

CHAPITRE I

Les Institutions en économie de la santé : élaboration d'un cadre théorique spécifique à l'Amérique latine

L'économie des institutions est le lieu d'un âpre débat théorique. Ce premier chapitre a pour but d'éclaircir les lignes de fracture entre les différents corpus théoriques, poser les bases conceptuelles de la thèse et construire un cadre analytique original qui permettra, dans les chapitres suivants, d'évaluer les systèmes de santé de deux pays en développement – la Colombie et le Venezuela – et, plus précisément, leur capacité à résoudre les problèmes inhérents à l'extension de l'accès aux soins dans les catégories les plus défavorisées de la population.

Tout d'abord, il semble important de souligner qu'un système de santé est un agrégat complexe d'acteurs dont la finalité - garantir à tout individu une prévention et des soins efficaces - est l'un des objectifs majeurs de tout Etat de droit. La santé est, avec l'éducation et le logement, la raison d'être des Etats et c'est bien de ces fondements qu'ils tirent leur légitimité auprès des individus.

Ainsi, de par son caractère central et décisif dans la politique publique des Etats, l'analyse théorique des systèmes de santé se rattache inéluctablement à l'analyse des **institutions**, de leur architecture, de leur dynamique, de leur soutenabilité et de leur impact.

En effet, l'institution comme objet d'étude à part entière permet : 1) de sortir de l'écueil de la lecture strictement néo-classique des interactions entre les agents d'un système de soins - interactions dont le concept de "rationalité" est le cœur - et 2) de rendre pleinement compte de la complexité des problématiques posées par les systèmes de santé d'un point de vue organisationnel et, dans notre cas précis, des enjeux qui en résultent en termes de « développement ».

De ce point de vue, tout un pan de l'analyse économique, érigé en "opposition", ou du moins en « remise en question » de l'analyse néo-classique, étudie la nature et l'évolution des institutions dans le champ économique d'une société. Nous retenons, en premier lieu, la lecture néo-institutionnelle, bâtie en particulier sur les apports de Douglass North et, surtout, ses prolongements récents, notamment le concept "*d'idéologie*" qui semble pouvoir se transposer aux cas vénézuélien et colombien. Deux frontières théoriques viennent ensuite délimiter notre analyse de l'économie institutionnelle : la théorie des conventions et la théorie de la régulation.

Ce cadre théorique ainsi défini, puis appliqué dans le chapitre 2 à l'analyse de deux systèmes de santé très différents, aura pour but de contribuer à préciser les contours d'une vision institutionnaliste de l'économie de la santé et, plus spécifiquement, dans l'esquisse de catégories particulières aux pays en développement (PED).

En effet, l'analyse des PED requiert, incontestablement, un socle théorique adapté et ce, pour trois raisons principales :

- 1) la dynamique des systèmes institutionnels des PED n'est pas encore **stabilisée**, contrairement aux pays développés qui connaissent, sans aller jusqu'à parler d'achèvement, du moins un équilibre dans le fonctionnement de leurs institutions ;
- 2) une instabilité évolutive des institutions qui est à la fois cause et conséquence de divers facteurs, notamment des soubresauts politiques caractéristiques des sociétés en développement, élément qui ne peut en aucun cas manquer à cette étude ;
- 3) ces variables supplémentaires introduites dans le raisonnement seront les marqueurs de comportements spécifiques à l'échelle des individus, comportements que nous vérifierons en pratique dans une analyse de terrain approfondie.

Dynamique instable, caractéristiques propres et conséquences particulières sur les comportements des agents, ces trois éléments plaident en faveur d'une analyse institutionnelle spécifique des PED. Il s'agira donc, dans un premier temps, de proposer une définition générique des institutions en étudiant les différentes conceptions érigées par les principaux courants de l'économie institutionnelle (Section 1) puis d'étendre l'examen des institutions au champ de l'économie de la santé, dans ses frontières actuelles (Section 2). Enfin, une démarche conceptuelle originale sera proposée, démarche qui sera l'outil utilisé lors des développements ultérieurs de la thèse.

Section 1 Les Institutions entre Idéologie, Convention et Régulation : le point sur la théorie

Le terme « Institution » est défini selon le dictionnaire comme étant une « *norme ou une pratique socialement sanctionnée qui a valeur officielle, légale ; organisme visant à les maintenir* » (Larousse, 2011). En admettant cette seule acception du terme institution, la réflexion se heurte à une limite évidente : l'Institution est-elle une « norme », un « organisme » visant à faire appliquer cette norme ou les deux à la fois ? L'ambiguïté du concept nous amène donc naturellement à en définir plus clairement les contours.

Comment définir l'institution ? La réponse à cette question est primordiale dans la mesure où d'elle dépend l'analyse des architectures institutionnelles et celui de leur impact sur les populations et donc, indirectement, de leur « performance » dans une société donnée. Le marché, lieu de l'interaction entre agents économiques, est l'une des institutions les plus caractérisable et est, de ce fait, l'objet d'étude privilégié de la science économique. Cependant, l'institution ne se limite pas à cette forme d'interaction sociale. En effet, l'analyse sociologique, dont le concept d'institution est central – Durkheim définit même la sociologie comme « science des institutions » (Durkheim, 1975) – apporte une définition plus précise du concept d'institution :

« Toute institution (...) se présente d'abord comme un ensemble de pratiques, de tâches particulières, de rites et de règles de conduite entre des personnes. Mais l'institution est aussi l'ensemble des croyances ou des représentations qui concernent ces pratiques, qui définissent leur signification et qui tendent à justifier leur existence » (Lagroye, 2002).

« *Croyances et représentations* » : dans le champ de l'économie, au-delà des frontières de l'économie standard, plusieurs courants de pensée ont tenté d'intégrer ces notions comme déterminants du comportement humain, théoriquement « rationnel » et « maximisateur », selon le dogme de l'orthodoxie économique. C'est à partir de cette simple remise en question que s'érige le concept d'institution, s'interposant entre

l'individu désormais imparfaitement rationnel et la complexité de son environnement l'appelant systématiquement à faire des *choix*. Pour comprendre les grands principes sur lesquels s'érige la recherche économique en matière d'institutions, il conviendra d'analyser, dans un premier temps, les grands courants institutionnalistes (Section 1) depuis les fondements néo-institutionnalistes (A) jusqu'aux développements plus hétérodoxes (B). La littérature étant très dense, le choix sera fait de porter l'éclairage sur, d'une part, les concepts-clés de chaque courant de pensée et, d'autre part, l'analyse faite par chacun de la *dynamique des institutions*, c'est à dire les processus par lesquels une institution est amenée à se transformer, se développer, améliorer son action voire même étendre son champ d'action.

A. Les bases posées par le courant néo-institutionnaliste ne suffisent pas à expliquer les dynamiques institutionnelles inhérentes à un système

L'institutionnalisme historique de Commons et Veblen (1920) avait pour particularité d'adopter une démarche holiste allant à contre-pied du modèle néo-classique. Ce courant a été occulté, après la seconde guerre mondiale, par le courant béhavioriste, axé sur les comportements observables. C'est au début des années 1980, aux Etats-Unis, que le "néo-institutionnalisme" s'est développé. En réaction aux limites des analyses béhavioristes, il réintroduisit les institutions comme variable explicative autonome. Dès lors, il se distingua de l'institutionnalisme historique par une acception plus complexe des institutions : une prise en compte de la rationalité individuelle, des règles formelles et informelles, des coûts de transaction, des droits de propriété, du déterminant technologique, etc. Cependant, bien que ce courant soit à la base d'une littérature très riche qui permet, entre autre, de mieux saisir l'inertie des dynamiques institutionnelles et qui élabore un cadre interprétatif ouvert aux institutions et à leurs effets, en incluant notamment le concept d'« idéologie », il n'en définit pas moins une approche dynamique des institutions fondée en grande partie sur la complexité des transactions, ce qui, en plus d'être une vision uniquement endogène et a-historique, ne tient pas compte des rapports de pouvoir *entre* les nations.

1. Une dynamique générale des institutions se heurtant au « *path dependence* »

Le dilemme évoqué plus haut, à savoir l'ambiguïté du concept d'institution, est levé par Jütting Johannes (2003) en distinguant trois grandes approches économiques des institutions : la première fondée sur l'opposition entre formel et informel ; la deuxième fondée sur un ordre hiérarchique des institutions ; la troisième fondée sur la distinction de plusieurs « aires » (économie, politique, légale et sociale). L'intérêt de l'analyse Northienne est qu'elle fait la synthèse de ces différentes catégories. Nous suivons en cela Prévost (2010) et partons de la définition généralement admise en économie, c'est à dire celle donnée par Douglass North, l'un des fondateurs de la nouvelle économie institutionnelle :

« Les institutions sont les contraintes établies par les hommes qui structurent les interactions humaines. Elles se composent des contraintes formelles (comme les règles, les lois, les constitutions), de contraintes informelles (comme les normes de comportement, des conventions, des codes de conduite imposés) et des caractéristiques de leur application »
(North, 1994).

Ainsi, la variation dans le caractère plus ou moins formel des règles amène à distinguer les règles informelles, parmi lesquelles les « conventions » seront analysées avec plus de précision ultérieurement, des règles formelles qui se divisent en trois sous-catégories (Masahiko Aoki, 2006) : 1) les règles politiques (lois, constitution, etc.) 2) les règles économiques et 3) les contrats, qui sont intimement liés aux règles du droit de la propriété et sont avant tout des « accords de volonté », par opposition aux conventions qui ne font pas l'objet, la plupart du temps, de choix conscients de la part des individus.

Plus fondamentalement, il apparaît que les institutions sont, pour North, une tentative de la part des hommes d'échapper à l'incertitude qui, pour lui, « *n'est pas une condition exceptionnelle : c'est la condition sous-jacente qui est responsable de la structure de l'organisation humaine tout au long de l'histoire et de la préhistoire* » (North, 2005, p.32). Dans ce monde d'incertitude (ou « non-ergodique »), les êtres humains érigent les

institutions afin de mieux maîtriser leur environnement et le rendre ainsi plus « prévisible ». Mais ces institutions, en tant qu'environnement à part entière, échappent en partie à leur contrôle ce qui conditionne la nature de la croissance économique.

Ainsi, à partir du postulat Northien de division des règles formelles et de leur interaction dans le but de réduire l'incertitude, nous retrouvons, *in fine*, au cœur du débat sur les institutions, la question de la **croissance**, essentielle en économie.

En effet, selon North, les règles économiques « suppléent » l'ordre politique dans les ajustements institutionnels nécessaires au fonctionnement efficace d'une société donnée, c'est à dire à sa croissance économique. A l'instar de Williamson (1985), le lien entre ordre politique et ordre économique est analysé par North au moyen des **coûts de transactions**, concept occupant une position centrale dans les branches de la nouvelle économie institutionnelle (NEI) que ces deux auteurs représentent. En effet, bien qu'utilisant des niveaux d'analyse différents, les deux auteurs s'accordent sur l'importance de la réduction des coûts de transaction dans le développement économique (Chabaud, 2004) : les institutions modifient les coûts de transaction supportés par les agents, facilitant ainsi les échanges et donc la performance économique. En d'autres termes, dans cette vision, une hiérarchie institutionnelle existe parmi les agents qui « font » les règles institutionnelles et, entre eux, une interaction spécifique se met en place pour parvenir à un état de développement optimal. Et dans cette interaction, « *c'est le politique qui définit et fait respecter les droits de propriété* » (North, 1995), ce qui, dans le raisonnement de North, implique la définition de nouvelles explications à la croissance économique, et donc, *in fine*, son éloignement progressif des thèses néoclassiques.

Car au départ, North souhaite compléter la théorie néoclassique qui présente la recherche des opportunités d'enrichissement fondées sur l'accumulation du capital - c'est à dire la recherche de profits - l'investissement dans le progrès technique et l'investissement en capital humain comme seules sources de croissance. North explique que si c'était le cas, l'ensemble des sociétés connaîtraient une croissance équivalente car étant soumises à égalité à cette même tendance des individus à vouloir augmenter leur richesse par ces trois biais. Or, l'un des premiers constats de la science économique - et

qui est pratiquement sa raison d'être - est celui de l'existence de différents niveaux de richesse et de croissance économique entre les pays. Donc logiquement, l'existence de ces différentiels de croissance conduit North à rechercher d'autres moteurs à la croissance économique : il conceptualise ainsi, au cœur des mécanismes économiques qui produisent la croissance, les contraintes dites "informelles" (traditions, codes de conduite, etc.) et les contraintes formelles, c'est à dire les **institutions**.

L'importance du cadre institutionnel, ainsi théorisé par North, se concrétise dans l'interrelation qui se noue entre les structures économiques et les structures politiques : ces dernières seraient plus ou moins performantes à créer un contexte propice à l'échange économique et "harmonieux" pour les agents économiques, agents qui seront, en conséquence, plus ou moins à même de redéfinir à leur tour les contours du cadre institutionnel (*Cf.* Schéma 1.1 : les institutions dans la théorie néo-institutionnaliste). Dans l'esprit de North, le développement des structures institutionnelles et celui des structures économiques sont donc corrélés. En ce sens, son raisonnement apparaît clairement comme étant une tentative **d'endogénéisation des dynamiques institutionnelles**, par opposition aux présupposés néoclassiques selon lesquels les institutions constituent une contrainte « exogène » au cadre rationnel, cadre à l'intérieur duquel les individus adoptent des comportements « optimisateurs ».

Dans le schéma ci-dessous, l'interrelation entre l'ordre politique et l'ordre économique est représentée par les flèches à double sens. Cette interrelation entre les deux ordres définit le contour des règles institutionnelles, représentées ici par le cône rouge. Ces règles institutionnelles impactent le sentier de croissance d'une société donnée représenté ici par le dernier compartiment du cône. Notons que la forme conique du schéma permet de représenter, d'une part, l'idée de hiérarchie que North établit entre l'ordre politique et l'ordre économique, et d'autre part, le fait que l'architecture institutionnelle influence le comportement des agents (flèche rouge) mais repose également sur eux : leur consentement étant l'une des conditions essentielle à la stabilité des institutions.

Schéma 1.1 : Les institutions dans la théorie néo-institutionnaliste

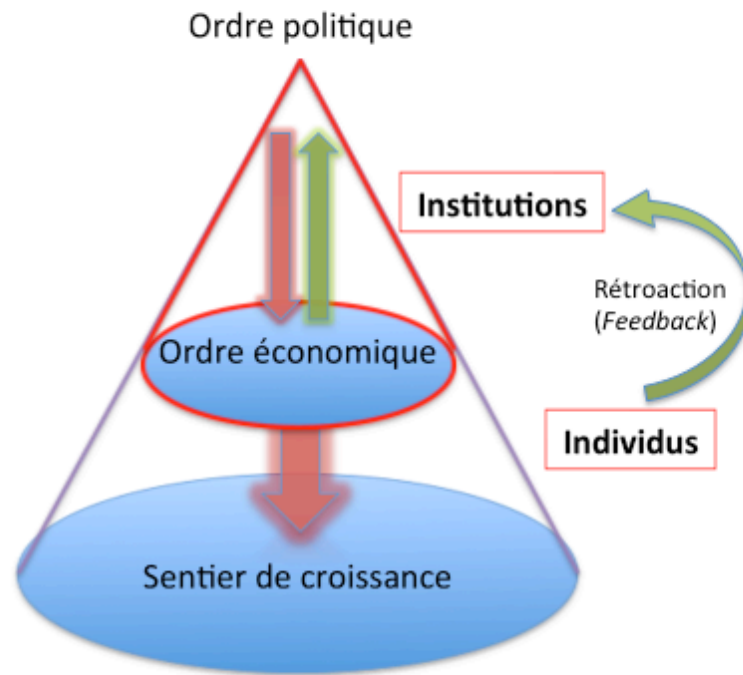


Schéma réalisé par l'auteur à partir de : Douglass North (1990), *Institutions, Institutional Change and Economic Performance*, éd. Cambridge University Press; et : Douglas North (1991) « Institutions », *Journal of Economic Perspectives*;

A cette dynamique institutionnelle, North ajoute le concept de « *path dependence* » (« dépendance du sentier ») initialement introduit dans la littérature économique sur le changement technique par David (1985) et Arthur (1989) et qui caractérise le poids des choix effectués dans le passé sur les décisions présentes. En arrimant sa théorie à ce concept, North s'inscrit clairement dans une approche évolutionniste des institutions, c'est à dire prenant en compte les reconfigurations et adaptations des institutions au contexte historique, sans toutefois éluder le caractère parfois sous-optimal d'un système institutionnel et de son maintien dans le temps. Pour North (1990), cette analyse s'explique, entre autre, par le fait qu'un changement implique cinq effets fondamentaux : 1) une perte de l'amortissement et des rendements croissants des investissements de départ (« *increasing returns* ») ; 2) devoir investir à nouveau ; 3) reprendre le processus d'apprentissage ; 4) risquer de ne plus être coordonné avec les autres institutions et enfin ; 5) il faudrait prévoir les nouveaux comportements, modifier les « anticipations adaptatives ». Cette analyse explique pourquoi les institutions, même inefficaces, sont

stables et résistent au changement : elles sont soumises à des effets d'auto-renforcement, elles n'évoluent que progressivement et même les changements pouvant apparaître comme étant une discontinuité (révolution, etc.) sont interprétés par North comme étant le résultat de l'imbrication (« *embeddedness* ») des contraintes informelles de la société.

