des experts, souvent dues moins aux critères sociaux retenus qu'aux modèles de prévision mobilisés. Des divergences plus fortes encore apparaissent entre les préconisations des experts et les décisions prises par les décideurs publics. En effet, les décideurs procèdent souvent par « cannibalisme », à savoir qu'ils vont puiser dans les études qui leur sont proposées des données, des raisonnements ou des arguments qu'ils recombinent dans leur propre optique. Seule l'économie financière connaît une application quasiment en court-circuit entre certaines méthodes d'aide à la décision et les décisions prises.

Rôle de l'État

Si l'État est un acteur spécifique chargé de définir des règles du jeu du système économique, sa représentation a beaucoup évolué. En ce qui concerne ses actions, l'État édicte moins des mesures de politique

économique qu'il ne crée des institutions assurant une coordination satisfaisante des acteurs. Face à des situations de monopole naturel, il n'impose plus autant de contraintes aux entreprises publiques mais crée des agences de régulation. En ce qui concerne ses représentations, l'État apparaît moins comme un acteur en surplomb du système économique que comme un acteur comme les autres. Ainsi, il substitue à un modèle macroéconomique donnant une image de son environnement un modèle de jeu où il apparaît en interaction stratégique avec les autres. En ce qui concerne ses préférences, l'État passe de ses objectifs classiques d'efficacité et d'équité à des objectifs plus dynamiques comme l'adaptabilité du système économique. En tenant compte de l'évolution technologique, il s'efforce de ramener le sentier de l'économie sur une trajectoire stable.

Les modèles sont également diffusés vers les autres sciences sociales, qui fonctionnent selon des méthodes sensiblement différentes. Tout d'abord, elles recueillent d'autres formes de données, souvent plus qualitatives, en particulier à travers des monographies de certains secteurs. Ensuite, elles mettent en œuvre moins des modèles que des récits, qui certes procèdent aussi à des raisonnements, mais de façon plus linéaire et moins systématique. Enfin, elles proposent des théories qui s'avèrent souvent contradictoires, et introduisent même des concepts irréductibles les uns aux autres. Les modèles économiques exercent un effet de fascination sur les autres disciplines, au regard des possibilités ouvertes par les outils mathématiques. Ils exercent dans le même temps un effet de répulsion dû aux réductions de sens que les formalismes introduisent. Néanmoins, les modèles pénètrent de plus en plus les sciences sociales en apportant les bénéfices d'une simulation aisée.

Les modèles jouent enfin un rôle pédagogique indéniable auprès des étudiants, des médias et même du public. D'une part, ils fournissent aux acteurs des informations importantes sur l'environnement économique dans lequel ils sont insérés. Ils alimentent les processus de décision individuels en fournissant des ordres de grandeur de certaines variables ou de certains phénomènes. D'autre part, ils façonnent les

structures mentales des acteurs en imposant une vision du monde spécifique. Ils contribuent à la diffusion de concepts (valeur ajoutée, coût d'opportunité) comme de processus (externalités positives ou négatives, bulles spéculatives). Cette diffusion est à l'origine du phénomène d'autoréalisation des croyances, à savoir le fait que les croyances endossées par les acteurs peuvent provoquer ce qu'elles décrivent. Une variable anticipée est autoréalisatrice si la croyance en sa valeur provoque cette valeur (anticipation d'une pénurie de sucre). Une relation postulée est autoréalisatrice si la croyance en son existense provoque son apparition (croyance en une relation entre prix et taches solaires).

Pour aller plus loin, consulter la bibliographie en ligne.

Chapitre 12

Histoire de la pensée économique : les grands auteurs et leurs doctrines

GUIDO HÜLSMANN

Professeur à l'université d'Angers

La science économique est une discipline relativement jeune qui germe au Moyen Âge et se développe vigoureusement à partir du 18e siècle. C'est une science qui répond à un besoin pratique : celui de bien organiser les sociétés humaines. Elle devient la reine des sciences sociales au 19e siècle, lorsqu'en appliquant les préceptes des économistes de l'époque, on a remporté un succès spectaculaire, facilitant entre autres une croissance inouïe des revenus réels par tête et une croissance sans précédent des populations.

L'étude de son histoire présente quatre enjeux principaux.

- *Premièrement*, elle permet aux chercheurs contemporains de se situer par rapport à la division du travail qui se fait entre les générations. Pourquoi étudions-nous aujourd'hui tel ou tel problème avec telles ou telles méthodes ? Pour répondre à cette question, il faut remonter dans le temps.
- *Deuxièmement*, elle nous permet de relativiser nos préoccupations actuelles. Lire un ouvrage publié dans un passé plus ou moins lointain est souvent comme une bouffée d'air frais qui nous fait sortir des sentiers battus.
- *Troisièmement*, elle peut nous éviter de la peine en nous rendant conscients des erreurs typiques et récurrentes.

- *Quatrièmement*, l'histoire de la pensée ne s'étudie pas sans lire les textes des grands auteurs ; et ceci est *une* manière d'apprendre la science elle-même, peut-être même la meilleure manière. En effet, les grands textes ont souvent une rigueur et une cohérence logique qu'on trouve dans peu de manuels de micro et de macroéconomie actuels, dans lesquels des modèles, parfois sans rapport ou contradictoires, se juxtaposent. Ils invitent à la réflexion sur les questions de principe qui se sont posées de leur temps tout comme aujourd'hui. Le *jargon* des économistes est en mutation permanente, et il y a de ce fait un risque de ne plus comprendre les anciens. En revanche, le fond des problèmes ne change ni souvent ni beaucoup.

Aussi, dans le présent chapitre, allons-nous faire revivre 20 grands auteurs, dans l'espoir de donner au lecteur l'envie de les lire. Il a donc fallu sélectionner 20 grands économistes – *exactement* 20 – non pas 25, ni 50, ni 100. Notre choix a été douloureux et n'est sans doute pas juste à certains égards. Toutefois, il n'est pas *plus* injuste que les choix alternatifs.

Les auteurs sont présentés dans l'ordre de leur date de naissance. Cependant, il faut garder à l'esprit qu'ils sont souvent contemporains et s'influencent les uns les autres, même si cette influence n'est pas toujours explicite. Nous avons tâché de retracer les connexions intellectuelles les plus importantes entre eux. Mais ici encore, il existe un danger qu'il convient de souligner.

L'histoire de la pensée – quelle que soit d'ailleurs la discipline – n'est pas un processus linéaire par lequel la qualité des idées augmente sans cesse, sans jamais diminuer. Il n'est pas toujours vrai que les théories les plus récentes « incorporent » toute la sagesse des anciens, et soient, de ce fait, supérieures aux précédentes. Bien au contraire, il y a des périodes de progrès et des périodes de régression. Il arrive assez fréquemment que les acquis intellectuels se perdent, qu'une génération oublie les leçons apprises par les générations antérieures et qu'elle doive réapprendre avant de continuer à bâtir.

PLATON

(427-347 avant. J.-C.)

Au commencement : la division du travail

Le père de la philosophie occidentale dédaigne les problèmes économiques. Sa pensée sociale concerne la définition d'une bonne cité *politique*. Mais pour la définir, dans son livre sur la *République*, Platon doit d'abord analyser la division du travail.

Il reconnaît que la cité n'est pas forcément construite à partir d'un plan unique. Elle peut émerger des actions individuelles sans coordination centrale. C'est ce cas de figure que Platon étudie de plus près, notamment pour en repérer les limitations. Il examine d'abord le cas de la « cité simple » puis celui de la « cité luxueuse ».

Les hommes s'associent pour mieux satisfaire leurs besoins. C'est par ce motif économique qu'ils décident de vivre ensemble et de s'aider mutuellement. Or, afin de s'organiser, la division des tâches et la spécialisation s'imposent pour trois raisons : pour exploiter les différents talents individuels ; parce que la répétition des gestes perfectionne tout métier ; et parce que certaines tâches, devant être accomplies au bon moment, nécessitent alors la présence permanente d'un responsable. Résultat général : la productivité des efforts augmente, et les efforts sont faits de manière esthétique et avec une plus grande facilité.

La division du travail ne se limite cependant pas aux biens de consommation. Elle couvre aussi les outils de l'agriculture, les matières premières et l'élevage. Pour assurer ces activités, le nombre de citoyens doit augmenter ; et la taille de la cité croît encore lorsque la division du travail s'internationalise. Les citoyens ont intérêt à coopérer avec d'autres cités en raison des imperfections naturelles de l'endroit où leur cité est érigée. Cette coopération se fait par l'échange, chaque cité cédant les marchandises dont elle est bien pourvue pour acquérir celles qui y sont relativement rares.

Qu'implique cet échange pour les quantités produites ? Platon souligne qu'il incite les citoyens à produire certains biens au-delà de leurs

propres besoins. L'échange n'est pas dissocié de la production, bien au contraire, il détermine les types de biens qui sont produits ainsi que leurs quantités. Aristote soutiendra plus tard une thèse opposée, à savoir que l'échange ne concerne que le « surplus » dont on n'aurait pas besoin et qui résulte plus ou moins accidentellement d'une production particulièrement abondante. Mais sur ce point il a tort, et Platon a raison. L'échange élargit la division du travail, crée des industries nouvelles telles que les transports maritimes et agrandit la taille de la cité.

Platon se tourne après vers l'analyse de la « cité luxueuse » qui surgit lorsque les gens cèdent aux tentations de la vie et abandonnent le style de vie simple de leurs ancêtres. Ils s'assoient sur des coussins et à table, mangent des desserts, utilisent pommades et parfums, couchent avec des prostituées et aiment la pâtisserie, le théâtre et les concerts. Pour satisfaire ces nouveaux besoins, il faut des produits nouveaux ; la division du travail est encore élargie, et la cité s'agrandit.

Mais cette croissance « immodérée » rencontre bientôt un obstacle lorsque la surface de la cité ne peut plus être étendue sans envahir le territoire des cités voisines. Platon ne considère pas la possibilité d'un achat de ces terrains supplémentaires pour éviter un affrontement violent. Pour lui la guerre est, dans ce cas, certaine. Ainsi l'opulence n'est-elle pas seulement la cause de la décadence des individus et de la société ; elle est également la source de la guerre.

NICOLAS D'ORESME
(1325-1382)

Le mal de l'inflation

L'étude systématique des lois économiques commence au Haut Moyen Âge ; les premiers économistes sont les théologiens scolastiques de l'école de Paris. Le premier d'entre eux qui écrit un traité scientifique entièrement consacré à un sujet économique s'appelle Nicolas d'Oresme. Vers 1360, il rédige son *Traité sur l'origine, la nature, le droit et les mutations des monnaies* qui résume et développe les idées des scolastiques de son temps. C'est la naissance de la science économique.

Au cœur de son analyse monétaire se trouve le problème des « mutations » de la monnaie, c'est-à-dire des altérations du contenu métallique des pièces et de leur dénomination. Ces altérations ont lieu depuis l'aube des temps et sont bien documentées pour l'Antiquité et le Moyen Âge. Leur effet le plus visible est de changer le pouvoir d'achat de l'unité monétaire, en particulier de le diminuer – il s'agit alors d'une forme primitive d'inflation. Tandis que le *Traité* concerne uniquement cette forme primitive, nombre de ses thèses peuvent être généralisées et s'appliquent également aux Temps modernes.

Oresme soulève d'emblée une question centrale : l'inflation est-elle utile pour la communauté ? Il répond par la négative, soutenant que l'inflation ne rend la monnaie ni plus ni moins utile pour les échanges. L'économie peut bien fonctionner quel que soit le niveau des prix, et donc quelle que soit la masse monétaire nominale.

Mais si tel est le cas, une autre question se pose évidemment : pourquoi les altérations de la monnaie existent-elles ? Et en particulier, pourquoi chercher à augmenter la masse monétaire ? Oresme répond que ces altérations n'ont pas les mêmes conséquences pour les différents membres de la communauté. Elles profitent à certaines personnes au détriment d'autres. Les gagnants prospectifs des altérations de la monnaie ont un intérêt matériel à les mettre en place. En règle géné-

rale, ces gagnants sont les gouvernements. Oresme dit : « Il me semble que la cause première et dernière pour laquelle le prince veut s'emparer du pouvoir de muer les monnaies, c'est le gain ou profit qu'il peut en avoir, car autrement, c'est sans raison qu'il ferait des mutations si nombreuses et si considérables. » Puis il rajoute encore ces précisions :

> *« Tout ce que le prince en retire de gain, c'est nécessairement aux dépens de la communauté. Or, tout ce qu'un prince fait aux dépens de la communauté est une injustice et le fait, non d'un roi, mais d'un tyran, comme dit Aristote… Si le prince peut, à bon droit, faire une mutation simple de la monnaie et en retirer quelque gain, il peut, pour une raison analogue, faire une plus grande mutation et en retirer plus de gain… Ainsi le prince pourrait enfin attirer à lui presque tout l'argent ou les richesses de ses sujets et les réduire à la servitude, ce qui serait faire entièrement preuve de tyrannie et même d'une vraie et parfaite tyrannie, comme il ressort des philosophes et des histoires des anciens. »*

Oresme souligne que les altérations de la monnaie ne sont pas simplement un jeu de redistribution en faveur du gouvernement au détriment du reste de la communauté. Elles entraînent des pertes *globales* – le jeu est à somme négative. Une monnaie en altération fréquente perturbe le commerce, surtout le commerce international, et elle invite les faux-monnayeurs à profiter de la confusion générale. De plus, si deux monnaies différentes profitent du cours légal, les agents vont thésauriser celle qui vaut plus, de manière que seule la monnaie inférieure demeure en circulation. (Oresme anticipe ici la fameuse « loi de Gresham » : la mauvaise monnaie chasse la bonne dans un régime de cours légal.) Il conclut que les manipulations de la monnaie sont pires que l'usure et que, probablement, elles ont été une cause importante du déclin de l'Empire romain.

JUAN DE MARIANA
(1536-1624)

La valeur est subjective

Oresme avait résumé et développé les travaux des premiers scolastiques sur la monnaie. Deux siècles plus tard, le jésuite espagnol Juan de Mariana fait un travail analogue, proposant une synthèse des travaux de l'école de Salamanque sur le même sujet. Cette école regroupe des théologiens jésuites et dominicains qui, au cours des 16e et 17e siècles, créent une analyse morale des métiers et des firmes en essor à cette époque. Définir une éthique des banques et du commerce de gros présuppose qu'on comprenne leur fonctionnement, ainsi que les mécanismes d'une économie monétaire en pleine mutation (exploitation des mines d'or et d'argent en Amérique latine à partir des années 1520).

À un moment caractérisé par le désarroi complet des finances de la couronne d'Espagne et par l'hyperinflation (1598-1600), Mariana analyse ces problèmes dans son ouvrage *De la mutation des monnaies*, publié en 1605. Il traite des altérations de la monnaie, examinant leur légitimité en tant qu'outil des finances publiques, et leurs conséquences économiques.

Mariana souligne qu'une altération de la monnaie fonctionne comme un impôt supplémentaire qui est payé par toute personne ayant été propriétaire de monnaie *avant* que l'altération fasse monter les prix. Aussi les altérations doivent-elles sans exception être approuvées par la communauté, car les biens des citoyens n'appartiennent pas au roi et leurs encaisses monétaires ne font pas exception. Mariana rejette d'ailleurs la justification de l'impôt d'inflation par la nécessité des activités de l'État, préconisant une réduction radicale des dépenses publiques.

Même si la communauté donne son accord à la mutation de la monnaie, cela n'empêchera pas que des effets néfastes se manifestent. Mariana fait valoir en particulier que le niveau des prix augmente dans la même proportion que la valeur de la monnaie chute. Il se fonde sur

la théorie quantitative de la monnaie formulée par Martin Azpilcueta Navarro (1493-1587), un autre membre de l'école de Salamanque. Celui-ci écrit en 1556 :

> *« Comme on voit par expérience, en France, où il y a moins de monnaie qu'en Espagne, le pain, le vin, les vêtements et le travail salarié coûtent bien moins ; et même en Espagne, aux temps où il y avait moins de monnaie, les choses à vendre et le travail salarié ont été cédés pour bien moins qu'après que la découverte de l'Inde la comblait d'or et d'argent. C'est parce que la monnaie vaut plus là où elle manque que là où elle se trouve en abondance. »*

Selon Mariana, l'expérience prouve qu'il est impossible pour le gouvernement d'empêcher cette augmentation des prix lorsque la masse monétaire est en croissance. L'imposition de prix plafonds est inefficace.

Mais il existe une conséquence encore plus grave : l'inflation fait dévier les prix par rapport à l'« estimation commune », qui est la véritable source des valeurs monétaires. Cette source de valeur est subjective. Mariana s'appuie ici sur une théorie élaborée par d'autres membres de l'école de Salamanque : « La valeur d'un article ne dépend pas de sa nature essentielle, mais de l'estimation subjective par les hommes, même si cette estimation est malavisée » (Diego de Covarrubias y Leyva). La valeur monétaire d'une marchandise et son juste prix en particulier n'ont rien à voir avec « le travail, les coûts et le risque assumés par la personne trafiquant cette marchandise… On trouve le juste prix, non pas par la comptabilité des coûts, mais par estimation commune » (Luis Saravia de la Calle).

RICHARD CANTILLON
(1680-1734)

L'économie est un circuit

À partir de la seconde moitié du 17e siècle, la technique de l'altération des pièces monétaires pour remplir le trésor de l'État tombe en désuétude avec la montée irrésistible de la création monétaire par les banques. La nouvelle donne nécessite alors une modification des anciennes théories monétaires. Ce travail, commencé par les scolastiques de l'école de Salamanque, est également avancé par des auteurs laïcs, notamment par des praticiens de la banque tels que John Law et Richard Cantillon.

Ce dernier écrit vers 1730 un traité sur *La Nature du commerce en général*, dans lequel il aborde l'analyse monétaire et bancaire à partir d'une étude préalable des forces économiques qui façonnent la taille de la population et sa distribution dans l'espace. Publié vingt ans après la mort de son auteur, l'ouvrage est le premier traité d'économie générale. Il fait apparaître les contours d'une nouvelle science transversale des richesses.

Cent ans après la découverte de la circulation sanguine, Cantillon démontre que tous les secteurs de l'économie sont similairement liés – par les flux monétaires. Il y a un véritable circuit de dépenses et de recettes :

> *« La circulation de cet argent se fait en ce que les propriétaires dépensent en détail, dans la ville, les rentes que les fermiers leur ont payées en gros articles, et que les entrepreneurs des villes, comme les bouchers, les boulangers, les brasseurs, etc., ramassent peu à peu ce même argent, pour acheter des fermiers, en gros articles, les bœufs, le blé, l'orge, etc. Ainsi toutes les grosses sommes sont distribuées par petites sommes, et toutes les petites sommes sont ensuite ramassées pour faire des paiements de grosses sommes aux fermiers, directement ou indirectement, et cet argent passe toujours en gage tant en gros qu'en détail. »*

Les agents décisifs de l'économie sont, d'un côté, les propriétaires fonciers et, de l'autre, les entrepreneurs. Les premiers décident de la manière dont les terres sont utilisées et donc du nombre de travailleurs qui trouvent subsistance dans le pays, les seconds ajustent la production à la demande, tout en assumant l'incertitude de l'avenir.