Cette idée de résistance au changement dans un système peut également se comprendre, sous l'angle de la psychologie sociale, comme étant une dynamique spécifique aux groupes, comme l'ont théorisé Anzieu et Martin (2000) : « *Toute modification apportée à l'équilibre d'un système entraîne, au sein de celui-ci, l'apparition de phénomènes qui tendent à s'opposer à cette modification et à en annuler les effets* ». Ce concept d'inertie ou, plus précisément, de **non-réversibilité institutionnelle** est également souligné par Pierson (1994), qui explique : « *Les politiques publiques créent des incitations qui encouragent les individus à agir de telles façons que les politiques suivantes se trouvent par la suite enfermées sur une voie de développement* ». Ainsi, bien que les limites de ce concept de « *policy lock-in* » aient été démontrées dans l'analyse des systèmes sociaux européens (Palier, 1999), il semble toutefois permettre d'éclairer la dynamique des institutions dans le cas des systèmes de santé de deux pays en développement tels que la Colombie et le Venezuela.

En effet, il apparaît logiquement que cette non-réversibilité institutionnelle est d'autant plus forte que : 1) la complexité du système est grande, ce qui est le cas des systèmes de santé étant donné leur échelle, le nombre d'acteurs économiques en présence et surtout la complexité de leur interaction ; et 2) les fondamentaux démocratiques d'où sont issues les institutions sont fragiles et instables, ce qui implique que les sphères décisionnelles sont perméables aux pressions des réseaux et de la corruption qui peuvent ainsi exercer une force de résistance aux changements institutionnels. Cette caractéristique définit les contours d'un cadre institutionnel spécifique aux pays en développement, du moins pour la Colombie et le Venezuela que nous décrirons avec précision dans le chapitre 2.

Par ailleurs, le concept de dépendance au sentier a deux conséquences fondamentales sur l'analyse des institutions (Erhel, 2004) : la première, c'est qu'il est impératif

d'analyser les institutions en dynamique, autrement dit, il est nécessaire de prendre en compte leur histoire ; la seconde, c'est qu'un choc identique, par exemple un changement politique violent, n'aura pas nécessairement les mêmes conséquences sur deux architectures institutionnelles différentes. Ces deux propriétés sont particulièrement intéressantes pour étudier la dynamique de modèles nationaux hétérogènes.

Ainsi, le concept de « *path dependence* » semble s'adapter en partie à une analyse du contexte latino-américain et contribuer à expliquer, d'un point de vue institutionnel, les « trappes à sous-développement », c'est à dire le maintien de formes institutionnelles inefficaces consécutives à un effort trop important à fournir pour modifier les structures existantes.

Cette idée de maintien d'institutions parfois inefficaces amène North à s'interroger sur la **rationalité** des individus.

2. Le concept de rationalité limitée : le trait d'union entre l'analyse néo-institutionnelle et la théorie des conventions

Les développements théoriques sur les institutions trouvent leurs racines dans la remise en question du concept d'*homo œconomicus*. En effet, en plaçant ce qu'il appelle le « marché politique » au cœur de sa lecture des dynamiques institutionnelles, North, qui reprend sur ce thème les travaux de Simon, se place à contre-courant des thèses néoclassiques qui, d'une part, évacuent la question des institutions de leur champ d'étude, mais qui surtout se basent sur le concept de rationalité parfaite ou substantive des individus pour appréhender les mécanismes économiques. Alors que les néoclassiques considèrent que l'individu cherche à tout prix à maximiser son profit et qu'il dispose pour cela de toutes les informations nécessaires, Simon, à travers la notion de **rationalité procédurale ou limitée**, et North (1993), par l'introduction du concept « **d'idéologie** », montrent que les motivations des individus sont diverses et qu'elles ne sont pas systématiquement compatibles avec l'idée de maximisation de la fonction d'utilité.

Simon (1991) part du principe que les individus ont une capacité cognitive limitée et que, par conséquent, ils ne peuvent pas saisir les millions d'informations qui composent leur environnement. Dans ce contexte, les individus ont une représentation parcellaire et surtout *subjective* de la « réalité ». Etant donné que l'action d'un individu dépend de la représentation qu'il se fait d'une situation, le fait que cette représentation soit subjective aura deux conséquences principales (Parthenay, 2004) : 1) le milieu organisationnel dans lequel se trouve la personne doit être pris en compte, et 2) il faut tenir compte de l'écart entre l'action et la réalisation des buts. L'approche de la rationalité retenue par Simon observe donc les quatre contraintes fondamentales suivantes :

« Premièrement les acteurs peuvent avoir (et la plupart du temps auront) une information incomplète ou erronée sur la situation et les changements potentiels de la situation au cours du temps (...) Deuxièmement, même si l'information est complète, un acteur peut être incapable (et généralement sera incapable) de calculer toutes les conséquences de l'action (...) Troisièmement, les acteurs n'ont généralement pas qu'un seul but, et il peut y avoir des incompatibilités entre les buts, la réalisation de l'un d'entre eux interférant avec la réalisation des autres (...) Quatrièmement, un acteur peut ne pas parvenir à atteindre un but en raison de son ignorance des moyens d'action. » (Simon, 1991b, p.2).

Pour ces quatre raisons, Simon (1976) va produire le concept de rationalité procédurale sur lequel va se baser par la suite la Nouvelle Economie Institutionnelle pour expliquer l'importance des institutions dans la mise en place d'une coordination plus efficace entre les individus. La rationalité procédurale est une analyse de la rationalité à partir des procédures par lesquelles l'action d'un individu s'exerce. Dès lors et comme le souligne Parthenay (2004), la compréhension d'une action relève d'une interrogation sur le « comment » plutôt que sur le « pourquoi ». En effet, dans une situation où un choix complexe doit être fait, un individu met en place une « *procédure délibérative* », c'est à dire qu'il va concevoir différentes solutions possibles, différentes alternatives au problème qui se pose à lui. L'individu choisira, non pas la solution optimale dans l'absolu, mais la solution qui lui semblera simplement satisfaire ses besoins, son « *niveau*

d'aspiration », en fonction de ses aptitudes. Il s'agit donc d'une rationalité dans laquelle l'agent modifie, au cours de sa délibération, l'objectif qu'il considère comme satisfaisant, en fonction des difficultés rencontrées ou des nouvelles opportunités qui s'offrent à lui.

Autrement dit, dans ce cadre d'analyse et par opposition à la rationalité substantive, la rationalité limitée des individus ne leur permet pas : 1) d'envisager l'ensemble des solutions possibles à un problème en raison de la limitation de ce que Simon appelle les « *capacités computationnelles des agents* » et, 2) parmi le sous-ensemble de solutions qu'ils envisagent, leur choix d'actions possibles est limité par leurs capacités et leur niveau d'aspiration.

Ainsi, étant donné la capacité cognitive limitée des individus, ceux-ci prennent des décisions qui sont rationnelles par rapport aux objectifs qu'ils se fixent. Cette rationalité, même limitée, permet de mettre en évidence les procédures de décision des agents et amène au constat suivant : la rationalité limitée conduit les individus à faire la plupart du temps des choix sous-optimaux. Au final, Simon s'oppose frontalement à *l'homo œconomicus* dans la mesure où, pour lui, dans le comportement d'un agent, l'optimalité est l'exception.

C'est sur ce socle que va s'ériger le questionnement sur l'efficacité d'une organisation, d'une société et, par extension, le rôle fondamental des institutions dans un système, à savoir : **diminuer l'incertitude des agents et améliorer leur coordination**. Ainsi, North (1990), à la suite de Williamson (1975), reprend et approfondit le concept de rationalité limitée développé par Simon qu'il ajoute, avec quelques nuances¹, à la théorie des droits de propriété et des coûts de transactions pour en augmenter la puissance explicative et ainsi rendre mieux compte du rôle des institutions.

En effet, l'essentiel de la théorie des institutions de North (1990) repose sur l'hypothèse de l'information incomplète et des perceptions subjectives des agents. En ce sens, North réaffirme clairement un présupposé individualiste. De ce fait, les concepts d' « *idée* » et

¹ Comme le souligne Claude Parthenay (2004), la rationalité n'apparaît chez Simon que dans les procédures alors que dans la Nouvelle Economie Institutionnelle, et notamment chez North, elle existe dans sa totalité et englobe également l'intention des agents.

d' « *idéologie* » (ou ce que Williamson appelle « *atmosphère* ») acquièrent de l'importance car ils façonnent les constructions mentales utilisées par les individus, ainsi que par les organisations, pour prendre leurs décisions. Pour lui, deux variables essentielles du comportement humain invalident le concept de rationalité parfaite : 1) les motivations des individus : il explique que coexistent, avec le calcul maximisateur, des comportements altruistes et des contraintes que les individus peuvent s'imposer volontairement ou qui s'imposent à eux, comme par exemple les idéologies qui jouent, selon North, un rôle majeur dans le choix des individus, et 2) l'interprétation partielle de l'environnement : il rejoint sur ce point Simon en faisant des limites cognitives des individus une barrière au choix optimal dans un contexte donné. Et c'est précisément la rationalité limitée des agents qui explique que les idéologies occupent une place importante dans le processus de choix.

Par conséquent, les institutions sont essentielles car elles déterminent l'impact de ces idées et idéologies sur les performances économiques. En effet, les structures mentales peuvent s'opposer aux structures économiques et sociales ce qui, dans ce cas précis, constituerait un cadre défavorable au développement économique qui, de surcroît, se reproduirait dans le temps du fait du sentier de dépendance. Ainsi, les institutions sont centrales dans l'analyse car elles déterminent le « *champ des possibles* » et guident ainsi l'action des individus : en ce sens, les institutions sont une condition permissive à la rationalité.

North parachève son raisonnement en combinant ses postulats sur le comportement humain avec la théorie des coûts de transaction, ce qui rend indispensable l'analyse de l'articulation entre les deux théories : « *Ma théorie des institutions est construite sur une théorie du comportement humain combinée à une théorie des coûts de transaction* » (North, 1990, p.27). Chez North, c'est donc bien la combinaison des deux théories qui permet de comprendre pourquoi les institutions existent et pourquoi elles jouent un rôle majeur dans le « bon fonctionnement » de la société. En effet, pour lui, les institutions fournissent la structure des échanges, l'espace des interactions possibles qui déterminent l'ensemble des coûts de production, composés chez North des coûts de transformation et des coûts de transaction.

Les coûts de transaction sont : « *le prix à payer pour mesurer les attributs de ce qui est échangé* », c'est à dire la valeur intrinsèque des biens qui font l'objet d'une transaction, mais ils comprennent aussi les coûts de protection des droits, d'application et de contrôle des accords. Théoriquement, le rôle des institutions est de procurer ces informations et ces garanties afin de limiter l'asymétrie d'information entre les parties qui échangent. En effet, cette asymétrie a un prix et donc un effet négatif sur la croissance : c'est le principe des coûts de transaction. Il est important de noter que, pour North, le niveau des coûts de transaction est déterminé par la complexité des échanges et le degré de spécialisation des agents dans un système économique. Par conséquent, un système dans lequel les acteurs économiques sont hautement spécialisés nécessite des institutions « sûres » afin que les transactions donnent lieu au moins de doute possible pour les agents : dans ce cadre, l'institution tient pleinement son rôle de « réductrice d'incertitude » par la fourniture de contrats, de règles politiques et juridiques parmi lesquelles le **droit de propriété**, droit fondamental en économie de marché que les institutions ont pour rôle capital de définir et de protéger.

Symétriquement, lorsque la spécialisation des acteurs est faible, les transactions étant moins complexes, la nécessité du cadre institutionnel est moins forte. Dans ces sociétés, les rapports humains sont principalement régis par les règles informelles ou des codes sociaux très spécifiques. North (1991) illustre cette analyse par l'exemple du souk qui est caractéristique du lieu d'échange pour lequel il n'existe pas d'institutions qui assurent contre l'asymétrie d'information. Ainsi, il apparaît clairement que cette corrélation qu'établit North entre le niveau de spécialisation des acteurs économique et le niveau des coûts de transaction et donc, par extension, de la nécessité relative d'une architecture institutionnelle performante tient lieu d'éclairage partiel à la question du sous-développement.

Par ailleurs, articulé chez North à la question des coûts de transaction, le **progrès technique** explique également la genèse et donc la raison d'être des institutions. En effet, pour North, les innovations ont engendré de façon « séquentielle » une organisation toujours plus complexe, justifiant l'existence des institutions. Dans son article « *Institutions* » (1991), North identifie trois innovations majeures qui ont contribué à accroître cette complexité : 1) l'augmentation de la mobilité du capital ; 2)

l'abaissement des coûts d'information ; et 3) la possibilité de s'assurer contre le risque avec la transformation progressive de l'incertitude en risque, c'est à dire le passage d'une situation où les événements sont incertains et contre lesquels les agents ne peuvent pas s'assurer, à une situation où, au contraire, les événements sont probabilisables et contre lesquels il est possible de s'assurer.

Par conséquent, pour North (1991), au cours de l'histoire, ces trois innovations ont rendu indispensable l'invention de nouveaux instruments institutionnels dont deux forces économiques sont principalement à l'origine : les économies d'échelle et le développement de mécanismes permettant de faire exécuter les contrats à moindre coût. Ces deux forces ont un fonctionnement réciproque : le volume croissant des échanges, et donc de la production, nécessite la création de mécanismes efficaces garantissant le respect des contrats et, réciproquement, cette garantie de respect des contrats améliore les perspectives de profit ce qui amène une croissance des échanges. Cette vision du développement, que North explique avec plus de précision dans son ouvrage « *Le processus du développement économique* » (2005), est une vision qu'il applique au monde entier, sur temps long, c'est à dire depuis le X^{ème} siècle jusqu'à l'hégémonie occidentale, et dont la caractéristique principale est la prise en compte des coûts de transaction liés à la mise en place et au respect des droits de propriété.

Ainsi, l'analyse de North constitue le socle de la théorie néo-institutionnelle et vise, par l'introduction de l'histoire et du changement institutionnel, à dépasser la vision statique de l'économie standard en introduisant le triptyque théorique des coûts de transaction, des droits de propriété et du changement technique. Mais le cœur de son analyse réside dans l'étude du comportement humain et le refus de considérer la rationalité parfaite comme étant à la base des interactions économiques. C'est donc cette rationalité limitée, non substantive, qui va constituer le point d'achoppement d'un autre courant institutionnaliste, plus hétérodoxe dans son approche : la théorie des conventions.

B. Les prolongements hétérodoxes à la définition du rôle des institutions : de la théorie des conventions à la théorie de la régulation

Nous montrerons dans cette partie comment la théorie des conventions et la théorie de la régulation permettent d'appréhender la question des institutions, comment elles peuvent éclairer la question du développement de deux pays comme la Colombie et le Venezuela et quels sont leurs points de divergence avec le courant néo-institutionnaliste. L'objectif étant toujours d'identifier le cadre théorique cohérent, les outils analytiques pertinents, qui nous permettront d'avoir une lecture claire et la plus complète possible de l'architecture institutionnelle des systèmes de santé colombien et vénézuélien.