Dans son analyse monétaire, Cantillon affirme que toute variation de la masse monétaire entraîne une variation du niveau des prix et par conséquent du niveau des dépenses. Il insiste cependant sur le fait que cela *ne* veut *pas* dire que tous les prix seront également affectés. En général, c'est même l'inverse qui sera le cas, précisément parce que les dépenses se succèdent dans le circuit monétaire. Toute variation de la masse monétaire fait alors augmenter les prix de différentes marchandises à *différents* points dans le temps et à *différents* degrés. Ce sont là les « effets Cantillon » des variations de la masse monétaire.

Cantillon anticipe sur les économistes classiques en mettant en garde contre des conceptions qui attribuent à la masse monétaire un pouvoir créateur de richesses. En particulier, il nie que les augmentations de la masse monétaire entraînent nécessairement une diminution des taux d'intérêt et par conséquent une stimulation de la production. De toute manière, les augmentations artificielles de la masse monétaire ne peuvent pas durer car elles provoquent une augmentation des prix, entraînant une diminution des exportations par rapport aux importations. La circulation monétaire artificielle sera vite réexportée. Pour des raisons analogues, Cantillon relativise les effets bénéfiques du papier-monnaie et du crédit bancaire. Leur avantage « est bien moins considérable et moins solide qu'on ne pense généralement ».

ÉTIENNE BONNOT DE CONDILLAC
(1714-1780)

La logique de l'économie

L'an 1776 voit la publication de deux ouvrages magistraux. Adam Smith et Étienne de Condillac synthétisent un siècle de réflexion à propos des politiques économiques et arrivent essentiellement aux mêmes conclusions. Leurs centres d'intérêt divergent cependant. Pour Adam Smith, c'est l'analyse des mécanismes de la croissance qui est prépondérante ; pour Condillac, c'est la cohérence logique du raisonnement économique.

Philosophe célèbre, Condillac ne publie qu'un seul ouvrage d'économie : *Le Commerce et le Gouvernement, considérés relativement l'un à l'autre.* Il se propose notamment de faire de l'analyse économique une véritable science :

> *« Chaque science demande une langue particulière, parce que chaque science a des idées qui lui sont propres. Il semble qu'on devrait commencer par faire cette langue ; mais on commence par parler et par écrire, et la langue reste à faire. Voilà où en est la science économique, dont l'objet est celui de cet ouvrage même. C'est, entre autres choses, à quoi on se propose de suppléer. »*

Condillac présente la science économique de manière axiomatico-déductive, en partant des phénomènes les plus fondamentaux pour après considérer des questions de plus en plus spécifiques. Il commence par une discussion de la valeur, puis se tourne vers les prix, ensuite vers les marchés organisés, et finalement vers d'autres sujets. Dans sa théorie de la valeur, il développe la conception de l'école de Salamanque, insistant sur le fait que la valeur subjective ou estimée d'une chose est « fondée » sur son utilité – sa capacité *objective* à servir « à quelques-uns de nos besoins ». Or, le fait crucial est que cette utilité d'une chose dépend de sa quantité :

> *« Dans un lieu aride [l'eau] a une grande valeur ; et on l'estime en raison de l'éloignement et de la difficulté de s'en procurer. En pareil cas un voyageur altéré donnerait cent louis d'un verre*

> *d'eau, et ce verre d'eau vaudrait cent louis ; car la valeur est moins dans la chose que dans l'estime que nous en faisons, et cette estime est relative à notre besoin ; elle croît et diminue, comme notre besoin croît et diminue lui-même. »*

Appliquant cette conception à l'analyse des échanges, Condillac souligne que les biens échangés n'ont *pas de valeur égale*. Chacun cède un bien qui, pour lui, a une moindre valeur pour obtenir un bien qui, pour lui, a une valeur plus élevée. L'échange est toujours bénéfique pour chacune des parties concernées.

Après l'analyse du commerce tel qu'il se présenterait dans un marché libre, Condillac se met à étudier les effets des interventions politiques. Il appelle ces interventions des « atteintes portées au commerce ». Son analyse comparative entre commerce et gouvernement cherche à dégager un point fondamental de logique, à savoir le caractère nuisible des guerres, des droits de douane et des quotas, des impôts sur la production et la consommation, des privilèges légaux pour certains agents, des altérations de la monnaie et des dettes publiques.

L'interventionnisme n'est pas toujours motivé par les projets propres du gouvernement. Parfois ce sont des groupes extérieurs, notamment des groupes de commerçants, qui pressent l'État d'intervenir en leur faveur. Mais peu importe l'origine des interventions, celles-ci engendrent une dynamique perverse car le gouvernement est tenté de poursuivre des interventions de plus en plus amples, dans le vain espoir de rectifier les échecs de ses interventions précédentes.

ADAM SMITH
(1723-1790)

Nature et causes de la richesse des nations

Le plus célèbre ouvrage économique fonde un mouvement intellectuel mondial qui bouleverse les politiques économiques au cours du siècle suivant. C'est le mouvement des économistes classiques.

L'auteur de l'*Enquête sur la nature et les causes des richesses des nations*, l'Écossais Adam Smith, semble être peu concerné par les questions d'épistémologie et de logique. Mais dans sa grande synthèse de la pensée économique du 18e siècle, il présente tout un programme scientifique et politique à contre-courant d'une orthodoxie qu'il appelle « mercantiliste ».

Selon la théorie mercantiliste, le niveau des dépenses monétaires est un facteur crucial de la richesse globale d'une nation. Plus on augmente ces dépenses, plus de bras sont employés, plus de biens sont produits et donc consommables. Puisque les dépenses faites pour les biens de production dépendent, en fin de compte, des dépenses attendues pour les biens de consommation, il s'ensuit que chaque nation peut s'enrichir en stimulant le niveau de ses dépenses à la consommation. À l'inverse, tout ce qui ralentit ces dépenses risque de suffoquer la production et de porter ruine à la nation. L'épargne et la thésaurisation sont condamnables d'un point de vue global.

Le niveau des dépenses est surtout limité par la masse monétaire circulant à l'intérieur de la nation. La politique économique devrait chercher, par des monopoles et des subventions, à stimuler les exportations (pour faire venir en échange davantage de monnaie), et à freiner les importations par des droits de douane, des quotas et des prohibitions de certains biens étrangers (pour empêcher des fuites de monnaie vers l'extérieur).

Le problème est que, tant qu'on utilise des monnaies métalliques, cette politique ne saurait être appliquée par toutes les nations en même

temps, car les excès d'exportation des uns sont les déficits des autres. La plus grande richesse des uns est la cause de la ruine des autres. Alors, on se dispute les marchés et on mène des guerres.

Selon Smith, ce raisonnement est futile. D'abord, il ne faut pas confondre la richesse avec le niveau des dépenses monétaires. La richesse d'une nation se mesure par le produit annuel de ses citoyens soit consommé par eux-mêmes, soit échangé contre des marchandises étrangères. Toute hausse de la masse monétaire au-delà des besoins du public sera rejetée par le marché, sans toucher le produit réel. Puis Smith démontre dans le détail que les restrictions des importations réduisent la division du travail et appauvrissent, de ce fait, *tous* les pays concernés. De même, la stimulation politique des exportations réduit le produit global de *tous* les pays, car elle détourne l'investissement des projets les plus rentables.

Smith admet seulement deux véritables causes de la richesse. La première est la division du travail qui génère un surplus par rapport à une production sans coordination entre les individus. La seconde, et la plus importante, est l'épargne ou le stock du capital accumulé. Il ne s'agit pas d'une somme monétaire, mais d'un fonds de biens réels, composé notamment de biens qui sont nécessaires à la survie des travailleurs. En effet, leurs salaires sont payés par ce stock. Par conséquent, plus il y a d'épargne, plus les salaires sont élevés ; plus grand sont alors l'emploi, la division du travail et la production.

La voie royale pour augmenter la richesse d'une nation consiste par conséquent à encourager les citoyens à mener une vie frugale, privilégiant l'épargne et réduisant la consommation. Cette voie peut être empruntée par *toutes les nations* en même temps ; et même si ce n'est qu'une seule nation qui opte pour la frugalité, elle en fait bénéficier toutes les autres. Elle agrandit non seulement sa propre richesse, mais *les* richesses de toutes les nations en même temps.

JEAN-BAPTISTE SAY
(1767-1832)

Les lois de la production

La thèse centrale d'Adam Smith est inouïe : le progrès matériel est possible pour tous et il relève du choix humain. En choisissant l'épargne, on peut accroître la *productivité par tête*. Il s'ensuit que la productivité n'est pas une affaire purement technologique, mais régie aussi par des lois économiques. Ces lois devraient occuper une place de choix dans l'analyse économique.

Le premier à tirer cette conclusion est Jean-Baptiste Say, l'auteur d'un *Traité d'économie politique* (1803), le manuel d'économie le plus utilisé au 19e siècle. Say subdivise la science économique en trois grands compartiments : production, distribution et consommation. Cette distinction se maintient jusqu'au 20e siècle, où elle est remplacée par celle entre micro et macroéconomie.

Say combine l'analyse smithienne avec la théorie de la valeur subjective. Il insiste sur le fait que la production ne vise pas, au fond, un objet matériel mais une satisfaction, donc un but immatériel. Le plus célèbre apport de Say à la théorie de la production est cependant la « loi des débouchés », qu'on appelle aussi la « loi de Say ». D'autres économistes l'ont découverte avant lui, mais il est le premier à saisir son importance et à la présenter comme une loi fondamentale de la production.

Il constate que, dans une économie de marché, le volume de chaque production particulière est limité par la demande c'est-à-dire qu'il est limité par le nombre de biens que d'autres personnes offrent en échange pour le produit en question. Bien que les échanges se fassent par l'intermédiaire de la monnaie, le producteur ne s'intéresse pas au nombre de billets qu'on lui propose en échange de son produit X, mais aux biens Y, Z, etc. qu'il peut acheter avec ces billets. Il désire un paiement réel et non pas un paiement nominal.

Afin d'augmenter la demande pour X, il faut accroître les quantités de Y, Z, etc. proposés en échange contre X. Mais cela veut dire qu'il faut augmenter la *production* de Y, Z, etc. pour augmenter la demande de X !

Il s'ensuit que chaque production crée un débouché à d'*autres* produits. Selon les mots de Say, « *un produit terminé offre,* dès cet instant, *un débouché à d'autres produits pour tout le montant de sa valeur* ». Quelle que soit alors la croissance de la production *globale*, la demande globale augmente toujours au même rythme que la production. La moitié de la production globale est pour ainsi dire échangée contre l'autre moitié.

Par conséquent, il ne peut pas y avoir de problème de surproduction *globale*. Il est possible de produire trop de certains biens *particuliers*. Il est aussi possible qu'il y ait des crises générales. Mais les causes de ces problèmes doivent être cherchées ailleurs que dans une production globale qui serait excessive par rapport à la demande globale.

Parmi les conséquences qui découlent de cette analyse, Say souligne en particulier que, « *dans tout État, plus les producteurs sont nombreux et les productions multipliées, et plus les débouchés sont faciles, variés et vastes* ». Il poursuit :

> *« [Chacun] est intéressé à la prospérité de tous, et la prospérité d'un genre d'industrie est favorable à la prospérité de tous les autres… Une nation, par rapport à la nation voisine, est dans le même cas qu'une province par rapport à une autre province, qu'une ville par rapport aux campagnes : elle est intéressée à la voir prospérer, et assurée de profiter de son opulence… [De même] l'importation des produits étrangers est favorable à la vente des produits indigènes ; car nous ne pouvons acheter les marchandises étrangères qu'avec des produits de notre industrie, de nos terres et de nos capitaux, auxquels ce commerce par conséquent procure un débouché. »*

DAVID RICARDO
(1772-1823)

Les mécanismes économiques

Pendant les guerres napoléoniennes, les prix ne cessent de monter en Grande-Bretagne. Pourquoi ? En raison des émissions de la Banque d'Angleterre ? Selon la théorie monétaire d'Adam Smith, cette réponse est inadmissible. Après tout, les billets bancaires ne circulent que dans les limites du besoin que le public en éprouve. La question suscite une fameuse controverse au sein du Parlement britannique, qui se termine en 1810 par la publication d'un rapport (*Bullion Report*)… qui donne tort à Smith.

David Ricardo est en première ligne parmi ceux qui mènent la théorie quantitative de la monnaie à la victoire. Il se met après à inventer un nouveau mécanisme monétaire : l'étalon change-or. Pour combiner les vertus du papier-monnaie avec ceux d'une monnaie métallique, il préconise une circulation de billets bancaires qui seraient convertibles, sur demande du porteur, en lingots d'or. Ainsi la monnaie aurait un *ancrage* dans l'or, mais l'or ne serait tout de même pas en circulation, car inutilisable sous forme de lingots dans les échanges quotidiens. Vingt ans après la mort de Ricardo, la Grande-Bretagne met son plan en pratique. L'étalon change-or s'internationalise à la fin du 19e siècle lorsque les autres pays adoptent le modèle britannique. Il s'éteint en 1971 avec la fin du système de Bretton Woods.

Grand théoricien monétaire de son temps, Ricardo rejette l'idée que la production monétaire saurait stimuler la production, mais il la rejette sans tomber dans l'erreur inverse qui consiste à nier toute influence de la monnaie sur l'économie réelle. La production monétaire modifie la composition du produit national. Elle change les types de biens produits et leurs quantités respectives. Mais elle n'accroît pas pour autant le résultat global.

En 1817, Ricardo publie ses *Principes d'économie politique et de la taxation*, chef-d'œuvre du raisonnement économique. Il se fonde d'emblée sur la théorie de la valeur travail proposée par Adam Smith selon laquelle « le prix réel de chaque chose, ce que chaque chose coûte réellement à l'homme qui veut l'acquérir, c'est le travail et la peine de l'acquérir ». Ce point de départ l'amène entre autres à la conclusion que le profit de l'entrepreneur est réalisé au détriment des salaires.

Cependant, un grand nombre de ses analyses ne dépendent pas de ce point de départ. Cela se confirme par exemple dans le fameux chapitre dédié au commerce international, dans lequel Ricardo présente le principe des avantages comparatifs. Il considère le problème de la coopération avec un partenaire qui est supérieur dans *tous* les secteurs de la production. Son exemple : le Portugal est *absolument* plus efficace que l'Angleterre dans la production de vins *et* de draps ; alors, comment peut-il y avoir une coopération entre les deux pays ? Sa réponse : la coopération doit, dans ce cas, exploiter les avantages *relatifs*. Si la supériorité du Portugal est encore plus grande dans la production du vin que dans la production des draps, alors ce pays devrait se spécialiser dans le vin et en échanger une partie contre des draps anglais. Ainsi les deux pays bénéficieront de quantités plus grandes des deux marchandises.

Ce principe n'est pourtant pas un aspect spécial des relations internationales. Il s'agit plutôt d'une loi économique générale qui s'applique également aux actions individuelles. Ricardo donne l'exemple suivant :

> *« Supposons deux ouvriers sachant l'un et l'autre faire des souliers et des chapeaux : l'un d'eux peut exceller dans les deux métiers ; mais en faisant des chapeaux il ne l'emporte sur son rival que d'un cinquième, ou de vingt pour cent, tandis qu'en travaillant à des souliers, il a sur lui un avantage d'un tiers, ou de trente-trois pour cent. Ne serait-il pas de l'intérêt de tous les deux que l'ouvrier le plus habile se livrât exclusivement à l'état de cordonnier, et le moins adroit à celui de chapelier ? »*

FRÉDÉRIC BASTIAT
(1801-1850)

Démystifier l'État

Les économistes classiques s'opposent à l'activisme de l'État en démontrant l'inefficacité de ses politiques. Une restriction des importations ne protège pas la production nationale ; les subventions à l'exportation ne stimulent pas la croissance ; la consommation de l'État n'est pas le moteur de la croissance. Bref, il ne suffit pas que l'État *veuille* faire du bien. Il doit agir raisonnablement car il est contraint par des lois économiques.

Dans un essai sur *La Loi* (1850), Frédéric Bastiat parachève la critique classique de l'étatisme en analysant le fonctionnement même de l'État. Qu'est-ce que c'est que l'État, en fait ? Dans la tradition occidentale, il est défini par sa *mission* : faire respecter la loi. Bastiat souligne que, quelle que soit sa mission, l'État est en premier lieu une *organisation sociale*. Autrement dit, il y a des êtres humains qui gèrent l'État, et ces personnes poursuivent forcément leurs objectifs propres. Il en résulte une dynamique qui tend à pervertir la loi :

> *« La Loi, c'est l'organisation du Droit naturel de légitime défense ; c'est la substitution de la force collective aux forces individuelles… pour faire régner entre tous la Justice… Mais la Loi est faite, le plus souvent, par un homme ou par une classe d'hommes. Et la Loi n'existant point sans sanction, sans l'appui d'une force prépondérante, il ne se peut pas qu'elle ne mette en définitive cette force aux mains de ceux qui légifèrent… On conçoit comment, au lieu d'être un frein à l'injustice, elle devient un instrument et le plus invincible instrument d'injustice. On conçoit que, selon la puissance du législateur, elle détruit, à son profit, et à divers degrés, chez le reste des hommes, la Personnalité par l'esclavage, la Liberté par l'oppression, la Propriété par la spoliation. »*

Pour empêcher que cette dynamique perverse se déclenche, il faudrait limiter d'emblée les compétences de l'État et notamment son budget. Mais ne renoncerait-on pas de cette façon aux effets bénéfiques de ses actions ? Pas selon les économistes. Ils admettent que les dépenses de l'État aient des conséquences positives pour certains bénéficiaires. Mais toutes ces dépenses ont d'abord été enlevées aux contribuables. Aussi toute augmentation des dépenses publiques va-t-elle de pair avec une *diminution* de certaines dépenses *ailleurs* dans le système économique. Les uns gagnent, d'autres perdent, sans qu'il n'y ait de gains nets.

Pourquoi est-il alors si difficile en pratique de freiner la croissance de l'État et la perversion de la loi qui en résulte ? Parce que, nous dit Bastiat, chaque citoyen espère être parmi les quelques heureux gagnants des activités étatiques :

> *« Donc, tous, à un titre quelconque, sous un prétexte ou sous un autre, nous nous adressons à l'État. Nous lui disons : "Je ne trouve pas qu'il y ait, entre mes jouissances et mon travail, une proportion qui me satisfasse. Je voudrais bien, pour établir l'équilibre désiré, prendre quelque peu sur le bien d'autrui. Mais c'est dangereux. Ne pourriez-vous me faciliter la chose ? Ne pourriez-vous me donner une bonne place ? Ou bien gêner l'industrie de mes concurrents ? Ou bien encore me prêter gratuitement des capitaux que vous aurez pris à leurs possesseurs ? Ou élever mes enfants aux frais du public ?" »*

Mais cet espoir est vain. La croissance de l'État ne bénéficie qu'à une minorité aux dépens du grand nombre. Dans son article *L'État* (1848), Bastiat propose de le définir de la manière suivante : « L'État, c'est la grande fiction à travers laquelle tout le monde s'efforce de vivre aux dépens de tout le monde. »

KARL MARX
(1818-1883)

L'exaltation du travail

La division du travail crée de multiples bénéfices matériels. Mais les relations sociales ne sont pas toujours équitables. *Homo homini lupus* – il y a pour l'homme le danger d'être spolié par d'autres hommes. Selon les économistes libéraux autour de Bastiat, la plus grande source de spoliation, c'est l'État. Par contre, les socialistes ont repéré des formes de spoliation, au sein même du marché. Ce sont notamment les capitalistes qui exploitent les travailleurs.