1. La théorie des conventions ou le dilemme du dépassement de la vision strictement stratégique vers une méthode d'analyse interprétative

La théorie néo-institutionnaliste et la théorie des conventions (TC) reposent sur une même base théorique, à savoir : **l'individualisme méthodologique**. En effet, l'objet de la TC est de comprendre comment, des individus confrontés à des situations d'incertitude parviennent à faire converger leurs décisions individuelles et ainsi ajuster leurs comportements les uns par rapport aux autres, c'est à dire : harmoniser leurs comportements individuels, autrement dit se *coordonner* (conventions internes) ou *coopérer* (conventions externes).

En posant en ces termes la question de la coordination et, par ailleurs, en reprenant à son compte le concept de rationalité limitée tel que l'a défini North, la TC concède, *a priori*, le même rôle aux institutions : celui « d'atténuateur » d'incertitude. En effet, il se trouve que les deux corpus théoriques mobilisent approximativement la même définition du terme « institution » (Daudigeos, 2009). Pour la TC, « *la notion d'institution est ici à prendre dans un sens très large qui recouvre toutes les formes de médiation sociale : coutumes, conventions, règles, normes, organisations et institutions au sens strict* » (Boyer, 2003). Pourtant, si l'on suit les « niveaux » formulées par Daudigeos (2009), une nuance apparaît entre les deux théories dans le sens que prend le concept d'institution. L'auteur explique, en effet, que le terme « institution » peut prendre quatre

sens différents, en sachant que les niveaux sont « inclusifs », c'est à dire que le second niveau inclut le premier niveau, et ainsi de suite :

- 1) Institutions = types particuliers d'organisations.
- 2) Institutions = larges secteurs de la vie sociale (famille, économie, science, religion)
- 3) Institution = toute « forme sociale » qui donne des règles d'action collective plus ou moins consciente (cognitive, normative, coercitive).
- 4) Institutions = règles sociales fondamentales et **tacites** qui permettent la coordination entre les individus.

Dans cette grille d'analyse, représentée de manière plus explicite dans le Schéma 1.2 ci-dessous (Schéma 1.2 : Les quatre niveaux de définition du terme « institution »), à mesure que l'on passe d'un niveau à un autre, le sens du terme « institution » admet une définition de plus en plus large qui inclut progressivement des critères cognitifs propres à chaque individu. En ce sens, l'auteur identifie le troisième niveau comme correspondant à la définition que donnent les néo-institutionnalistes du concept d'institutions, et il caractérise les troisième et quatrième niveaux comme participant, l'un et l'autre alternativement, à la définition qu'en donnent les auteurs conventionnalistes. La nuance se situe donc dans le fait que la TC se situe à deux niveaux d'analyse, selon ce qui peut être appelé un « *pluralisme vertical des conventions* » (Eymard-Duvernay, 2003) : un niveau « macroéconomique » dans lequel les conventions nécessitent une légitimité publique et ont vocation à concerner une société toute entière, comme c'est le cas, par exemple, pour la confiance dans une monnaie (Aglietta et Orléan, 1982) et, par ailleurs, une définition plus « microéconomique » des institutions laissant une place plus importante au caractère « tacite », implicite, des institutions et se focalisant sur une échelle locale voire individuelle des interactions humaines. Pour le dire encore différemment, il est possible de comprendre cette double définition comme étant une caractéristique de la TC qui est d'isoler deux niveaux d'intervention : celui des représentations et celui des règles.

Schéma 1.2 : Les quatre niveaux de définition du terme « Institution »

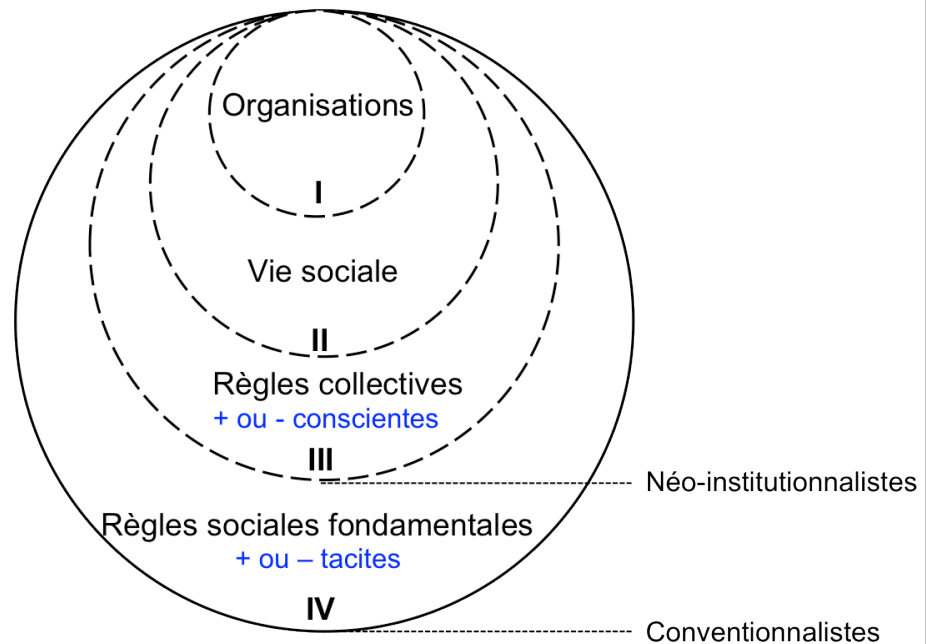


Schéma réalisé par l'auteur à partir de : Daudigeos Thibault, Valiorgue Bertrand (2009), « L'apport de la théorie des conventions à la théorie néo-institutionnelle des organisations »,

La raison fondamentale de ce double constat (la nuance qui sépare les deux corpus théoriques dans la définition des institutions et la double définition des institutions dans la TC) tient au fait que la convention est, en elle-même, une catégorie spécifique d'institution. En effet, l'effort de conceptualisation des conventionnalistes porte sur une institution particulière : les conventions.

La convention est une **représentation commune**, stabilisatrice, qui permet de régler un problème d'incertitude et de coordonner l'action d'un groupe d'individus. Ces individus interprètent leur environnement, agissent en conséquence et, ce faisant, ils formulent une règle qui n'est pas donnée au départ. Dans ses travaux, Lewis (1969) définit cinq composantes à la convention :

- 1) « *chacun se conforme à la convention* » ; dans une même situation où les agents doivent résoudre un problème de système d'information, le dispositif proposé sert de « modèle de réponse collectif » ;

- 2) « *chacun anticipe que tout le monde s'y conforme* », chaque individu anticipe que tous partagent et se conforment au même modèle de réponse collectif ;
- 3) « *chacun préfère que la conformité soit la plus générale possible* », le dispositif conventionnel n'a de sens que si une grande proportion de personnes font le même choix ; plus cette condition est remplie, plus l'utilité de chacun est élevée ;
- 4) « *il existe au moins une régularité alternative* », le dispositif n'est conventionnel que si il est le résultat d'un choix de la part des individus ; pour cela, chacun doit pouvoir avoir au moins une alternative au dispositif dominant ;
- 5) « *les quatre premières conditions sont Common Knowledge (CK : connaissance commune)* » ; CK veut dire que chacun sait que l'autre sait qu'il sait, etc. Cet instrument permet à Lewis de lier convention et rationalité des comportements mais est fortement critiqué (Dupuy, 1989) puisqu'il vide de son sens la notion même de rationalité. L'hypothèse de CK revient en effet à se coordonner avec son « clone » étant donné que le savoir détenu par chaque individu est absolument identique à celui de tous les autres. Le relâchement de cette hypothèse permet de résoudre ce paradoxe (Batifoulier, 2001).

C'est lorsqu'une situation vérifie ces cinq conditions que l'on peut utiliser le concept de « convention ». Les conventions ont donc des propriétés totalement différentes aux contrats ou aux institutions classiques, fondées sur des règles explicites, puisqu'elles sont **tacites** et **implicites**. Elles se posent donc comme mode de coordination « alternatif » aux schémas traditionnels proposés par les néo-institutionnalistes et remettent en cause, plus radicalement que ces derniers, les postulats du modèle classique puisque le choix que fait un individu peut être basé sur des éléments exogènes qui lui échappent.

Ainsi, pour Lewis, la convention est une **régularité de comportement** où chacun se conforme au comportement qu'il croit que l'autre aura. De cette façon, il associe subtilement l'arbitraire et le rationnel : « *La convention est arbitraire, mais il est rationnel de la suivre si chacun fait de même, car l'objectif est de se coordonner en adoptant le même comportement que les autres* » (Lewis,). Mais dans ses analyses, Lewis

utilise principalement la théorie des jeux qui reste, pour lui, l'outil privilégié du choix rationnel dans un contexte d'interaction entre plusieurs individus. En ce sens, deux principales visions sous-tendent la théorie des conventions : la vision humienne et la vision wébérienne (Batifoulier, 2001).

Dans la vision humienne, la convention est suivie par intérêt commun : elle définit en quelque sorte un « ordre spontané » qui émerge de l'intérêt qu'ont les individus à la suivre. Cette vision se rattache donc à une définition de **l'intérêt général**. La vision wébérienne, quant à elle, suppose l'existence d'un ordre préétabli et qui prescrit les comportements à suivre et à ne pas suivre. Cette vision donne une dimension prescriptive à la convention dans la mesure où elle introduit la **sanction** comme garantie de la stabilité de la convention. Ces deux visions donnent lieu à deux approches de la convention : l'approche stratégique et l'approche interprétative. Ces deux approches, qui sont le signe d'une ligne de fracture forte au sein de cette école de pensée (Raveaud, 2004), articulent différemment les deux éléments-clés : l'intérêt général et la sanction.

L'approche stratégique se focalise sur une vision de la convention en terme d'ordre spontané *via* la notion d'intérêt général et modélise pour cela la convention avec le langage de la théorie des jeux. En revanche, dans l'approche interprétative, la convention est à la fois une règle de comportement et un modèle d'évaluation de la règle. En d'autres termes, les individus ont, dans cette deuxième approche, une capacité d'interprétation des règles et de remise en question de celle-ci, ce qui rend plus réaliste l'analyse mais complexifie également le problème car ce jugement critique va être à l'origine de l'évolution des règles : le changement est *endogénéisé*. Dans la vision stratégique, la question du jugement de la situation par les individus est évacuée : la convention est forcément un équilibre de Nash car la règle est définie explicitement et l'intérêt stratégique du joueur est choisie *ex-ante* par le modélisateur via la matrice des gains. D'ailleurs, cette matrice des gains, qui est l'essence même de la vision stratégique, a finalement pour conséquence de rétablir la vision utilitariste et maximisatrice de l'individu, ce qui équivaut à un éloignement de l'un des buts initiaux de la théorie des conventions qui est de rompre avec le cadre théorique dominant. Ainsi, la théorie des jeux explique l'institution ou la convention par l'émergence d'un équilibre au sein d'un

jeu répété, lui-même basé sur un ensemble de règles exogènes (Aoki, 2002). Pour Searle (date), cette approche par le raisonnement stratégique aboutit à une « désubstantialisation » de l'action collective, qui est pourtant le cœur de l'économie des conventions. Ainsi, les théoriciens des jeux ne peuvent modéliser que des acteurs non-interprétatifs, ils ont donc une vision plus réductrice et moins « réaliste » des choix individuels que l'approche interprétative des conventions.

Ces deux approches amènent naturellement à deux interprétations différentes de la notion de « développement », qui est évoquée en terme de « dynamique des conventions ». L'approche stratégique aborde la question en terme de théorie des jeux évolutionnistes : une convention serait amenée à changer en fonction d'un processus de *mutation* des individus, c'est à dire en fonction du nombre d'individus choisissant une convention alternative dont la proportion est modélisée permettant ainsi de calculer un indicateur de « résistance » d'une convention aux « attaques » des individus « mutants » (auteur, date). Nous partons du principe que cette vision très restrictive et « mécaniciste » du développement ne permet pas d'appréhender l'ensemble des interactions dans lesquelles les individus s'inscrivent.

La vision interprétative des conventions, plus réaliste, conçoit le développement comme la conséquence des représentations que se font les individus du bien commun et cette interprétation à laquelle ils se livrent leur permet d'ajuster leurs actions en fonction de l'interaction émergente qui est observée : « *Ce sont les modalités de l'interprétation qui sont affectées puisqu'elles visent non plus l'adaptation a priori d'un comportement mais bien plus l'ajustement des conduites dans le cours même de l'action, donnant ici tout son poids au caractère procédural de la rationalité* » (Livet, 1994). Ainsi, l'interprétation, influencée par l'*idéologie*, et surtout l'apprentissage individuel et collectif, sont les clés de compréhension de la modification des règles, règles qui sont modifiées au fil d'un processus d'apprentissage mais dont les conventionnalistes n'ignorent pas leur lien étroit avec la notion de *pouvoir* et les conflits que suppose un changement de règles du fait de la recomposition de l'ordre social qu'elle induit (Rebérioux, 2001).

En somme, ces deux visions très différentes se différencient par l'importance qu'elles donnent à l'individu ou au collectif dans le processus de changement de convention.

Dans le cas de la vision stratégique, l'individu calculateur est au centre du processus de transformation (ou « mutation ») et dans le cas de la vision interprétative, le collectif influence l'individu dans la remise en question de la convention et donc, par l'interprétation que l'individu va faire de la convention, celui-ci va pouvoir à son tour modifier les règles qui régissent son action, c'est à dire le cadre institutionnel dans lequel s'inscrivent ses choix.

Ainsi, la théorie des conventions est partagée entre deux approches théoriques différentes, parmi lesquelles l'une d'entre elle nous est apparue comme insuffisante pour expliquer la dynamique des conventions et donc l'évolution du cadre institutionnel. Mais peut-être faut-il considérer, à l'instar de Raveaud (2005), qu'au-delà de ses différents théoriques internes, la théorie des conventions ouvre le champ à une approche empirique originale qui consiste à postuler que « *les individus sont partout et toujours plongés dans les institutions* » et que cette approche « *conduit le chercheur à identifier les cadres, dispositifs, juridiques, techniques, institutionnels, statistiques ou encore langagiers qui structurent les actions des personnes* ». L'auteur indique que ce qui caractérise la démarche empirique comme étant spécifiquement conventionnaliste, c'est l'attention particulière portée aux arguments échangés, aux justifications avancées par les personnes et à la pluralité des arrangements possibles. C'est uniquement de cette façon qu'une évaluation peut être portée sur les dispositifs institutionnels et leur interaction avec les catégories d'individus observées. Cette méthode, que prône l'auteur pour sortir du débat uniquement théorique, relève donc implicitement, et malgré lui, de l'analyse interprétative de la théorie des conventions puisqu'une place importante est laissée à l'interprétation que font les individus de leur cadre d'action. C'est vers cette méthode d'analyse empirique que nous orienterons nos recherches sur la comparaison des architectures institutionnelles.

2. De l'interprétation à la dynamique historique des institutions par la théorie de la régulation

Si l'analyse conventionnaliste est tiraillée entre deux approches dont la différence est de concéder plus ou moins d'importance à la micro-économie et au calcul stratégique des individus lors d'un choix, l'autre approche hétérodoxe analysant les dynamiques

institutionnelles, la théorie de la régulation, place son analyse, quant à elle, sur le plan de la macro-économie et accorde un rôle prépondérant au *politique* dans l'émergence des institutions.

En effet, pour la théorie de la régulation, « *les formes institutionnelles correspondent à la codification d'un ou plusieurs rapports sociaux fondamentaux* » (Boyer, 2003), par opposition aux « *micro-fondements des institutions et des organisations* » développés par les autres courants de pensée. Afin de développer cette démarche analytique originale, la théorie des conventions oriente son analyse dans trois directions :

Premièrement, la théorie de la régulation est une approche clairement institutionnaliste dont le but est de rechercher les « *fondements macro-institutionnels* » des économies de marché. Elle articule pour cela trois concepts centraux : régime d'accumulation, mode de régulation et formes institutionnelles. Le mécanisme est le suivant : l'analyse régulationniste identifie cinq formes institutionnelles (dites également « structurelles »), qui caractérisent une organisation sociale :

- 1) **La forme monétaire** (ou forme de la contrainte monétaire), la monnaie étant considérée comme un rapport social fondamental ;
- 2) **La forme du rapport salarial** (ou configuration du rapport salarial), le rapport salarial étant envisagé comme la modalité de rattachement du salarié à l'entreprise ;
- 3) **La forme de la concurrence** (ou mode d'organisation des « *relations entre les différents centres d'accumulation fractionnés* ») ;
- 4) **La forme de l'Etat** (ou forme des interventions de l'Etat dans le domaine économique) ;
- 5) **La forme de l'adhésion au régime international** (ou mode de relations entre l'Etat-nation et le reste du monde).