Mais en quoi consiste précisément ce rapport inéquitable entre capitalistes et salariés ? Dans *Le Capital* (1867), Karl Marx cherche à répondre à cette question et à donner ainsi un fondement scientifique au socialisme.

Il souligne que tout échange implique que les biens échangés aient une *valeur égale*. Quelle est la source de cette valeur ? Selon Marx, qui développe la théorie de la valeur-travail ébauchée par Smith et Ricardo, c'est la quantité de travail « socialement nécessaire » pour produire le bien en question, dans un contexte historique donné. Cette quantité de travail est en fait la *seule* source de valeur.

Par conséquent, on devrait s'attendre à ce que les revenus du travail soient les seuls revenus obtenus par la division du travail. Tel n'est cependant pas le cas dans les économies de marché, où l'on observe d'autres types de revenus, notamment les profits des entreprises – le surplus de leurs ventes par rapport à leurs coûts de production. Mais comment est-il possible de réaliser un tel surplus ? Après tout, la valeur du travail devrait être exactement égale à la valeur des produits de ce travail. Marx répond que les capitalistes ne paient pas aux travailleurs la valeur intégrale qui leur est due. Ils savent par exemple que les produits vaudront la somme de 120, mais au lieu de payer cette même somme à leurs ouvriers, ils paient seulement 100 et gardent la différence. Les ouvriers

doivent subir cette injustice parce que leurs adversaires contrôlent les moyens de production. Ils sont forcés de fournir quelques heures de travail sans compensation. La valeur produite pendant ces heures est la « plus-value » et se reflète dans un « taux de plus-value » – relation quantitative entre la plus-value et le « capital variable » investi dans les salaires des ouvriers. Voilà la source du taux de profit (taux de d'intérêt) observable sur le marché.

Cependant, dans la vie réelle, les taux d'intérêt se calculent non pas sur la base du seul capital investi dans les salaires (le « capital variable » de Marx), mais sur la base du capital total. De plus, il y a dans la réalité une tendance vers l'égalisation des taux d'intérêt à travers l'économie entière. Ces faits empiriques soulèvent une difficulté. En effet, à la lumière de la théorie marxiste, il devrait y avoir des taux d'intérêt *différents* dans chaque branche de l'industrie, en fonction du poids relatif du capital variable. Mais en réalité on observe un seul et même taux d'intérêt.

Marx ébauche une solution à ce problème dans le troisième et dernier tome de son chef-d'œuvre *Le Capital*, publié en 1894 à titre posthume sous la direction de son ami Friedrich Engels. Il y conclut que le taux d'intérêt unique observable sur le marché est un « taux moyen de profit » ; celui-ci *dérive* des taux de plus-value, mais il en dérive de manière *indirecte*, à savoir par la concurrence des entrepreneurs qui font l'arbitrage entre ces différents taux. Marx précise que la somme totale des plus-values obtenues dans les entreprises de la société entière est égale à la somme totale des profits. Il admet que cette égalité s'obtient *seulement du point de vue de ces agrégats* ; dans la réalité concrète, c'est-à-dire dans chacune des entreprises individuelles, il y a divergence entre plus-value et profit moyen. Et à la grande consternation de ses disciples, Marx considère que c'est bien le taux de profit moyen, et non pas le taux de plus-value, qui détermine tous les prix concrets.

LÉON WALRAS
(1834-1910)

La mathématique : reine de l'économie

Au début des années 1870, trois économistes donnent à l'analyse des prix un fondement complètement nouveau. Carl Menger en Autriche, William Jevons en Angleterre et Léon Walras en Suisse démontrent que les prix de *tous* les biens sont déterminés par les choix des consommateurs ; et que ces derniers sont à leur tour déterminés par la *loi de l'utilité marginale décroissante* : l'utilité d'une unité d'un stock de biens homogènes diminue lorsque la taille totale du stock croît.

Jevons et Walras privilégient les mathématiques pour exprimer leur découverte. Les mathématiques et la physique sont au 19e siècle les symboles du progrès humain. Grâce à elles, on a su confectionner les prodigieux appareils industriels. Ne faudrait-il pas les appliquer pour construire une société plus juste, plus efficace ? N'a-t-on pas besoin d'« ingénieurs sociaux » ?

Walras est le premier qui, dans ses *Éléments d'économie politique pure* (1874), présente un modèle mathématique du système économique entier. Le mot « pure » exprime l'hypothèse que la détermination des prix se fait « sous un régime hypothétique de libre concurrence pure ». Mais Walras précise que cette dernière « n'est pas le seul moyen de ramener le prix de vente au niveau du prix de revient… et l'économie appliquée doit se demander si elle est toujours le meilleur ».

Son premier défi est d'aller au-delà des modèles de marchés particuliers pour concevoir en termes mathématiques les interdépendances entre *tous* les secteurs de l'économie. Ce modèle général devrait déterminer les prix et les quantités de chaque bien – les services, les produits et les capitaux neufs (l'épargne) – en fonction les uns des autres. Walras démontre que ce problème peut être résolu car *le nombre des équations* qui décrivent l'offre et la demande de chaque bien en fonction de son prix *est égal à celui des variables inconnues*.

Ensuite, il s'agit de montrer que l'équilibre général entre toutes les variables représente un état *optimal*. Walras fonde son argument sur l'idée que le besoin qu'un agent éprouve pour un type de bien X dépend des *quantités partielles* de X que cet agent possède déjà. Le prix qu'il consent à payer pour une unité supplémentaire dépend de l'intensité du *dernier* besoin satisfait, l'intensité du besoin « marginal » *diminuant* avec la quantité disponible.

En équilibre général, les prix de tous les biens se proportionnent aux derniers besoins satisfaits *dans exactement la même proportion*. Si alors une tomate coûte 1 € et un feutre 2 €, le dernier besoin satisfait par le feutre est exactement deux fois plus intense que le dernier besoin satisfait par la tomate. Toute déviation de cet état d'équilibre implique dès lors une diminution de la somme des satisfactions. Si par exemple la quantité de feutres est plus grande qu'en équilibre général et celle des tomates plus petite, les besoins supplémentaires satisfaits par les feutres sont *moins* intenses que ceux qui ne sont plus satisfaits par les tomates manquantes. D'où il s'ensuit que l'équilibre général représente un optimum social.

Walras insiste sur le fait que ces résultats sont uniquement obtenus grâce à la mathématique :

> *« La mathématique seule peut nous apprendre la condition du maximum d'utilité. […] Et la mathématique seule peut nous apprendre pourquoi et comment, non seulement dans l'échange, mais dans la production, la capitalisation et la circulation, on arrive à des prix courants d'équilibre. »*

EUGEN VON BÖHM-BAWERK
(1851-1914)

Biens présents et biens futurs

Carl Menger rejette l'approche mathématique de Walras car celle-ci concerne les seules interdépendances entre les variables économiques, et non pas les rapports *causaux* entre eux. De plus, elle se fonde sur des stipulations fictives telles que la parfaite divisibilité de tous les biens ou la mesurabilité de l'utilité. Elle est donc moins exacte que la « méthode empirique » – l'analyse descriptive des relations entre les besoins humains et les valeurs des biens.

Cependant, Menger n'applique pas sa méthode pour étudier le fonctionnement d'une économie de marché dans son ensemble. Cette lacune est comblée par Eugen von Böhm-Bawerk qui, dans les années 1880, présente une théorie de la production fondée sur la loi de l'utilité marginale décroissante.

Dans les deux tomes de son chef-d'œuvre *Capital et intérêt* (1884, 1889), Böhm-Bawerk présente d'abord une théorie générale des coûts qui inclut la théorie classique – selon laquelle ces coûts déterminent les prix des produits – comme un cas particulier. Il démontre que les prix des facteurs (les coûts de production) résultent des prix payés pour leurs produits les moins bien rémunérés (les produits marginaux). Par conséquent, l'utilité marginale des biens de consommation est effectivement la cause *ultime* des coûts.

Or, dans le cas des facteurs de production *reproductibles*, la production de chacun de leurs produits est poussée jusqu'à ce que la rémunération marginale soit la même pour tous ces produits, cette rémunération étant encore égale à celle du produit marginal. Il s'ensuit que, si la rémunération du produit marginal augmente, alors non seulement il y aura d'abord une même augmentation des coûts, mais ensuite aussi une même augmentation des prix de tous les autres produits. Par conséquent, dans ce cas particulier, les prix de

ces autres produits sont *immédiatement* déterminés par les coûts des facteurs, bien que déterminés de manière *ultime* par le prix du produit marginal.

Tous les projets de production qui utilisent un même facteur sont alors liés en tant que causes et effets par la « loi des coûts ». Mais il y a un mécanisme causal encore plus fondamental qui englobe toutes les quantités et tous les prix. Ce mécanisme résulte du fait que tous les projets de production nécessitent du travail et que les travailleurs ne peuvent pas attendre la fin de la production pour consommer les fruits futurs de leur travail. Ils consomment *déjà pendant la production*. D'où il résulte que chaque projet de production est en concurrence avec tous les autres pour être alimenté par le fonds limité des biens de consommation *présents*.