Ces cinq formes institutionnelles sont au centre de l'analyse régulationniste car elles vont façonner de manière décisive le *régime d'accumulation*, concept qui peut être défini comme étant l'ensemble des mécanismes par lesquels la croissance d'une économie, et donc l'accumulation du capital, se maintiennent dans le temps. Par exemple, ce régime,

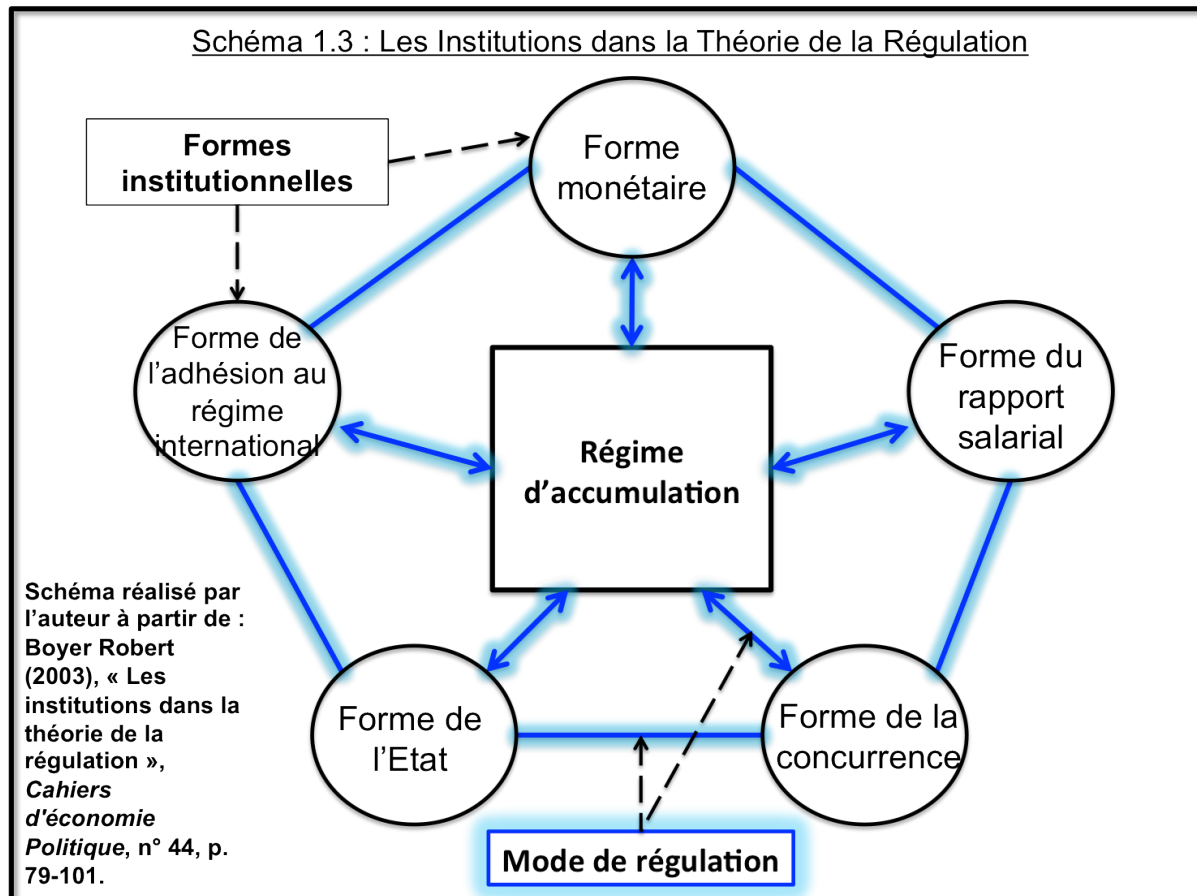
s'il a pu être caractérisé par une croissance des moyens de production au début du siècle (régime dit « extensif ») a muté, à partir de 1945 du fait des reconfigurations institutionnelles, en un régime basé sur les gains de productivité (régime dit « intensif »), puis basé sur les gains de productivité *avec* consommation de masse (« régime d'accumulation fordiste » qui correspond aux « *trente glorieuses* ») pour passer finalement dans les années 1990, du fait de forces exogènes aux Etat-nations (Bowles, 1990), à un régime d'accumulation « patrimonial » ou « financiarisé ».

Ainsi, dans l'analyse régulationniste, les cinq formes institutionnelles façonnent le régime d'accumulation, ce qui veut dire qu'elles déterminent, *in fine*, les modalités de maintien de la croissance économique et donc du profit. Précisons toutefois que la relation n'est pas unilatérale : les évolutions économiques, comme par exemple les crises structurelles, peuvent amener à des reconfigurations institutionnelles de grande ampleur, reconfigurations dont le rôle est de répondre aux conflits sociaux que la théorie de la régulation prend en compte dans son analyse. Lorsque les cinq formes institutionnelles se coordonnent les unes aux autres, lorsqu'elles dépassent les crises structurelles et qu'elles façonnent un régime d'accumulation stabilisé au travers d'une forme historique stabilisée des rapports de pouvoir, alors émerge un **mode de régulation**.

Le mode de régulation est l'ensemble des mécanismes, procédures et comportements qui permet aux formes institutionnelles de former un système cohérent, c'est à dire : « *reproduire les rapports sociaux fondamentaux ; soutenir et piloter le régime d'accumulation en vigueur ; assurer la compatibilité dynamique d'un ensemble de décisions décentralisées, sans que soit nécessaire l'intériorisation par les acteurs économiques des principes de l'ajustement de l'ensemble du système* » (Boyer, 2003). Autrement dit, le mode de régulation est l'ensemble des canaux par lesquels les formes institutionnelles se coordonnent entre elles et impactent le régime d'accumulation. Sur cette base, le schéma suivant (Cf. Schéma 1.3 : les Institutions dans la Théorie de la Régulation) se veut être un résumé de la vision régulationniste des institutions.

Deuxièmement, les institutions sont replacées dans leur contexte historique : *le temps long de l'Histoire* est pris en compte comme étant une composante majeure de

l'évolution du régime d'accumulation au sein des formes institutionnelles. Cette prise en compte de l'Histoire dans la dynamique institutionnelle place la théorie de la régulation à contre-pied des autres champs de recherche qui concentrent leurs analyses sur les équilibres institutionnels de court terme.



Troisièmement, la théorie de la régulation, à l'instar de l'approche interprétative de la théorie des conventions, réfute l'hypothèse d'anticipations rationnelles des individus et considère, au contraire, que les agents n'ont qu'une connaissance partielle et locale des réalités qui gouvernent leurs relations. C'est le sens de la séparation très claire entre régime d'accumulation (ou régime de croissance) et mode de régulation (ensemble de comportements qui « pilotent » le régime de croissance) par l'intermédiation des formes institutionnelles : « *il n'est pas de passage du macro au micro sans référence aux catégories intermédiaires que sont les formes institutionnelles. (...) Les institutions, en se situant au niveau méso-économique ont pour propriété de résumer les informations pertinentes pour les acteurs, au-delà du système de prix* » (Boyer, 2003). Nous retrouvons

ici une propriété des institutions commune à l'ensemble des approches institutionnalistes : elles sont, en dernière analyse, un « *réducteur d'incertitude* ».

Ainsi, avec ces trois éléments - formes institutionnelles façonnant le régime de croissance, prise en compte de l'Histoire et rejet des anticipations rationnelles - la théorie de la régulation érige un cadre conceptuel octroyant un rôle déterminant au *politique*. En effet, pour les régulationnistes, l'Histoire des institutions est celle des rapports de force entre salariat et fraction du capital industriel (puis patrimonial) dans la détermination de la forme hiérarchique dominante (Lipietz, 1991). Cette prise en compte des conflits sociaux comme élément fondamental du processus de changement institutionnel se fait donc en opposition avec les déterminants choisis par les autres approches institutionnalistes : déterminant technologique, économies d'échelle et dépendance au sentier pour les néo-institutionnalistes, ou même coévolution entre technologie, institution et normes pour les approches évolutionnistes, etc.

Par voie de conséquence, avec cette prise en compte des rapports de classes, concept émanant directement du marxisme, les régulationnistes portent l'éclairage sur le *rapport salarial* comme concept-clé dans l'explication des dynamiques institutionnelles et du passage d'un mode de régulation à un autre (Husson, 2001). Au final, le cœur de la question soulevée par la théorie de la régulation est bien d'analyser quelles conditions permettent d'assurer la viabilité d'une *pluralité* de régimes d'accumulation (Boyer, 2002) qui sont, par nature, soumis à des déséquilibres et à des contradictions, mais surtout, de quelle façon les modes de régulation se succèdent les uns aux autres. Dans cette optique, les institutions tiennent lieu d'axe de régulation du système capitaliste car, en tant que formes stabilisées des rapports de pouvoir et intermédiaires entre les individus et le régime d'accumulation, elles permettent de réguler les conflits sociaux sans les faire disparaître, et peuvent, lorsque les contradictions deviennent trop profondes, modifier leur architecture et ainsi transformer le régime d'accumulation : il y a alors passage à un nouveau mode de régulation.

Ainsi, ce rapide survol de la théorie de la régulation a permis de comprendre que les institutions y jouent un rôle central : enserrées dans une perspective historique de long terme incluant les dynamiques sociales elles sont, plus globalement, analysées dans le

but de comprendre le processus d'accumulation du capital, sa multiplicité, ses contradictions et son évolution. Du fait de sa portée et du caractère transdisciplinaire de son analyse, l'approche régulationniste permettra sans nul doute d'apporter, dans notre comparaison des architectures institutionnelles de la Colombie et du Venezuela, une lecture historique de leurs évolutions respectives, lesquelles s'enracinent incontestablement dans des problématiques politiques. L'histoire des luttes sociales en Amérique latine ainsi que l'évolution conjointe des architectures institutionnelles valide, sur cette question précise du déterminant politique, les analyses de la théorie de la régulation.

Conclusion de la Section 1

Pour conclure, les trois courants de pensée qui délimitent notre propre analyse des architectures institutionnelles que nous nous proposons de comparer, à savoir le néo-institutionnalisme, le conventionnalisme et le régulationnisme, apportent chacun un éclairage différent à la question des institutions, celle de leur frontière et de leur évolution. Une synthèse des trois corpus théorique - auquel nous avons ajouté la vision néo-classique - est présentée dans le tableau ci-dessous.

Cadre théorique	Nature des institutions	Origine des institutions	Dynamique des institutions
Néo-classique	Exogènes	-	-
Néo-institutionnaliste	Modalités efficaces de coordination suppléant ou palliant les mécanismes de marché	Résultat agrégé du calcul coûts/avantages, contrat entre les agents	Dépendance au sentier avec toutefois rétroaction des individus sur les institutions
Conventionnaliste	Les conventions sont un type d'institution différente du contrat car elles sont informelles, extérieures aux agents et collectives	Solution aux problèmes de coordination dus à une rationalité limitée et procédurale	Théorie des jeux évolutionniste / Interprétation et apprentissage collectif
Régulationniste	Compromis sociaux régulant les conflits sans les faire disparaître : formes stabilisées des rapports de pouvoir	Contingence historique, conséquences des pratiques sociales (structuralisme)	Produit des contradictions inhérentes au régime d'accumulation et conduisant à une trop forte pression sociale

Tableau réalisé par l'auteur sur la base de : SIMONNET Jean-Paul (2009), « Ecole de la régulation et institutionnalisme », Sans éd.

Ainsi, conscients désormais des tensions idéologiques divisant l'analyse des institutions, entre individualisme méthodologique et holisme ou entre vision stratégique et portée interprétative des processus cognitifs - tensions frappant de plein fouet la théorie des conventions, comme scindée en deux par cette ligne de fracture - notre démarche consistera dès lors à, premièrement, analyser la place laissée aux institutions dans les différents courants de l'économie de la santé et, dans un second temps, à construire, sur la base des concepts forgés par les grands courants de pensée institutionnalistes, un cadre conceptuel original, efficace dans l'exercice délicat d'une analyse comparative de deux architectures institutionnelles totalement différentes et axé, de surcroît, sur ce qui constitue un enjeu vital pour tout pays en développement : l'accès aux soins.

Section 2 L'économie de la santé et la place des institutions

Les différentes analyses des institutions ayant été posées, il s'agit désormais de connecter ces analyses à **l'économie de la santé**. En effet, une société s'organise autour de sous-systèmes dont l'agrégation forment un tout appelé « système ». Parmi ces sous-systèmes, celui de la santé est l'objet d'un intérêt croissant de la part des économistes. Le système de soins a été en effet l'enjeu principal des travaux des cinquante dernières années, de par l'importance croissante des problématiques de santé et de vieillissement au sein des sociétés post-industrielles. Plus profondément, cet intérêt accru pour le mode d'organisation des systèmes de soins semble être la conséquence naturelle de deux phénomènes inter-reliés : les découvertes majeures dans le domaine de la médecine et la croissance fulgurante, durant ces cinquante dernières années, des découvertes technologiques² qui se répandirent, par ricochet, dans le domaine médical. Cette croissance vertueuse du savoir et des moyens technologiques qui permettent l'application de ces connaissances explique le besoin, depuis cinquante ans, de penser l'organisation des systèmes de soins et de donner aux progrès acquis dans la science médicale les moyens organisationnels de se diffuser dans l'ensemble des sociétés.

² Ces découvertes technologiques dans tous les domaines étant elles-mêmes issues en grande partie des grands conflits armés qui ont marqué ce siècle ainsi que l'effort de recherche et développement réalisé par les Etats dans ce domaine en temps de paix.

Cette section a pour but d'ériger un cadre conceptuel nous permettant de saisir tous les aspects du processus d'institutionnalisation de l'accès aux soins dans des quartiers pauvres de villes comme Bogota et Caracas. Pour ce faire, il apparaît comme essentiel dans un premier temps, de définir clairement les liens théoriques unissant l'économie de la santé et l'analyse des institutions, ainsi que les contours précis de la notion « d'accès aux soins », centrale dans ce travail de thèse. A cette fin, une étude des différents courants de l'économie de la santé - depuis les thèses néoclassiques sur lesquelles s'érige la discipline, jusqu'à la prise en compte des institutions dans le choix des individus en matière de soins - s'avère importante pour permettre d'en cerner le périmètre théorique actuel. Ensuite, il s'agira de construire un raisonnement propre nous permettant de cerner le caractère spécifique de l'Amérique latine et donc des deux économies en développement qui font l'objet de cette étude, contribuant ainsi à définir les contours d'une démarche institutionnaliste en économie de la santé, à la fois originale et transversale, et qui nous permettra ensuite de comprendre le double impact des institutions sur les individus (et des individus sur les institutions) lors d'un processus d'institutionnalisation des soins médicaux dans des quartiers jusque-là inaccessibles aux institutions.

A. L'évolution progressive de l'économie de la santé : les insuffisances de l'analyse néoclassique dans sa prise en compte des institutions

Les systèmes sociaux, et en particuliers les systèmes de santé, se prêtent à l'analyse des institutions dans la mesure où ils sont, par essence, une construction institutionnelle, tant sur le plan des « représentations symboliques » - l'enjeu étant la vie elle-même, la santé tient naturellement une place centrale dans toute société humaine, toutes époques confondues - que sur le plan des structures, étatiques ou non, en charge de l'octroi des soins à la population et forcément d'ampleur nationale.

Pourtant et paradoxalement, l'économie de la santé s'est construite à ses débuts sur des bases théoriques néoclassiques, c'est à dire en excluant totalement l'importance des institutions, au sens large, au seul bénéfice des choix optimisateurs d'agents dits

« rationnels ». C'est cette construction théorique puisant ses sources dans les analyses standards qu'il convient d'étudier afin de comprendre les actuelles lignes de fracture divisant l'économie de la santé, lignes que l'on retrouve avec acuité dans le cas de la comparaison de la Colombie et du Venezuela.

En effet, les économistes de la santé considèrent que la discipline est née en 1963, avec l'article du prix Nobel d'économie, Kenneth Arrow (1963). Cet article, tout comme l'ensemble des premiers travaux concernant ce nouvel objet de recherche demeurent, comme l'explique Lise Rochaix (2009), à la recherche de légitimité auprès des économistes et restent donc calqués sur les modèles néoclassiques de concurrence pure et parfaite, sans véritable prise en compte de l'incertitude ou de l'ignorance du patient qui caractérise ce secteur. Mais c'est surtout au cours des années 1980, où la portée des thèses néolibérales s'amplifia dans le monde entier, conduisant à une période prolongée de restrictions budgétaires, qu'apparut de manière explicite le dilemme auquel les acteurs du système de santé sont confrontés : celui de besoins sans cesse étendus face à des moyens stagnant ou décroissants. Le calcul économique devint alors le moyen de faire face aux choix d'allocations de ressources et aux nouvelles problématiques de maîtrise des coûts.

L'économie de la santé évolua, d'une part, dans ce contexte de **tensions idéologiques** sur le rôle et la place de l'Etat dans la société, mais, d'autre part, dans un contexte de pression budgétaire croissante liée au vieillissement de la population des pays développés. Dans ce contexte très particulier, l'évolution du périmètre de l'économie de la santé peut se résumer comme étant une prise en compte croissante et progressive des différents acteurs d'un système de soins et du cadre institutionnel dans lequel ils interagissent : à une analyse strictement microéconomique (1), s'est substituée progressivement une démarche pluridisciplinaire, incluant le rôle du médecin, puis celui des patients et de l'Etat dans la formulation de la demande de soins (2). Nous verrons pourtant en filigrane que, bien que les modèles incluent de plus en plus de paramètres dans la formulation de la demande de soins, ils restent limitées par l'acceptation de l'hypothèse de rationalité parfaite et de comportements calculateurs. Nous verrons ensuite que les comportements des individus sont en fait largement tributaires d'un cadre institutionnel dont nous essayerons de définir le fonctionnement et la nature.

1) Du modèle néoclassique de la demande de soins...

A son origine, l'article de Kenneth Arrow est une analyse du comportement des consommateurs de soins. Cette étude est le point de départ de la microéconomie de la santé dont l'objet est d'analyser et modéliser les comportements individuels des acteurs dans le champ de la santé (Béjean, 2009). En tant que premier et principal outil des économistes pour interpréter le fonctionnement du secteur de la santé, en tant que « *point d'ancrage des études qui se sont développées depuis une quarantaine d'années* » (Chambaretaud, Hartmann, 2004) et surtout en tant que fondement des politiques de santé dans la plupart des systèmes de soins en vigueur, il semble primordial d'en décrire les caractéristiques essentielles.

La première hypothèse qui est posée est celle de la rationalité des consommateurs de soins. Ainsi, voulant assouvir une demande de soins, les consommateurs vont à la rencontre des producteurs afin d'acquérir des biens et services de santé. L'objectif visé est une amélioration de l'état de santé du consommateur de soins. De cette façon et à l'instar du reste de la science économique, les deux catégories traditionnelles que sont l'offre et la demande sont prises en compte comme les seules catégories permettant la compréhension des phénomènes de consommation médicale. Le marché est le lieu de rencontre exclusif de ces deux catégories qui fonctionnent de manière symétrique : le consommateur, parfaitement rationnel, formule sa demande en fonction de son revenu et des prix. A partir de ces deux variables, le consommateur adopte un comportement optimisateur, c'est à dire qu'il est capable de choisir les quantités de marchandises et de services qui maximisent leur utilité, sous une contrainte de revenu.

Il est important de noter que Arrow, dans ses travaux, note immédiatement le caractère imparfait de l'information détenue par le consommateur et convient qu'elle est une caractéristique essentielle de la défaillance du marché des soins médicaux (Arrow, 1963). Notons toutefois que cette « ignorance » du patient sera plus tard relativisée car, d'une part, le savoir médical n'étant pas une science exacte, le médecin comme le patient peuvent être amenés à commettre des erreurs de diagnostic (Béjean 1994, Domenighetti 1995), mais, d'autre part, les soins médicaux appartiennent à une catégorie très spécifique de biens appelés les « biens d'expérience » (Rochaix, 1989),

c'est à dire que les patients sont capables d'avoir un jugement, au moins subjectif, sur la qualité des soins prodigués et pourront, par la suite, diffuser cette information (c'est l'effet « réputation ») à d'autres individus (Tirole, 1993).

Le producteur, quant à lui, est également un agent optimisateur dont l'objectif est de maximiser son profit en mobilisant ses ressources disponibles. Les principes microéconomiques s'appliqueraient donc au secteur de la santé conformément aux autres sphères du champ économique. En ce sens et à l'instar de Thomas (2001) nous identifions trois modèles apportant chacun une progression dans la compréhension des mécanismes qui sous-tendent l'action des individus, dans le cadre de cette vision strictement néoclassique du comportement des acteurs du système de santé.

Le **premier modèle** appliquant ces principes est celui de Phelps et Newhouse (1974). Il considère la santé sous l'angle de la consommation et se base sur des mécanismes relativement simples : il étudie l'évolution de la demande des consommateurs par rapport au coût engendré pour obtenir des soins. Le point de départ du modèle est la charge pesant sur le malade sachant que ses dépenses sont couvertes par une forme très simple d'assurance : le consommateur paye une prime fixe d'assurance puis est couvert pour un pourcentage fixe de coût des soins indépendamment de la quantité qu'il consomme, le taux de remboursement étant déterminé par un contrat.

Dans ce modèle, le consommateur prends en compte deux variables, avant de formuler sa demande de soins : 1) la somme déboursée pour les soins, notée cP (avec c la fraction non remboursée par l'assurance et P le prix d'une unité de service de soins) ; et 2) ce que les auteurs appellent « la valeur du temps » (*The price of time*) notée Wt et qui représente le coût d'opportunité du temps passé dans les services de santé (avec t le temps et W le coût d'opportunité). Dans ce cadre de décision très particulier, le coût total d'une unité de soin pour un consommateur serait donc : $C = cP + Wt$.

Comme les auteurs le démontrent dans des travaux antérieurs sur l'élasticité-prix de la demande de soins, cette structure de la fonction de coût des soins de santé pour le consommateur révèle une forte sensibilité au taux de couverture assurantielle : plus la couverture sociale est grande, moins les individus ont à financer directement leurs soins

et donc, mécaniquement, plus la demande de soins est forte. En revanche, et les prolongements de Thomas et Montmarquette (2001) le confirment, la sensibilité de la demande de soins par rapport au « coût du temps » (ou coût d'opportunité du temps) est moins forte : un individu gravement malade ne pourra de toute façon pas utiliser son temps pour travailler ce qui implique que le coût d'opportunité décroît avec la gravité de la maladie. Dans ce modèle, les individus n'accordent donc, *a priori*, que peu d'importance au temps.

Par suite, le point d'intersection entre le coût marginal des soins et leur valeur relative pour le consommateur détermine la quantité d'équilibre que choisira l'individu dans le cadre d'un programme de maximisation de l'utilité sous contrainte budgétaire. Cette résolution par équation d'un problème d'une telle complexité amène cependant les auteurs à des prévisions générales relativement simples qui dépendent d'une variable principale : la partie non-remboursée du coût total des soins. Ainsi, dans le modèle de Phelps et Newhouse (1974), la demande de soins dépend principalement du coût des soins effectivement payés par les individus, le coût d'opportunité du temps ne jouant qu'un rôle secondaire dans leur décision. Par conséquent, la demande ne dépendant que du coût des soins, un système octroyant gratuitement des soins verrait la demande augmenter très fortement. Dans ce cas, la solution serait, pour les auteurs, une réduction de l'offre.

Nous voyons dans cette première analyse des interactions entre la demande et l'offre de soins - analyse fondatrice d'un courant de pensée très influent en économie de la santé : « la microéconomie de la santé » - une simplification extrême du comportement des acteurs, ramenés à un rôle d'agents optimisateurs et préoccupés, essentiellement, par des considérations d'ordre économiques. Le modèle prédit en effet que la demande de soins augmente lorsque l'état de santé de l'individu se détériore et lorsque son revenu s'accroît. Mais bien que cette relation paraisse intuitive, les études statistiques ne vérifient pas cette relation statistique (Béjean, 2009). La seule institution tenant une place réelle dans ce cadre d'analyse est le marché où se rencontrent l'offre et la demande. Cette analyse doit être bien sûr resituée dans le contexte des Etats-Unis des années 70, années qui connurent le triomphe de l'école de Chicago avec le début de l'application des politiques économiques néolibérales au Chili à partir de 1973 et,

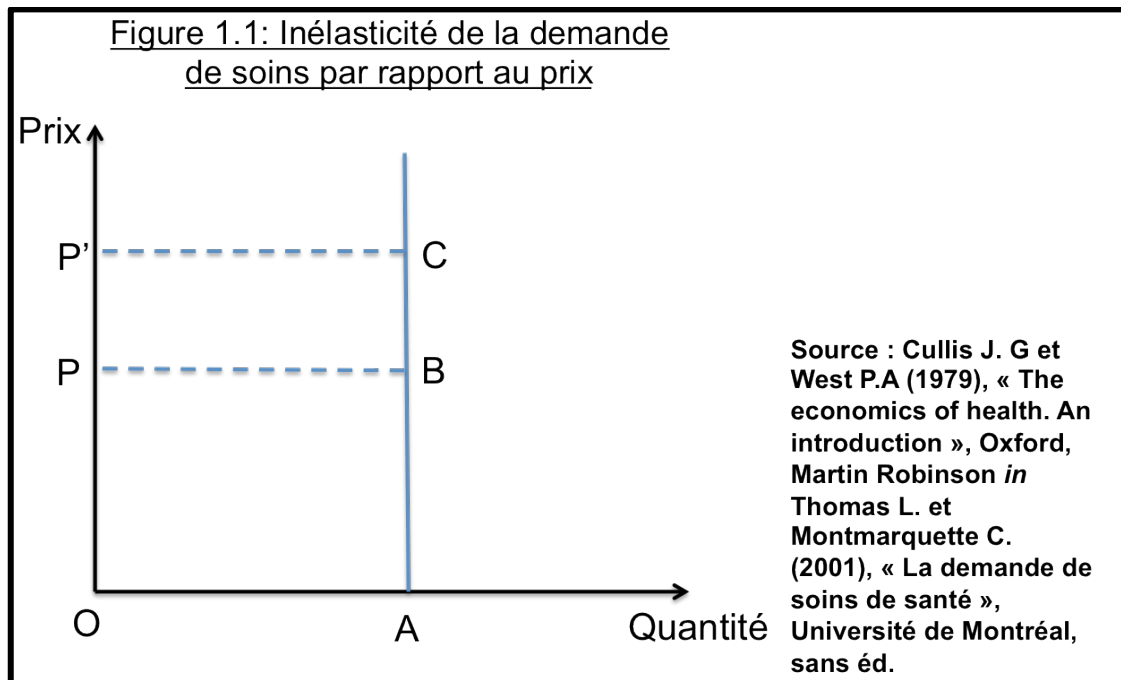
quelques années plus tard, l'attribution du prix Nobel d'économie à Milton Friedman. En ce sens, l'analyse de Phelps et Newhouse contribuera à établir que, le prix étant la variable dominante dans le choix des acteurs, seul un modèle basé sur la concurrence entre les offreurs du bien-santé est théoriquement en mesure de satisfaire la demande de soins. Cela aura un impact important dans les orientations choisies par de nombreux gouvernements concernant les mécanismes régulant leur système de santé, aboutissant à la privatisation d'un certain nombre d'entre eux et parmi lesquels le système de santé colombien en est une des nombreuses déclinaisons.

Cette vision des mécanismes régulant la demande de soins sera vivement critiquée par de nombreux économistes, y compris parmi ceux faisant partie de cette même branche néoclassique de l'économie de la santé, et introduiront progressivement d'autres éléments d'analyse contribuant ainsi à mieux cerner le fonctionnement des systèmes de santé. La **deuxième approche** que nous synthétiserons, l'approche par les besoins de Cullis et West (1979), est caractéristique de cette évolution.

En effet, cette approche tient compte d'une réalité évidente qu'avaient pourtant éludée les auteurs de l'approche par les coûts : un individu atteint d'une maladie consulte généralement un médecin, et c'est ce dernier qui, de part ses connaissances médicales, fait un diagnostic et prodigue les soins adaptés au patient. En d'autres termes, c'est le médecin qui offre la quantité de soins appropriés et choisit la nature de ceux-ci. Le patient, en revanche, n'a le choix que d'accepter ou de refuser cette offre. Cette hypothèse ramène le bien santé à une spécificité introduite initialement par Arrow et utilisée ici dans le cadre de l'approche par les besoins : le fait que la santé est avant tout un bien de confiance.

Ainsi, le modèle repose sur une première hypothèse essentielle : l'impossibilité pour le patient de prévoir son état de santé, en dehors de l'anticipation de sa dégradation naturelle due à l'âge. Cette incertitude des individus sur la probabilité de connaître une dégradation de leur état de santé et d'avoir recours au système de soins constitue ce qui est appelé le « *risque épidémiologique* » (Chambaretaud, Hartmann, 2004). C'est ce risque qui explique que la santé constitue un poste important dans la dépense des ménages (Arrow, 1963). La deuxième hypothèse fondamentale est que **la demande est**

totalemment inélastique par rapport au prix (Cf. Figure 1.1 : Inélasticité de la demande de soins par rapport au prix) du fait de la dépendance des individus vis à vis de l'offre des médecins. Cette deuxième hypothèse, très forte, a d'importantes conséquences sur le raisonnement global car elle permet de justifier le fait que des soins gratuits et universels améliorent l'équité au sein d'une population.

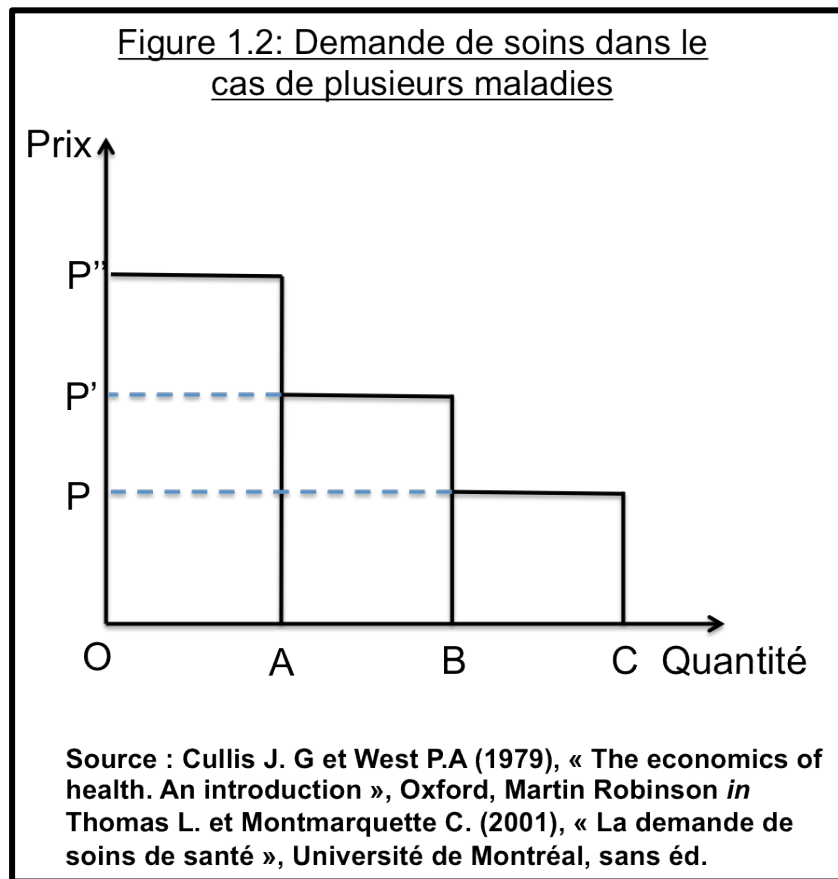


Sur la figure ci-dessus, il apparaît clairement que, quel que soit le prix des soins, P ou P' , la quantité consommée sera la même et équivalente à A . Le prix n'est ici pas limité par le coût marginal comme ce serait le cas pour une demande élastique au prix. De ce fait, un prix nul n'entraînant aucun réflexe de surconsommation étant donné l'inélasticité de la demande, les individus qui n'avaient pas les moyens de se faire soigner ont désormais accès aux soins. La solution serait donc *a priori* équitable, « à condition d'être financée par un impôt juste » (Thomas et Montmarquette, 2001).