Selon Böhm-Bawerk la distinction entre biens présents et biens futurs donne également la clef de compréhension des taux d'intérêt. Ceux-ci sont des prix payés dans les échanges intertemporels, donc des taux d'échange entre biens présents et biens futurs. Par exemple, si l'on prête 100 € maintenant en échange de 110 € dans un an, alors on échange effectivement une somme présente contre une autre somme future, à un taux d'intérêt de 10 %.

Les taux d'intérêt sont positifs parce que les biens présents valent en général plus que les biens futurs. Une pomme présente vaut plus qu'une pomme future. Cette préférence pour le présent ou « agio » est la source de l'intérêt monétaire.

Böhm-Bawerk insiste cependant sur le fait que le marché des crédits représente juste *un* des segments, d'ailleurs le moins important, du marché général où s'échangent biens présents contre biens futurs. Le segment le plus important est le marché des facteurs de production. En effet, *tous* les facteurs sont de véritables biens futurs qui, par la production, sont transformés en biens présents, et leurs prix sont nécessairement inférieurs aux prix de leurs produits. D'où il s'ensuit qu'un « taux d'intérêt originaire » se forme au sein même des firmes.

IRVING FISHER
(1867-1947)

L'idéal de la stabilisation

La théorie de Böhm-Bawerk est particulièrement bien reçue aux États-Unis, où plusieurs économistes tels que Frank Fetter, Frank Taussig, Herbert Davenport et Irving Fisher se mettent à la développer.

Économiste mathématicien, Fisher cherche non seulement à modéliser la théorie du capital et de l'intérêt, mais aussi à la généraliser par la prise en compte de l'influence de la monnaie. Dans sa *Théorie de l'intérêt* (1930), il analyse le choix d'investissement des firmes comme un problème d'optimisation intertemporelle.

Fisher considère que chaque firme choisit entre des projets d'investissements plus longs, qui vont générer des revenus dans un futur plus lointain T2, et des projets plus courts, avec des revenus dans un futur plus proche T1. Ce choix est soumis à la loi des rendements décroissants : chaque unité supplémentaire de revenu en T1 implique de renoncer à des revenus de plus en plus grands en T2, et inversement. Le *rapport* entre les deux revenus en T2 et T1 est un taux de rendement intertemporel. Plus on investit dans des projets courts, plus ce taux est élevé.

Comment ce choix devrait-il se faire alors ? L'entrepreneur pourrait le faire en fonction de sa propre épargne et de ses préférences personnelles, mais Fisher montre que cela ne serait pas optimal. Une solution optimale ne saurait être définie que par rapport aux taux de rendement intertemporels des *autres* entreprises, c'est-à-dire par rapport au marché. Il se peut que le taux de rendement réalisable dans une firme X soit supérieur au taux réalisable partout ailleurs dans l'économie. Aussi X devrait-elle investir au-delà de sa propre épargne et financer le solde par un crédit. À l'inverse, si l'investissement dans X portait un rendement inférieur par rapport au reste du marché, alors X devrait diminuer son investissement et prêter cette somme au marché.

Ce raisonnement amène Fisher à deux « théorèmes de séparation » : (1) le choix d'investissement optimal d'un entrepreneur est indépendant de ses préférences personnelles quant à l'épargne ; (2) il est également indépendant de la manière dont son investissement est financé (fonds propres ou dettes). En équilibre général, tous les projets d'investissement réalisent le même taux de rendement, et ce taux est effectivement le taux d'intérêt.

Fisher distingue soigneusement entre le taux d'intérêt *nominal* (monétaire) et le taux d'intérêt *réel*. Tous les paiements se font en monnaie, donc les taux effectivement payés sont des taux nominaux. Mais ces derniers *coïncident* avec les taux réels uniquement lorsque le pouvoir d'achat de la monnaie ne varie pas. Dans tous les autres cas, il faut *calculer* le taux réel (r) en déduisant du taux nominal (i), le taux de variation du pouvoir d'achat de la monnaie (π), selon l'« équation de Fisher » :

$$r = i - \pi.$$

Or, les variations du pouvoir d'achat de l'unité monétaire sont difficilement prévisibles. Le danger est grand de méprendre une hausse ou une baisse du taux *nominal* pour une hausse ou une baisse *réelle* (« illusion monétaire »), ce qui falsifierait les calculs d'investissement et entraînerait des gaspillages. Fisher recommande par conséquent une politique de stabilisation du pouvoir d'achat de la monnaie. D'un point de vue technique, cela présuppose qu'on puisse modifier librement l'offre globale de la monnaie. Il est nécessaire d'abandonner l'étalon-or pour adopter une monnaie purement fiduciaire et d'empêcher toute création monétaire par les banques.

LUDWIG VON MISES
(1881-1973)

Monnaie et civilisation

Böhm-Bawerk intègre le passage du *temps* dans l'analyse du fonctionnement de l'économie dans son ensemble, mais il néglige à intégrer également la monnaie. Comme la plupart des économistes classiques, il pense que la monnaie est en fin de compte neutre. Elle affecte le *niveau* des prix, mais non pas les *relations* entre les prix, et seules ces dernières comptent véritablement.

Le premier qui réussit à pleinement intégrer la monnaie dans la théorie de l'utilité marginale est Ludwig von Mises. Dans sa *Théorie de la monnaie et des moyens de circulation* (1912), Mises crée une synthèse entre l'analyse monétaire et l'analyse réelle qui démontre que la monnaie n'est jamais neutre. Les variations de l'offre et de la demande de monnaie ont un impact durable sur la distribution des revenus et sur la production. Une création monétaire excessive peut même causer des crises économiques car elle tend à baisser artificiellement le taux d'intérêt, ce qui crée des déséquilibres intertemporels.

Mises se tourne après vers l'analyse du socialisme, qui monte en puissance à l'issue de Première Guerre mondiale lorsque des gouvernements socialistes prennent la relève en Russie, en Autriche, en Hongrie et en Allemagne. Il étudie de près la planification centrale en pratique : le « socialisme de guerre » introduit en 1916 en Allemagne et en Autriche, un échec lamentable. Mais comment expliquer cet échec ? Mises est impressionné par le dévouement des Allemands au bien commun qui les rallie derrière leur gouvernement. Si le socialisme ne fonctionne pas avec une population si désintéressée, il doit exister encore un autre problème qui n'a rien à voir avec les motivations des agents.

En 1920, Mises propose alors une autre explication : *il est impossible d'effectuer des calculs économiques dans une société socialiste*. Le calcul économique des entreprises capitalistes est fondé sur les prix de marché.

L'on peut comparer les taux de rendement de différents projets d'investissement sur la base de leurs coûts et recettes monétaires. Mais les prix de marché résultent de l'échange. Ils n'existent que quand le bien en question peut changer de propriétaire. Or, en régime socialiste, il n'y a, par définition, qu'*un seul et unique* propriétaire de tous les *facteurs de production* : la société, représentée par le comité de planification. Par conséquent, il ne peut pas exister de prix pour les facteurs de production. Impossible d'effectuer des calculs tels qu'ils s'effectuent dans les régimes capitalistes.

Le socialisme est alors exactement le contraire de ce qu'il prétend être. Loin de permettre une direction rationnelle des investissements, il ne peut s'empêcher de gaspiller les ressources disponibles et d'appauvrir la population car il doit faire ses choix sans la boussole du calcul économique. Il ne *peut* pas fonctionner, même pas dans une société qui y est dévouée.

Mises élabore et généralise ces réflexions dans *Le Socialisme* (1922). Il montre notamment que le socialisme se marie mal avec la démocratie et qu'il n'est pas *a priori* plus moral que le capitalisme. Il n'y a pas non plus de tendance du marché libre à devenir monopolistique, ni d'évolution inéluctable vers le socialisme.

Enfin, dans son chef-d'œuvre *L'Action humaine* (1949), Mises se met à comparer le capitalisme, le socialisme et l'interventionnisme à la lumière d'une théorie générale du calcul économique. Il montre que, contrairement à ce qu'ont pensé les économistes depuis Adam Smith, le calcul monétaire n'est pas juste *une* forme du calcul économique qui, en principe, pourrait se faire également sous d'autres formes. C'est *le* calcul économique tout court. La rationalité et ses fruits sont logiquement et historiquement contingents. Seul le capitalisme admet une division du travail étendue et tout ce qui en dépend : un développement des arts et des sciences, et une grande civilisation.

JOHN MAYNARD KEYNES
(1883-1946)

L'économie politique du risque

Les idées des économistes classiques perdent en influence à partir de la fin du 19e siècle. Seule une minorité d'économistes professionnels se réclame encore de cet héritage dans les années 1920. Les universitaires ne croient pas au laissez-faire, mais il leur manque néanmoins une théorie générale qui pourrait se substituer à celle des classiques.

John Maynard Keynes comble cette lacune avec sa *Théorie générale de l'emploi, de l'intérêt et de la monnaie* (1936), où il se détourne des idées classiques à la fois sur le plan de l'analyse et celui des politiques. Tandis que le raisonnement classique écarte autant que possible la considération des dépenses monétaires, Keynes les met au centre de son analyse. Au laissez-faire classique, il substitue un agenda bien étoffé d'interventions de l'État. Celui-ci doit remplacer l'étalon-or par la gestion publique d'un papier-monnaie ; il doit être prêt à dépenser des sommes considérables importantes dans des travaux publics pour neutraliser les réticences des investisseurs ; et il doit redistribuer les revenus en faveur des ménages qui ont une plus grande « propension marginale à consommer ».

Keynes insiste sur le fait que le laissez-faire ne repose sur aucun fondement scientifique. Il s'appuie, selon lui, uniquement sur des postulats arbitraires d'ordre philosophique (droit naturel) et théologique (harmonie des intérêts). Sa validité n'a pas non plus été démontrée par la théorie classique car celle-ci n'est pas *générale*. Il n'est pas toujours vrai, comme le veut la loi de Say, que les hausses de la demande soient provoquées par des hausses précédentes de l'offre. Selon Keynes, il est parfois possible d'inverser cette causalité : une demande globale supplémentaire peut entraîner un surcroît de production.