En effet, les auteurs retiennent seulement deux solutions au financement : le financement « direct » au travers de cotisations ou dépenses personnelles, ou le financement « indirect » du système, c'est à dire le financement par l'impôt. Il exclut ainsi *de facto* une autre source de financement possible : le financement par l'utilisation

d'une rente, comme c'est le cas du Venezuela avec son pétrole. Mais cet élément n'interfère pas dans le raisonnement global qui consiste à dire que, en considérant que les individus peuvent être atteints par une seule maladie, et en faisant l'hypothèse forte d'une inélasticité de la demande par rapport au prix, la gratuité peut aboutir, sous certaines conditions, à un système efficace d'un point de vue de l'équité. Cette conclusion analytique, propre à l'approche par les besoins, pourrait éventuellement s'adapter aux conditions spécifiques des pays latino-américains que nous étudierons au chapitre deux, et notamment le cas du Venezuela où les soins sont totalement gratuits. Nous aurions alors une justification théorique à un choix très ambitieux de politique économique.

Mais les deux auteurs en viennent ensuite à complexifier leur modèle en introduisant l'hypothèse d'une gradualité des maladies en fonction de leur gravité. Cette complexification leur permet de nuancer leur propos. En effet, dans le cadre plus réaliste d'individus pouvant être touchés par plusieurs maladies au cours d'une période donnée, la gratuité des soins introduit une inefficience dans l'offre de soins. L'explication qui est donnée est relativement simple : en introduisant plusieurs maladies avec des niveaux de coût de traitement différents, **la demande redevient élastique par rapport au prix** du fait du choix qui s'offre aux individus (Cf. Figure 1.2 : Demande de soins dans le cas de plusieurs maladies). Ainsi, avec une demande redevenue dépendante du prix, les contraintes s'exerçant habituellement sur un marché apparaissent sur le marché des soins : une baisse du prix augmente ainsi mécaniquement la demande, au risque, si le prix est nul et donc inférieur au coût marginal, de déstabiliser le financement du système de santé.



La figure ci-dessus montre bien le principe introduit par Cullis et West dans leur approche par les besoins : à différents niveaux de maladies A, B ou C correspondent différents prix P'' , P' et P auxquels sont confrontés les consommateurs. Ils ont donc le **choix**, en fonction de leur budget, de dépenser les sommes P , P' et P'' dans les soins préconisés par le médecin. C'est ce choix qui, en permettant aux individus de se positionner sur une maladie plutôt qu'une autre - c'est à dire de se faire soigner telle maladie, généralement la plus grave, et non telle autre - redonne à la variable prix un impact sur le choix des individus et, ce faisant, annulerait les bienfaits de la gratuité due à l'inélasticité initiale de la demande.

Avec cette explication, l'approche par les besoins tend à converger vers les conclusions émises par les approches néoclassiques standard, elle élargit simplement le nombre de facteurs influençant la demande. Elle introduit surtout - et il nous semble qu'au delà de l'analyse de la demande, c'est le point essentiel à retenir - une nouvelle forme de régulation du système de soin, c'est à dire « d'institutionnalisme », en élevant la relation patient-médecin au rang de relation-clé permettant la compréhension des mécanismes

de fixation des quantités et des prix d'équilibre sur le marché de la santé. Cette hypothèse a permis de poser les premiers jalons de la théorie de la « demande induite », clairement définie par Rice (1983) comme étant la capacité du médecin à « *dispenser une prestation de service médical différente de celle que le patient choisirait si il disposait de la même information que lui* ». Cette définition donna lieu à un renouvellement de l'économie de la santé car, en effet, le pouvoir discrétionnaire de l'offreur de soins sur la demande fait du secteur de la santé un marché régi par des lois très particulières où la courbe de demande, en raison de l'asymétrie d'information, n'évolue pas de manière autonome : l'absence de stabilité de la courbe de demande remet en question le modèle économique néoclassique.

Mais ces développements ultérieurs sur la multiplicité des déterminants de la demande de soins se réaliseront surtout dans le prolongement des analyses de Grossman qui, grâce au concept de capital santé, et tout en restant dans la droite ligne des analyses standards, va parvenir à renouveler la recherche en économie de la santé et donner un nouveau souffle à la discipline.

2) ... à la demande de santé et le modèle de Grossman (1972)

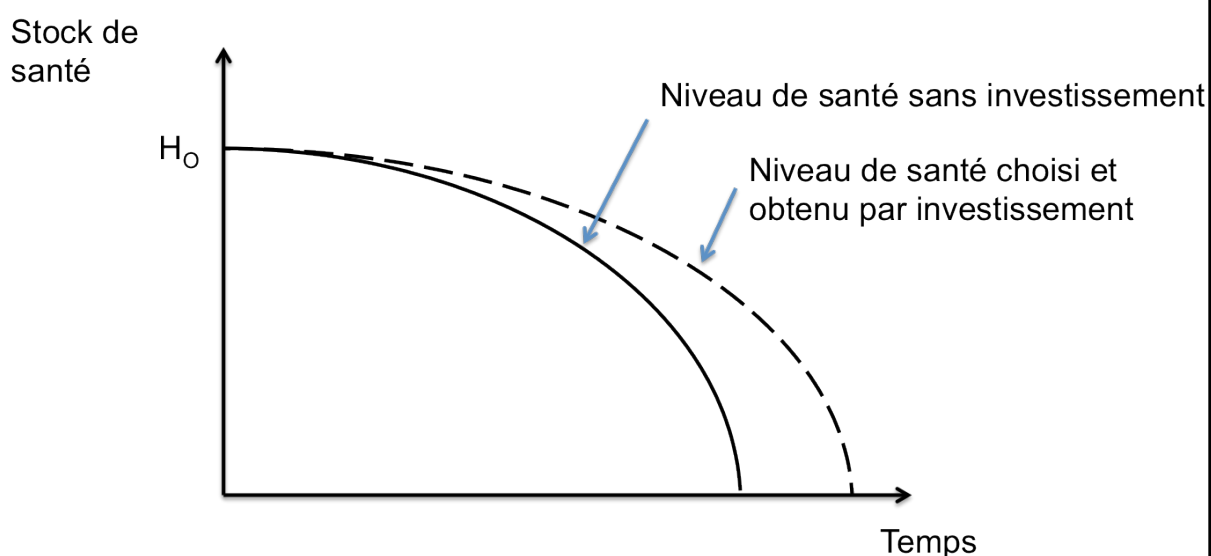
Le modèle de Phelps et Newhouse et celui de Cullis et West, étudiés précédemment, connaissent certaines limites que le modèle de Grossman a permis de surmonter (Menahem, 2000). En effet, du strict point de vue des comportements humains, ces modèles aident à comprendre comment se formule la demande de soins mais ne permettent pas d'évaluer le résultat de cette consommation sur les individus. La rationalité prêtée aux individus dans les modèles standards, rationalité qui est basée essentiellement sur les variables prix et contrainte budgétaire, ne peut expliquer à elle seule l'ensemble des comportements individuels. Grossman va donc introduire une différence d'efficacité productive entre les différentes consommations individuelles, lesquelles permettront d'optimiser ce qu'il va identifier comme étant le « capital santé ».

La notion de capital santé s'inscrit dans le cadre des analyses sur le capital humain, nées dans les années 60 avec les théoriciens de la « nouvelle économie domestique » (Becker, 1964). L'apport de Grossman sera de considérer l'état de santé de chaque individu

comme un capital dans lequel celui-ci devra « investir » afin de « l'optimiser » et accroître, *in fine*, sa longévité. Par conséquent - et c'est toute l'originalité de la démarche de Grossman - il découple la demande de santé de la demande de soins, distinction qui jusqu'alors n'était pas clairement explicitée : la demande de soins n'est que la conséquence de la demande de santé, c'est à dire de la volonté des individus de conserver leur capital santé. Ainsi, les individus ne consomment pas des biens santé uniquement pour satisfaire une demande qui augmenterait une utilité immédiate, mais dans le but indirect d'optimiser leur capital santé et donc, à terme, leur productivité. L'auteur, dans son article, l'explique sobrement de cette façon : « *la demande de soins de santé est une dérivée de la demande, plus fondamentale, de bonne santé* » (Grossman, 1972, p.248)

Ainsi, l'hypothèse principale du modèle est simple : la santé est un bien durable, assimilable à un « stock », dont les individus héritent à la naissance (« *initial stock* ») et qui se déprécie avec l'âge. La dépréciation de ce capital peut être freinée où même stoppée par des investissements, c'est à dire par la demande rationnelle de soins formulée par les individus. C'est là l'apport crucial de cette théorie : les individus sont parties prenantes dans la « gestion » de leur capital santé. La santé est pour chaque individu une variable endogène à leurs décisions, modes de vie, habitudes, etc. et est donc en quelque sorte une « production », et non plus seulement l'expression d'un besoin. Cette vision est résumée dans la figure 1.3 [Cf. figure 1.3 : hypothèse de Grossman, évolution du stock de santé avec le temps] où l'on observe clairement la dépréciation progressive du capital santé que seuls des investissements de la part de l'individu permet de compenser. Intuitivement, déjà, on observe que la charge de l'investissement en santé repose avant tout sur l'individu et sur la société et beaucoup moins sur le médecin. Par ailleurs, le modèle repose sur un ensemble de relations que nous allons tenter de résumer.

Figure 1.3 : Hypothèse de Grossman, évolution du stock de santé avec le temps



Source : Grossman M. (1972), « On the Concept of Health Capital and the Demand for Health », *The Journal of Political Economy*, Vol.80, N°2 (Mar.-Apr., 1972), p.232.

Tout d'abord, il convient de noter que la fonction d'utilité des individus est définie par Grossman comme étant dépendante de quatre variables : le stock de santé initial (H_0), le stock de santé à l'âge t (H), la consommation de services de santé (Φ_t) et la consommation d'autres biens et services (Z_t). Elle est de la forme : $U = U(H_0, \Phi_t, H_t, Z_t)$. A partir de ces quatre variables, les individus vont établir un arbitrage entre ce que va leur coûter un investissement en capital santé, aussi bien en coûts directs (services de santé) qu'en coûts indirects (temps, coût d'opportunité, etc.) et ce que va leur rapporter cet investissement en termes d'évolution de leur capital santé, c'est à dire le « rendement » de leur investissement (Fargeon, 2009).

Le rendement total d'investissement en santé est composé de deux éléments, qui séparent le raisonnement de Grossman en deux modèles différents (Thomas, 2001) : 1) un rendement « **psychologique** », dans la mesure où une unité supplémentaire dans le stock de santé d'un individu élève son bien-être. On parle dans ce cas de modèle de consommation ; et, 2) un rendement « **marginal monétaire d'investissement en santé** », noté γ_i , qui correspond au rapport entre le revenu (W_i), le produit marginal du

capital santé (G_i) et le coût marginal de l'investissement réalisé (π_{i-1}). On parle alors du modèle d'investissement. Grossman établit la relation suivante entre ces trois variables :

$$\Upsilon_i = \frac{W_i \cdot G_i}{\pi_{i-1}}$$

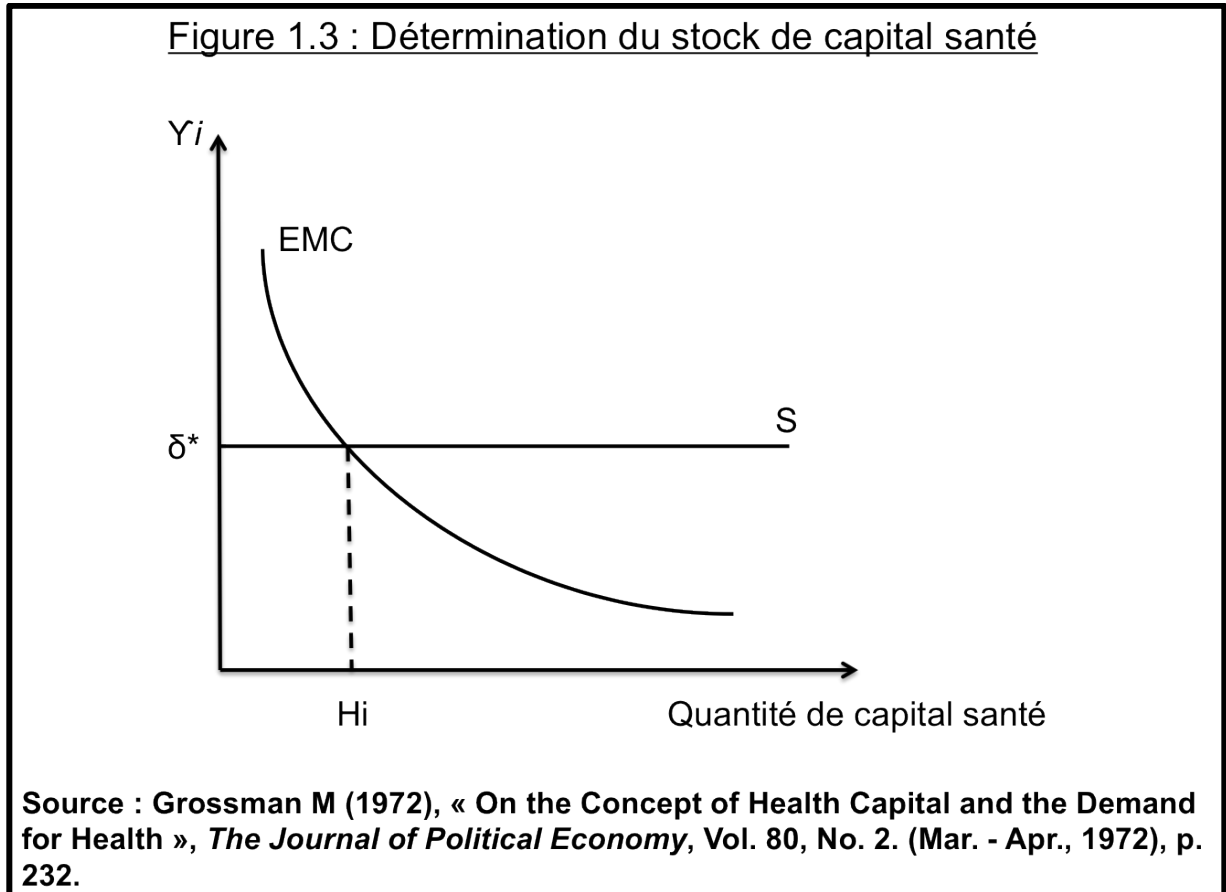
Cette relation signifie simplement que le rendement de la dernière unité monétaire investie dans la santé est important lorsque le coût marginal de cet investissement est faible (dénominateur), ou lorsque le revenu ou le produit marginal du capital santé est grand (numérateur). Cette relation logique constitue réellement le cœur de l'explication Grossmanienne des comportements des individus dans un système de santé. En effet, à partir de cette équation, l'auteur peut déterminer que les individus rationnels cherchent naturellement à optimiser ce rendement marginal monétaire d'investissement en santé. Cette première équation constitue ce qu'il appelle **l'efficacité marginale du capital santé** (EMS) qui décroît progressivement au cours du temps, ce qui veut dire qu'avec le temps, investir dans le capital santé est de moins en moins rentable pour un individu. La représentation graphique de cette équation représente la demande de capital santé déterminée, en dernière analyse, par la relation entre le stock de capital santé et son rendement d'investissement (Figure 1.3 : détermination du stock optimal de santé).

Mais Grossman détermine une seconde relation : il déduit mathématiquement que le rendement du capital santé (Υ_i) peut se comprendre également comme étant la somme du taux de dépréciation du capital santé (δ_i) et du taux d'intérêt (r_i) auquel renonce l'individu en faisant le choix d'investir dans son capital santé. Ainsi, l'auteur aboutit à la relation suivante : $\Upsilon_i = \delta_i + r_i$ qui s'apparente à la droite d'offre de santé et qui dépend du coût du capital et de son taux de dépréciation.

A l'optimum et logiquement, l'auteur obtient l'équation suivante :

$$\frac{W_i \cdot G_i}{\pi_{i-1}} = \delta_i + r_i$$

Cette égalité représente le point d'intersection entre la courbe de demande de santé (EMC) et la droite d'offre de santé (S), c'est à dire le stock optimal de capital santé pour un individu. La figure 1.4 suivante permet la lecture graphique de cet équilibre sur le marché de la santé :



Il est à noter que l'article de Grossman, pourtant central en économie de la santé puisqu'il introduit des innovations conceptuelles majeures, est très abscons dans sa forme et développe un raisonnement très mathématisé. A ce titre, l'auteur se situe bien dans la droite ligne de l'économie standard qui laisse une place prépondérante à l'hypothèse de rationalité et, par conséquent, aux formulations équationnelles des comportements humains. Ainsi, de la méthodologie de Grossman, plusieurs conclusions peuvent être tirées.

Tout d'abord, les modèles antérieurs introduisaient directement la consommation de biens médicaux dans la fonction d'utilité, puis, en appliquant simplement la théorie classique du consommateur, ils parvenaient à déterminer la demande de santé (Le Pen,

1988). L'apport le plus important de Grossman a été de mettre en évidence le fait que la consommation de soins n'est pas directement source d'utilité mais qu'elle permet de produire de la santé qui contribue au bien être de l'individu. Nous verrons dans les chapitres suivants que cette hypothèse dépend de la configuration institutionnelle dans laquelle s'inscrivent les individus et que cette vision de l'individu « auto-entrepreneur de son capital santé » ne s'adapte pas du tout à une architecture institutionnelle où l'offre de soins est limitée et saturée. En effet, même si les individus décident d'accorder de l'importance à leur capital santé, les investissements qu'ils sont censés faire dépendent fatalement du contexte immédiat de l'offre de soins, de l'environnement (qualité de l'eau, de l'air, des conditions de travail, etc.) et des liens institutionnels existant entre les individus et le système de santé. Autant de variables exogènes que le modèle de Grossman, quoique plus précis que les modèles étudiés précédemment, ne prend pas en compte. Cette relation existant entre qualité de vie et état de santé sera par la suite démontrée dans de nombreux travaux (Hertzman et al, 1994).

De ce fait, en se centrant exclusivement sur les comportements individuels face à la santé, le modèle de Grossman est généralement considéré comme étant **déterministe**. Et les déterminants que l'auteur développe, c'est à dire ceux qui ont un impact sur l'évolution du capital santé, sont : **le taux de salaire** - puisqu'un individu en bonne santé perdra d'autant moins de jours de salaires (coût d'opportunité) - **l'âge**, qui réduirait progressivement l'efficacité marginale du capital et enfin, **l'éducation**, puisque les individus ayant un haut niveau d'éducation seraient de meilleurs producteurs de santé. Ainsi, pour Grossman, les déterminants du capital santé sont uniquement personnels et en aucun cas exogènes.

Ainsi, le modèle de Grossman, dans la droite ligne de l'économie standard, place l'individu face à ses choix rationnels, confrontant ainsi le bien santé aux mécanismes de marché traditionnels. Si les conclusions de l'auteur vont plus loin que l'approche par les coûts de Phelps et Newhouse et que l'approche par les besoins de Cullis et West - laquelle laisse tout de même une place à la relation patient-médecin et offre donc, de ce point de vue, des débouchées à une grille d'analyse institutionnaliste - son apport en termes de capital santé, réutilisé ensuite par tout un pan de l'économie de la santé, est intrinsèquement a-institutionnel et déterministe. Au final, ce que l'on peut déduire de

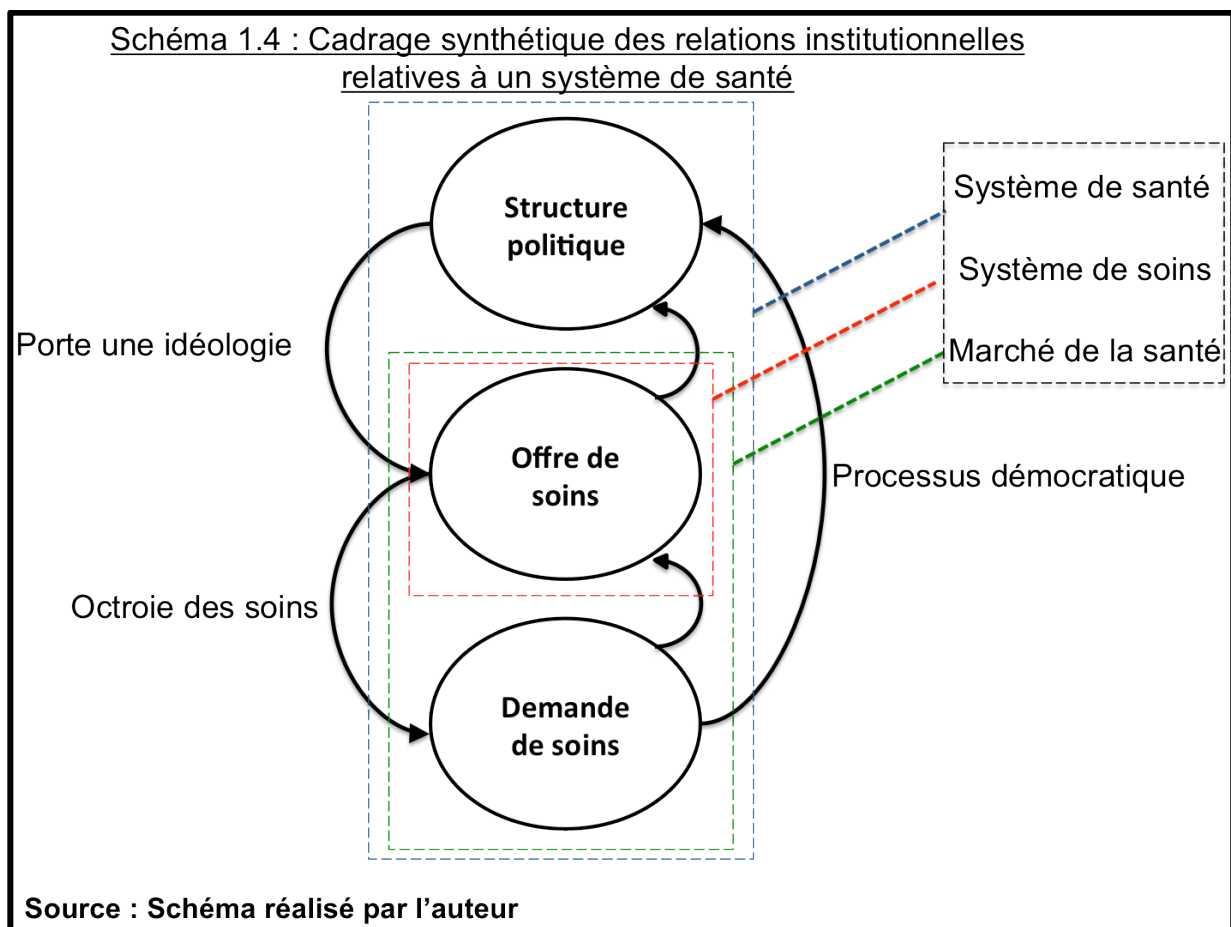
l'analyse de ces trois modèles fondamentaux en économie de la santé, c'est qu'avec eux, l'analyse des systèmes de soins s'est indéniablement enrichie, et notamment en termes de définition des déterminants de la demande, mais qu'elle s'est limitée à une grille de lecture centrée sur un individu supposé rationnel, opérant des choix dans un contexte systématiquement identique, quel que soit l'individu, et ayant une vision nécessairement instrumentale du bien santé : l'individu consomme des soins pour atteindre l'état de santé qu'il a choisi et arbitre rationnellement entre cette consommation et les autres. Il convient donc de comprendre que, si la plupart des mécanismes évoqués dans ces modèles sont viables et pour certains démontrés par la littérature, il apparaît comme étant impératif d'élargir l'analyse et, pour cerner le comportements des individus face à un système de santé, de prendre en compte les déterminants organisationnels et institutionnels.

B. Vers une approche institutionnaliste de la demande de soins : adaptation au contexte des pays latino-américains

Nous venons de voir que l'économie de la santé repose sur les bases et hypothèses de l'économie standard. Or, pour comparer deux systèmes de santé organisés autour de fondements idéologiques opposés et analyser le comportement des individus au sein de chacun d'eux, il est nécessaire, d'une part, d'introduire la dimension institutionnelle, absente (ou quasi-absente) dans l'analyse classique du marché de la santé et, d'autre part, de mettre en lumière la spécificité du contexte institutionnel des deux pays étudiés. Cette démarche en deux temps permettra, en filigrane, de construire un cheminement logique, une « grille de lecture » spécifique permettant la compréhension des mécanismes et des idéologies qui sous-tendent l'architecture institutionnelle des systèmes de santé colombien et vénézuélien.

1) Introduction de critères institutionnalistes dans les choix des individus : définition d'un cadre théorique spécifique

Nous avons vu, dans la section précédente, que trois courants économiques prenaient pleinement en compte les structures institutionnelles dans leurs analyses : le courant néo-institutionnaliste, le courant conventionnaliste et le courant régulationniste. La question est de savoir lequel de ces trois courants s'adapte le mieux à une analyse des systèmes de santé de deux pays comme la Colombie et le Venezuela. Afin de répondre à cette question, il est tout d'abord primordial de cerner les interactions principales existantes au sein d'un système de santé. Nous retenons, pour simplifier, trois niveaux institutionnels qui forment ce que nous identifions comme étant un ensemble de « circuits » : les patients (la demande de soins), les services de santé (l'offre de soins), et les structures politiques (Cf. Schéma 1.4 : Cadrage synthétique des relations institutionnelles relatives à un système de santé).



Le schéma ci-dessus représente le champ institutionnel global encadrant l'ensemble des interactions existantes entre les acteurs d'un système de santé. Nous avons départagé le schéma en trois périmètres institutionnels distincts :

- Le « système de soins » concerne ici l'ensemble des services octroyant des soins, tous niveaux d'attention confondus. Le système de soins se confond avec l'offre de soins.
- Le « marché de la santé » comprend les deux composantes classiques autour desquelles se structurent théoriquement les prix et les quantités offertes : l'offre et la demande de soins.
- Le « système de santé » est la définition la plus extensive que l'on peut attribuer à ce secteur puisqu'elle comprend la structure politique qui encadre le marché de la santé.

C'est à partir de ces trois dimensions que nous analyserons les inter-relations entre les différentes entités impliquées dans le fonctionnement du système de santé. Précisons que chaque entité du système de santé est elle-même traversée par des liens institutionnels internes : que ce soient des relations entre offreurs de soins, entre organes politiques ou entre patients, par exemple. Il convient de noter également que le système est ici analysé comme un « circuit », c'est à dire, pour résumer, que la structure politique définit l'organisation de l'offre de soins qui, elle-même, conditionne la demande de soins. A son tour, la demande de soins, c'est à dire les patients, influencent par différents canaux la structure politique, et ainsi de suite. Il est important de comprendre que **chaque relation est à double sens, aucune d'entre elle est unilatérale** : les patients ont un impact sur l'offre de soins (dans le cadre de la relation, par exemple, patient-médecin), qui elle-même entretient un rapport de force avec les structures politiques. Et, enfin, les structures politiques jouent, en dernier ressort, un rôle sur les représentations que se font les individus (la demande de soins) du système de santé, de son rôle et de sa place dans la société.

C'est cet ensemble d'inter-relations croisées entre les trois corps du système de santé qui définissent le cadre de notre vision institutionnaliste de ce système. En effet, au-delà des simples mécanismes d'auto-équilibre du marché de la santé, le système de santé

est ici compris comme une entité composite, dans laquelle les individus nouent un ensemble de relations, à la fois formelles (modalités législatives de coordination, type de rémunération en place, modalités de remboursement, etc.) et informelles (règles de comportement tacites émanant de représentations et valeurs portées par un système particulier), qui définissent et influencent la structure globale. Il n'y a, par conséquent, pas d'équilibre sur le long terme car le système tout entier évolue au rythme des relations institutionnelles que les acteurs nouent entre eux. Cette dynamique est rendue possible grâce au postulat selon lequel l'ensemble des interactions entre les acteurs du système forme un « circuit institutionnel ». En effet, les règles régulant le comportement des individus étant reliées entre elles, comme cela est représenté dans le schéma 1.4, celles-ci « s'autoalimentent » les unes aux autres et permettent une évolution permanente et plus ou moins progressive du système. Cette vision générale, que nous retenons, permet donc d'expliquer dans le détail, la nature des relations entre chaque corps du système de santé, mais également l'évolution institutionnelle globale. Parmi la multiplicité des relations existantes au sein du système de santé, nous développerons ici les principales.

La relation **patient-médecin** (qui est l'une des composantes de la relation offre-demande de soins, pour reprendre les catégories évoquées dans le schéma 1.4) est probablement la relation la plus étudiée en économie de la santé et celle qui permet de remettre le plus en question la vision standard de la rationalité individuelle amenant naturellement à l'équilibre du marché de la santé. A cet égard, plusieurs éléments peuvent être soulignés.

Tout d'abord, comme nous l'avons mentionné plus haut, la théorie de la demande induite par l'offre, introduite par Evans (1974) puis modélisée par Fuchs (1978), peut être mobilisée afin d'expliquer en quoi la position privilégiée des médecins, tant du point de vue des connaissances mobilisées « réellement » lors d'un acte médical, que du point de vue, plus subjectif, des représentations sociales qu'évoquent le statut de médecin, à plus forte raison dans une société en développement où le niveau de vie général de la population est bas, leur permet d'acquérir un pouvoir discrétionnaire sur les patients, mettant ainsi à mal l'idée d'une souveraineté totale des consommateurs de soins. Selon Evans (1974), de par sa position centrale dans le système de santé, le

médecin va élaborer des stratégies afin de maximiser sa fonction d'utilité qui inclut son revenu, son temps de travail et son pouvoir discrétionnaire.

Trinquard (2006) explique très bien le processus par lequel ce concept s'est construit puis à évolué. Dans les modèles initiaux de la demande induite - élément qui constitue un biais au bon fonctionnement du marché des soins (Evan 1974, Woodward et Warren-Bolton 1984) - la fonction de demande était considérée comme totalement endogène à l'égard des préférences de l'offre de soins. Autrement dit, l'offre de soins était considérée comme décidant entièrement, selon ses propres préférences et critères, pour la demande de soins. Cette hypothèse a par la suite été relâchée au profit d'une « demi-endogénéité » de la fonction de demande (Rossiter et Wilesky, 1984) où la demande induite émanant du médecin dépend de deux variables : l'état de santé du patient et un élément de création de demande variant selon plusieurs critères d'incitation. Dans les années 1990, ces modèles laissèrent la place à l'hypothèse de demande conditionnelle de Pauly (1980) selon laquelle le médecin fait une proposition au patient et ce dernier a le choix d'accepter ou de refuser l'offre. Dans ce modèle, le patient détient une information imparfaite dont il va se servir pour interpréter le choix qu'offre le médecin : il confronte son information privée à l'information diagnostique selon le principe de l'inférence bayésienne qui exprime de manière probabiliste le degré de confiance à une hypothèse (Trinquard, 2006). Ce modèle introduit, en d'autres termes, le recul qu'aurait tout patient face à l'information que lui livre le médecin et donc, le choix de chercher une seconde opinion (Rochaix, 1989) ou de refuser le traitement (Pitchik, 1987).