Keynes rejette également les nouvelles théories mathématiques censées démontrer qu'une « concurrence parfaite » crée un optimum social. Il

souligne que les prémisses de ces théories sont largement fictives : divisibilité parfaite de tous les biens, connaissance parfaite de l'avenir, ajustements instantanés, unités de production infiniment petites, etc.

Le laissez-faire est loin d'être une politique rationnelle. Mais quel est le fondement scientifique de l'interventionnisme préconisé par Keynes ? Son point de départ est une réflexion sur l'ignorance humaine. Les agents sont ignorants de l'avenir, qui va pourtant valider leurs choix présents. Pour s'immuniser contre les risques associés à l'investissement, ils demandent alors des primes de risque qui s'ajoutent aux coûts de production et qui, selon Keynes, en représentent peut-être la part la plus lourde.

La théorie des probabilités est impuissante devant le problème de l'ignorance de l'avenir car elle s'applique uniquement aux cas de risque quantifiable, alors que l'incertitude n'est justement pas quantifiable.

Ainsi le marché maîtrise-t-il mal ce problème. Dans un environnement instable – fluctuations fortes du niveau des prix, crises économiques –, les entrepreneurs refusent de prendre des décisions d'investissement. Résultat : une baisse de l'activité et donc une baisse du produit national. Et Keynes de conclure : « Les plus grands maux de notre temps sont les fruits du risque, de l'incertitude et de l'ignorance. » Il recommande alors un programme politique qui s'articule autour de quatre piliers : contrôle direct de la monnaie et du crédit par une institution centrale ; collection et dissémination d'informations sur l'économie ; contrôle complet de l'épargne et de l'investissement ; contrôle de la population en qualité et en nombre.

JOSEPH SCHUMPETER
(1883-1950)

La destruction créatrice

Les équations de Walras décrivent une économie statique en équilibre général. Pourtant, dans la vie réelle, l'économie est un processus en transformation permanente. Elle se caractérise en plus par la présence de profits et de pertes, de chômage et de faillites d'entreprises, etc. – autrement dit par des erreurs de choix qui n'existeraient pas en équilibre général. Comment expliquer cela ? Quel est le facteur dynamique qui crée des déséquilibres ?

Selon Joseph Schumpeter, ce facteur est l'entrepreneur. Dans sa *Théorie de l'évolution économique* (1911), Schumpeter présente l'entrepreneur comme un innovateur qui, par ses choix originaux, détruit la routine des simples directeurs d'entreprise qui, eux, ne cherchent qu'à s'ajuster à leur environnement. L'innovation concerne notamment les produits, les techniques de production, l'organisation du travail, les matières premières et les débouchés.

L'entrepreneur n'est pas forcément un capitaliste, et Schumpeter prend soin de ne pas confondre ces deux rôles. Toutefois, l'innovateur a besoin de capital pour mettre ses idées en pratique. D'où viennent alors ces fonds, si ce n'est des capitalistes ? Schumpeter invoque dans ce contexte la création monétaire par les banques. Le banquier crée le « pouvoir d'achat » qui permet à son client entrepreneur de réaliser son projet. Si l'innovation porte succès, l'entrepreneur gagne un profit et le banquier gagne un revenu d'intérêt. En cas d'échec, les deux partenaires perdent leur mise.

Schumpeter insiste sur le fait que ces types de revenu – intérêt, profits et pertes – sont des revenus *exclusivement dynamiques*. Ils n'existeraient pas en équilibre général. Cela implique à son tour que les déséquilibres sont le prix à payer pour la croissance. En effet, la croissance résulte essentiellement de l'accumulation du capital, voire de l'épargne. Mais

cette épargne dépend, entre autres, de la rémunération que l'épargnant peut espérer. Dans un contexte statique où les revenus d'intérêt n'existent pas, elle est nécessairement plus faible que dans un contexte dynamique. La croissance économique est par sa nature « cahoteuse », entraînant des faillites d'entreprises sur le plan microéconomique et des crises sur le plan macroéconomique. Impossible de séparer ces aspects, le progrès matériel va de pair avec les perturbations.

Dans *Capitalisme, socialisme et démocratie* (1942), Schumpeter élabore sa théorie du marché d'un point de vue plus large. Il souligne que la concurrence au sein d'une économie de marché ne ressemble pas à une situation stable, mais qu'elle est bien au contraire « un processus d'évolution », ou bien aussi « une méthode de transformation économique » qui ne pourrait jamais être stationnaire. L'innovation permanente par les entrepreneurs crée, encore et toujours, de nouveaux modes de coopération avec lesquels les anciennes entreprises ne peuvent plus concourir. « Ce processus de *Destruction Créatrice* constitue la donnée fondamentale du capitalisme ; c'est en elle que consiste, en dernière analyse, le capitalisme, et toute entreprise capitaliste doit, bon gré mal gré, s'y adapter. »

Propulsée par l'innovation, la concurrence met en jeu *tous* les entrepreneurs, quelle que soit leur branche d'activité. Elle ne concerne pas simplement le prix d'une marchandise identique produite avec des méthodes identiques. Elle « s'attaque, non pas seulement aux marges bénéficiaires et aux productions marginales des firmes existantes, mais bien à leurs fondements et à leur existence même ».

FRANK KNIGHT
(1885-1972)

La concurrence parfaite et son contraire

En précisant que la concurrence est un processus intégrant toutes les firmes, quel que soit leur secteur, Schumpeter s'oppose à la théorie de la « concurrence pure et parfaite », qui apparaît entre 1874 et 1921. Fondée sur les conceptions de Cournot et de Walras, cette théorie « néoclassique » reçoit sa forme canonique dans la pensée de l'économiste américain Frank Knight.

Dans son chef-d'œuvre *Risque, incertitude et profit* (1921), Knight affirme que la théorie économique a toujours été fondée sur l'hypothèse d'une compétition parfaite. Chez les économistes classiques, cette hypothèse « a été partiellement implicite et n'a jamais été formulée de manière adéquate ». Il s'agit alors de la rendre explicite et de détailler ses aspects, si nécessaire, par des abstractions « héroïques », c'est-à-dire entièrement fictives.

Quelles sont alors les caractéristiques de la concurrence parfaite ? Knight les énumère en neuf points fondamentaux, précisant en particulier les hypothèses suivantes :

1. les agents sont complètement rationnels, dans le sens qu'ils connaissent les conséquences objectives de leurs actes et qu'ils se comportent en fonction de ces conséquences (« connaissance parfaite ») ;
2. il y a une « mobilité parfaite » – pas d'obstacles physiques aux changements des plans et à leur exécution. Cela implique également que tous les biens soient parfaitement divisibles, que les activités productrices ne créent pas d'habitudes ou de dispositions psychologiques, et que tout produit soit immédiatement consommé et suivi d'une production nouvelle ;
3. il y a une parfaite et continuelle communication gratuite entre tous les individus. Chaque acheteur connaît les offres de tous les vendeurs, et inversement ;

4. tout individu est complètement indépendant de tous les autres – pas influencé par leurs besoins, préjudices, préférences, répulsions, et libre de toute valeur qui ne se manifeste pas dans les échanges. Toute forme de collusion est exclue et il n'y a pas le moindre monopole, ni la moindre tendance monopolistique.

Clairement, le modèle de la concurrence parfaite n'a rien à voir avec la réalité. Mais selon Knight, il est nécessaire de définir *d'abord* ce modèle pour identifier après le trait caractéristique de l'économie réelle, à savoir la présence de l'incertitude.

Knight distingue rigoureusement deux types de risque : (A) un risque quantifiable et donc assurable ; et (B) un risque non quantifiable et non assurable qu'il appelle « incertitude ». Dans la concurrence parfaite, il n'y a que le risque (A), qui ne dérange pas le postulat de la connaissance parfaite car les primes d'assurance contre ce risque font partie des coûts de production connus par avance. Dans le monde réel, par contre, il existe à la fois le risque (A) et l'incertitude (B), et cette dernière provoque des erreurs d'investissement qui se manifestent dans les profits et les pertes des entreprises.

Dans ses travaux ultérieurs, Knight s'oppose notamment à la théorie de Böhm-Bawerk, qui affirme que la production est un processus étalé dans le temps et fondé sur l'épargne. Selon Knight, il n'est pas nécessaire de « financer » la subsistance des travailleurs engagés dans la production par une épargne faite au préalable. Leur subsistance résulte de la *production courante* de biens de consommation. On n'a pas besoin de faire d'abord ces derniers pour se tourner après à la fabrication des autres biens. Il s'agit plutôt de deux productions *simultanées*.

RONALD COASE
(1910-)

Les coûts de l'organisation

La théorie de la concurrence parfaite détermine les relations entre les prix et les quantités des différents biens, telles qu'elles existeraient en absence de toute erreur de la part des agents. Elle ne dit cependant pas grand-chose sur les organisations sociales, qui jouent pourtant un rôle primordial dans la vie réelle. Pourquoi y a-t-il des organisations comme les entreprises, ou firmes ? Pourquoi y a-t-il des « employés » travaillant pour un « patron » qui vend ensuite leurs produits ? Pourquoi ne se coordonnent-ils pas entre eux par des accords et contrats individuels, afin d'éviter l'intermédiation du patron ?

La raison principale est, nous dit l'économiste anglais Ronald Coase dans son article sur « La nature de la firme » (1937), qu'il est *coûteux* de se coordonner. D'abord, il faut identifier des partenaires convenables ; puis chacun doit se mettre d'accord avec tous les autres (pour 5 associés, cela impliquerait 10 négociations ; pour 10 associés : 45 négociations ; pour 15 associés : 105 ; et ainsi de suite) ; et finalement l'on doit veiller sur l'exécution des contrats et imposer leur application par la force, si nécessaire. Les activités qui préparent, définissent et appliquent les contrats représentent selon Coase des coûts d'un type particulier, à savoir les *coûts de transaction*.

Ces coûts seraient nuls dans une concurrence pure et parfaite, caractérisée par la parfaite information de tous les agents. Dans ce cas, il ne pourrait y avoir ni de firmes, ni d'autres organisations. Les individus n'auraient pas besoin d'intermédiaires pour se concerter et achever les projets les plus complexes.

Dans la vie réelle, cependant, les coûts de transaction ne sont pas nuls. Ils sont même prohibitifs pour ceux qui chercheraient à s'organiser sans intermédiaires, même lorsqu'il s'agit des projets de petite et moyenne tailles. Aussi y a-t-il un rôle à jouer par les organisations.

Par exemple, les firmes permettent d'économiser ces coûts parce que : (1) le *nombre de négociations* est drastiquement réduit (pour un patron avec quatre employés : quatre négociations, avec neuf employés : neuf négociations, etc.) ; et que (2) il est dès lors possible de profiter de la *spécialisation* dans la préparation, définition et application des contrats.

Dans un article célèbre sur « Le problème des coûts sociaux » (1960), Coase applique cette théorie au cas des conflits juridiques. D'un point de vue économique, ces conflits concernent en règle générale les droits de propriété, comme dans l'exemple du blé d'un fermier qui peut être endommagé par les étincelles d'un train qui passe. Le fermier avait-il le droit de planter son blé sans nuisances ? Ou est-ce que le propriétaire de la ligne ferroviaire avait le droit d'opérer sa ligne et d'émettre des étincelles à gauche et à droite ?

Coase insiste d'abord sur le fait que ce problème n'existerait pas dans un monde sans coûts de transaction, car dans ce monde, chaque ressource serait éventuellement acquise par l'individu qui y attache la plus grande valeur. Mais dans le monde réel, cette affectation n'a souvent pas lieu en raison desdits coûts. L'économie réelle est selon Coase « inefficiente » par rapport au monde sans coûts de transaction. Son inefficience peut être réduite, comme on l'a vu, par les firmes ; mais les firmes jouent ce rôle uniquement tant qu'elles réalisent des profits. Dans tous les autres cas, on a besoin de l'État. Celui-ci devrait affecter les ressources à *ceux qui les auraient acquises* en l'absence des coûts de transaction. Cependant les interventions de l'État entraînent aussi nécessairement des coûts, même lorsqu'elles réussissent à corriger certaines défaillances du marché et donc à créer des gains sociaux. Il convient alors que l'État n'agisse que si les coûts de son intervention sont inférieurs aux gains attendus. Avec ce raisonnement original, Coase inaugure la sous-discipline de l'« économie du droit ».

MILTON FRIEDMAN
(1912-2006)

La monnaie réhabilitée

Dans la théorie keynésienne, les dépenses globales à la consommation sont la cause la plus importante de la croissance. Cependant, pour chaque individu, la « propension marginale à consommer » diminue lorsque son revenu augmente. Pour surmonter cette barrière à la croissance, les keynésiens préconisent un programme interventionniste : augmentation des dépenses de l'État et redistribution des revenus des ménages riches (qui ont une faible propension à consommer) vers des ménages pauvres (qui ont une plus forte propension à consommer).

Cette théorie se heurte cependant à un paradoxe empirique. Pour chaque période donnée, les propensions à consommer varient en fonction des fourchettes de revenus, les ménages riches consommant relativement moins que les ménages pauvres. Mais si l'on regarde la consommation des ménages individuels *à travers le temps*, on constate que leur propension à consommer varie assez peu, bien que leur revenu augmente et diminue.

Cela s'explique, selon l'économiste américain Milton Friedman, par le fait que les dépenses à la consommation se font en fonction de ce que les gens pensent être leur revenu à long terme – leur « revenu permanent » – et non pas en fonction des oscillations transitoires de ce revenu. Friedman modélise cette hypothèse sous forme d'une fonction de consommation et montre que celle-ci résout effectivement ledit paradoxe.

Il en tire deux conclusions. Premièrement, le capitalisme ne souffre pas après tout d'une tendance vers la stagnation pour cause de sous-consommation. Deuxièmement, il est difficile pour la politique budgétaire de l'État de corriger les choix des consommateurs à court terme, car ces derniers ne vont pas tenir compte des variations temporaires de leur revenu.

Friedman s'oppose également à un autre élément de l'orthodoxie de son temps : l'idée que la masse monétaire soit une variable entièrement *dépendante*, qui suit les variations d'autres variables telles que les coûts de production ou le déficit budgétaire du gouvernement. Friedman conçoit la monnaie comme étant une véritable marchandise que les individus détiennent dans leurs portefeuilles. Parfois la masse monétaire réagit aux variations des autres variables, mais dans un grand nombre de cas elle est un élément actif, et alors les augmentations de son offre provoquent des augmentations des revenus nominaux et des prix. Dans son *Histoire monétaire des États-Unis, 1867-1960* (ouvrage coécrit avec Anna Schwartz et publié en 1963), Friedman montre que les variations de la masse monétaire *mènent* les variations des autres variables : au bout de neuf mois environ, elles sont suivies par des augmentations de la production et des revenus. Mais ces effets ne sont pas durables. La seule conséquence permanente est la hausse du niveau des prix.

Puisque la monnaie n'a pas d'impact durable sur la croissance, Friedman ne voit aucun sens dans une politique monétaire discrétionnaire. On peut se passer d'autorités monétaires qui, par leurs choix, modifient la masse monétaire. Il préconise alors de mettre à leur place un système automatique, fondé sur une seule règle : la masse monétaire doit être augmentée chaque année par un même pourcentage fixé d'avance, par exemple 2 ou 3 %.

La « politique » monétaire est réduite au seul choix d'une règle fixe. Dans les années 1980, Friedman propose de la réduire encore davantage. Il est parvenu à la conclusion que les avantages des systèmes métalliques – tels que l'étalon-or – ont été sous-estimés par les économistes du 20e siècle, et qu'il ne serait pas déraisonnable de réfléchir aux possibilités de se passer complètement du contrôle public de la monnaie.

GARY BECKER
(1930-)

L'économie au-delà du marché

Avec la « révolution marginaliste » déclenchée par les travaux de Menger, Jevons et Walras, le domaine de l'analyse économique s'est considérablement élargi. Cette dernière n'est plus simplement la science de la richesse matérielle. Elle est dès lors la science des choix individuels, tandis que l'analyse des causes de la richesse est seulement une de ses applications.

Les économistes se sont vite rendu compte de cette nouvelle perspective et des nouveaux défis qu'elle implique. Au début du 20e siècle, on voit la naissance d'une nouvelle espèce de chercheurs : les économistes-sociologues. Ludwig von Mises est le premier à esquisser les contours d'une science générale de l'action humaine. D'autres économistes suivent la même orientation avec des outils et des interrogations différentes.

Le plus célèbre économiste-sociologue de notre temps, Gary Becker, se penche d'abord sur les mécanismes économiques de la discrimination. Cette dernière se manifeste notamment dans le traitement des agents minoritaires lorsqu'ils obtiennent des salaires inférieurs à leurs produits marginaux. Becker insiste sur le fait que la discrimination crée des opportunités de profits pour d'*autres* firmes. Si par exemple un patron raciste paye des salaires trop bas à ses employés noirs, alors un de ses concurrents pourrait les embaucher à des salaires plus élevés, tout en réalisant encore un profit. D'où il s'ensuit que la concurrence diminue la discrimination. Dans ses études empiriques, Becker constate que la discrimination ne joue un rôle vérifiable que dans les cas de faible concurrence, notamment lorsque les minorités représentent un très faible pourcentage de la population.

Becker se consacre après à l'analyse du « capital humain », le produit de la formation et de l'éducation en particulier, en examinant par

exemple le rôle des habitudes de consommation ainsi que l'impact d'une formation générale sur les avancements de carrière.

Puis, il développe une analyse du crime fondée sur l'hypothèse que les criminels n'agissent pas par impulsion, mais par choix bien considéré. Par exemple, un voleur compare l'utilité de son butin avec les coûts qui résulteraient en cas de détection, pondérés par la probabilité de la détection. S'il trouve que l'utilité subjective de son crime est supérieure aux coûts respectifs, il se décidera tout à fait rationnellement en faveur du crime.

Pour diminuer la criminalité, le législateur doit agir sur les paramètres principaux. Il peut augmenter les coûts du crime (amendes, longueur et conditions des séjours en prison…) et la probabilité de détection (patrouilles plus fréquentes, caméras de surveillance…). Becker précise cependant qu'il ne serait pas efficace de supprimer le crime complètement. La lutte contre le crime est elle-même coûteuse. On peut toujours diminuer les infractions par l'emploi de moyens de plus en plus importants, mais, à partir d'un certain seuil, le gain en crimes évités devient inférieur aux coûts mis en œuvre. Il y a un taux de criminalité « socialement optimal ».

Le chef-d'œuvre de Becker est son *Traité de la famille* (1981) où il utilise le raisonnement économique pour expliquer le temps alloué aux différentes tâches de la « production ménagère » (élever des enfants, préparer des repas, nettoyer la maison et les vêtements…), le choix des partenaires sur le « marché du mariage », le taux de fertilité des femmes et le taux de divorce. Quant à ce dernier, par exemple, Becker affirme que les couples aisés tendent à divorcer moins fréquemment car les coûts en sont plus élevés.