Cette vision, bien qu'elle utilise en grande partie les outils de l'économie standard, parvient à élaborer une critique cohérente de la concurrence entre offreurs de soins et à introduire une réflexion sur la nécessité d'une régulation - c'est à dire une intervention des pouvoirs publics - dans cette relation institutionnelle-clé, centrale pour tout système de santé : la relation patient-médecin. Cependant, il convient de noter que les études statistiques ne confirment pas clairement l'hypothèse de demande induite et cela, pour plusieurs raisons (Fargeon, 2009) : la difficulté à séparer, dans une prescription ou un acte médical, ce qui relève du soin nécessaire pour le patient et ce qui relève de l'induction ; la difficulté, également, d'imputer au phénomène de la demande induite la relation observée empiriquement entre la densité de médecins et le niveau

d'honoraires : en effet, d'autres théories expliquent cette relation, comme par exemple la saturation de l'information qui, dans un contexte de trop forte densité de médecins, annulerait l'effet de réputation, etc.

Mais au-delà de la question de l'induction de la demande par l'offre, le problème d'asymétrie d'information et d'incertitude, intrinsèque au bien santé et fondamentale dans la défaillance du marché des soins médicaux (Arrow, 1963), tient un rôle crucial dans la définition des modalités de la relation patient-médecin. Ces modalités ont été analysées sous l'angle de la théorie des contrats.

Cette théorie est une autre analyse faite des relations entre les individus et de la résolution des problèmes de coordination. Elle postule que le patient, n'ayant pas prise sur l'ensemble de l'information le concernant, tient un rôle de « principal » qui délèguerait ses choix au médecin, mieux informé, et qui tiendrait le rôle « d'agent » cherchant à maximiser l'utilité du patient. Cette relation principal-agent déjouerait les problèmes de coordination - risque moral ou sélection adverse, principalement (Rochaix, 2009) - par l'intermédiaire de contrats dont le rôle serait d'aiguiller le médecin vers une solution « gagnant-gagnant ». Trois critiques principales peuvent être émises à l'égard de cette théorie (Fargeon, 2009). Tout d'abord, le médecin est pris dans un ensemble de relations complexes et d'intérêts divergents (structure politique - structure de soins - patients) qui font que ses décisions dépendent de critères multiples (état de santé du patient, mode de rémunération, système de remboursement, critère psychologique, etc.) que ne saurait prendre en compte totalement la solution contractuelle. Par ailleurs, le contrat est défini par le financeur et non par le principal ou l'agent, ce qui éloigne d'autant plus ses termes de la réalité de la relation patient-médecin. Enfin, le contrat est conçu *ex ante*, or le domaine médical comporte une part importante de conséquences imprévisibles et même d'inconnu dû aux limites de la science médicale que le contrat, figé dans le temps de sa propre fixation, ne peut pas prendre en compte.

Ainsi, la relation patient-médecin est déterminée par une asymétrie d'information qu'ont pris en compte les deux modèles que nous venons d'étudier. Mais ces deux modèles ont pour particularité de maintenir la question institutionnelle dans le cadre

d'un choix optimal opéré par un autre acteur, le médecin, plus éclairé que le patient. Il s'agit donc tout simplement d'un transfert de souveraineté du patient vers le médecin, avec déplacement des questions décisionnelles et interactionnelles vers ce dernier, et d'une réduction de la question institutionnelle à un contrat qui régulerait théoriquement les comportements stratégiques. Par ailleurs, comme le souligne Brunat (2010), ces modèles s'appuient sur l'approche standard de la demande faisant des soins médicaux un bien comme les autres : divisible, révélateur de préférences et s'échangeant sur un marché. Or, la configuration de l'accès aux soins est conçue, dans notre démarche, comme la composante et la finalité d'un circuit institutionnel mettant en jeu des stratégies au sein de plusieurs périmètres décisionnels inter-reliés. Le marché de la santé est l'un de ces périmètres, mais pas le seul.

A cet égard, et bien que la relation patient-médecin soit centrale, notre analyse peut se porter sur de nombreux autres liens institutionnels qui, ensemble, font du système de santé un ensemble dynamique.

La relation entre la **structure politique et l'offre de soins** peut également être analysée comme une relation à double sens. La structure politique est comprise ici comme étant l'ensemble des entités jouant un rôle dans la définition de l'organisation, des mécanismes et des règles formelles qui sous-tendent le fonctionnement du système de soins : budget, lois, modes de rémunération, dosage entre système contributif, assistanciel et assuranciel, place des politiques préventives, choix des axes prioritaires, etc. En ce sens, la structure politique est au cœur de la construction du système de soins et porte, de ce fait, une *idéologie* qui détermine l'ensemble de ces paramètres et qui explique les variations dans la nature et l'intensité des relations institutionnelles d'un système à un autre. Cependant, cette relation n'est pas unilatérale dans la mesure où le système de soins offre de multiples résistances aux réformes politiques. Tout d'abord et de façon évidente, le système évolue selon l'inertie et le poids de la structure déjà existante. Aucun changement politique ne peut tout refonder radicalement : les processus réformateurs, y compris ceux qui ont pour volonté de modifier les règles institutionnelles en profondeur, sont tributaires de l'histoire du système de soins préexistant et de ses héritages tant formels qu'informels.

Par ailleurs, les corps constitués, c'est à dire les ensembles plus ou moins cohérents regroupant les professions médicales, sont en tension permanente avec les structures politiques, tant du point de vue de la gestion concrète du système de soins (budget, salaires, effectifs, etc.) que du point de vue de la cohérence ressentie des politiques de santé (privatisation, taux de remboursements, etc.). De ce point de vue, l'asymétrie d'information entre les structures politiques (plus éloignées et donc moins confrontées à l'environnement réel du système) et le système de soins (parfaitement informé) semble également être un facteur-clé pouvant expliquer les manques de cohérence institutionnelle dont peut être l'objet un système de santé. C'est grâce à cette tension permanente que l'offre de soins peut impacter les structures politiques : par la confrontation de l'idéologie au *principe de réalité*.

Cette idée du principe de réalité, qui est tout simplement l'écart entre les faits que prévoit une théorie et les faits réellement constatés, ce principe est également transposable à une troisième relation identifiée dans le système de santé : la relation entre **la demande de soins et les structures politiques**. En effet, il apparaît que cette relation qui boucle le circuit institutionnel, lui donnant cette dynamique endogène, est également à double sens. D'un côté, et par l'intermédiaire d'un certain nombre de médiations, les structures politiques s'attachent à modeler les représentations que se font les individus du système de santé en place - ou mis en place - afin de rendre légitime à leurs yeux l'idéologie portée et, d'un autre côté, les individus, également citoyens, expriment leur ressenti au travers d'un système démocratique et ont, théoriquement, le pouvoir de modifier, ou non, les choix réalisés par les structures politiques.

C'est ainsi que, dans cette vision institutionnaliste particulière des systèmes de santé, que nous qualifierons de « synthèse », les comportements et choix des individus sont certes encadrés par des règles formelles et informelles, mais les interactions multiples nouées entre les trois corps du système sont systématiquement réciproques et variables en intensité selon les systèmes de santé et l'idéologie sous-jacente. De ce fait et par rapport aux trois courants institutionnalistes identifiés dans la première section, cette vision se veut être une vision transversale, tenant compte à la fois des concepts Northiens d'*idéologie* et de *path dependance*, avec un attachement fort à certains principes néo-institutionnalistes comme la rétroaction des individus sur les institutions,

mais en retenant également le principe fondamental et conventionnaliste de rationalité limitée des individus - quel que soit le périmètre institutionnel de leur action - et d'apprentissage collectif, de même qu'est retenue comme pertinente l'analyse régulationniste des dynamiques institutionnelles sur temps long, prenant en compte les rapports de force sociaux et politiques comme variable explicative de l'instabilité des constructions systémiques.

Ce positionnement théorique trans-courant, qui devrait permettre l'analyse de deux systèmes de santé latino-américains, doit impérativement s'appuyer sur une analyse préalable des particularités du continent en matière d'architecture institutionnelle. Cette dernière analyse de ce premier chapitre permettra de positionner cette thèse sous l'angle de la **soutenabilité** des constructions institutionnelles.

2) Architectures institutionnelles des systèmes de santé : poids du réseau et poids de l'histoire comme singularités latino-américaines

La « fragilité institutionnelle » serait « la » maladie latino-américaine, rapportaient dans un article, Borner, Brunnetti et Weder (1992). Cette idée que l'instabilité macro-économique dont a souffert l'Amérique latine depuis plusieurs décennies est due, entre autre, à une fragilité institutionnelle est une opinion partagée par l'ensemble des spécialistes de la région. La question centrale est d'identifier la nature de cette fragilité et les causes historiques profondes spécifiques à l'Amérique latine, identifiées par les auteurs institutionnalistes.

En effet, North (2005) part de l'hypothèse de **non-ergodicité du monde** - un monde ergodique étant un monde où, à partir de la connaissance des unités de base, il est possible de déduire le fonctionnement de la structure économique globale, ce qui est une hypothèse fréquemment utilisée en physique moléculaire ou en chimie et qui est également l'hypothèse principale de l'économie néoclassique - pour expliquer la nécessité de s'intéresser à la dynamique des structures économiques et à l'évolution des principes fondamentaux inscrits dans l'esprit des individus. Pour lui, le monde étant en permanente évolution, il est essentiel de comprendre les changements à l'aune de ce qui en est le moteur : les individus et, surtout, « l'intentionnalité du changement » - il

s'oppose ainsi aux théories évolutionnistes et à l'idée d'un changement aveugle dans l'évolution. De ce point de vue, pour comprendre la dynamique d'une société, il est nécessaire d'en étudier l'évolution historique et en particulier l'évolution des institutions, c'est à dire des incitations formelles ou informelles qui ont poussé historiquement les individus à agir dans un sens plutôt que dans un autre et à construire progressivement la société telle qu'elle se présente actuellement : « *Pour changer la règle du jeu, vous devez comprendre l'histoire* » (North, 2005).

Dans un article écrit avec l'historien spécialiste de l'Amérique latine, William Summerhill, North (1999) compare le développement des institutions - et la différence de trajectoires de celles-ci - en Amérique du Nord et en Amérique du Sud. Il explique qu'avec la forte prégnance de la matrice institutionnelle léguée par les espagnols lors de la période de la colonisation, la construction de règles de comportements en Amérique latine a été conditionnée, depuis deux cents ans, par les valeurs hispano-chrétiennes qui donnèrent lieu à la forme très spécifique d'administration et d'organisation étatique observable aujourd'hui. Les schémas de pensées issus de cette civilisation se développèrent sur le continent latino-américain comme un héritage et un *path dependence* à partir duquel s'organisèrent les modalités de résolution des problèmes collectifs.

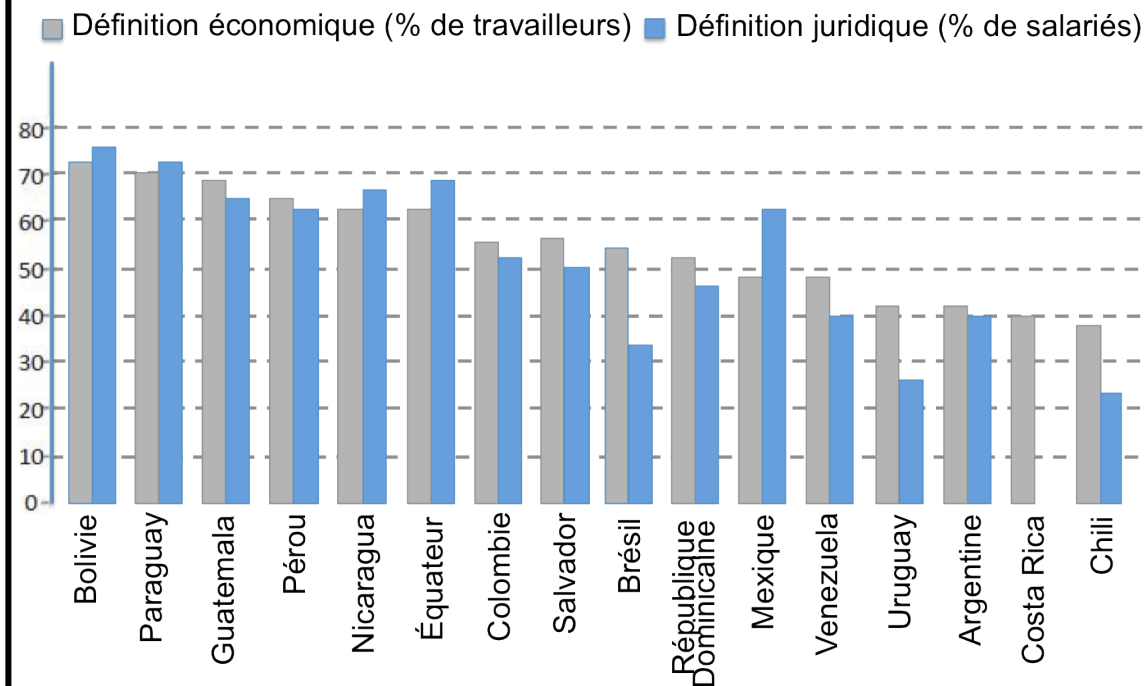
Ce qui est très intéressant c'est que, pour l'auteur, le patrimoine latino-américain est basé sur les **échanges personnels**. Il explique que les échanges entre les individus sont la clé de voûte de l'évolution des sociétés humaines. Au fil des millénaires, ces échanges se seraient plus ou moins transformés, selon les régions du monde, passant d'une coopération entre individus « de confiance » (famille, clan, etc.) à un type d'échange « impersonnel ». C'est à dire que les individus prennent progressivement le « risque » d'échanger avec des individus se situant en dehors du cercle restreint des connaissances, construisant, chemin faisant, un cadre institutionnel incitatif permettant « l'extension de la taille du marché », clé de la richesse des nations selon Smith. Or, « *pour augmenter la taille du marché, il faut passer de l'échange personnel à l'échange impersonnel* » (North, 2005).

Ce passage de l'échange personnel à l'échange impersonnel est l'étape que, contrairement à l'Amérique du Nord, l'Amérique du Sud n'aurait pas pleinement franchi. L'auteur explique que l'échange personnel est « génétiquement » ancré en l'Homme depuis des millions d'années et qu'il a fallu cinq ou six siècles à l'occident pour développer progressivement des institutions capables de créer un monde d'échanges impersonnels, c'est à dire de mettre en place les règles et leur systèmes d'application, juridique et judiciaire, capables de faire respecter la mise en œuvre des contrats dans le temps et dans l'espace.

Cette propriété énoncée par North reste très théorique mais semble tout de même pouvoir trouver plusieurs points d'ancrage dans la réalité sud-américaine.

Tout d'abord, l'omniprésence du **secteur informel** comme problématique commune à l'ensemble du continent s'inscrit pleinement dans le cadre d'analyse institutionnel que nous posons : le fait que des règles informelles régissent des aspects entiers de la vie des individus est une preuve tangible de l'échec des structures politiques à tisser un cadre institutionnel suffisamment incitatif pour permettre le dépassement du seuil des échanges personnels, pour reprendre le concept de North. Par exemple, selon les données de l'OCDE (2008) et bien que cette réalité soit, par définition, très difficile à estimer, le travail informel, dans sa définition juridique, c'est à dire pour les emplois ne donnant pas lieu à des droits à la retraite (notion que nous réutiliserons dans le chapitre 3 avec cette définition), s'élèverait aux alentours de 50% pour la Colombie et de 40% pour le Venezuela [Cf. figure 1.4 : le travail informel en Amérique latine].

Figure 1.4 : Le travail informel en Amérique latine



Source : Perspectives économiques de l'Amérique latine 2009, OCDE 2008

Cette question du secteur informel - qui s'étend même parfois dans certains domaines en opposition au secteur formel - est donc centrale en Amérique latine et pose avec acuité le problème de l'échec des structures étatiques à « formaliser l'informel » (Lopez-Castaño, 1987). Dans le domaine de la santé et selon des observations que nous avons pu faire *in situ* et sur lesquelles nous reviendrons avec précision au chapitre 3, il n'est pas rare de rencontrer ce phénomène de travail informel dans le domaine médical, en marge du système de santé : médecins généralistes manifestement « autoproclamés », médecine parallèle ou « non-conventionnelle ». Au-delà des individus encouragés par une impunité juridique ressentie ou réelle, ces pratiques s'inscrivent bien souvent, comme l'explique North et avec lui certains auteurs latino-américains (Montealegre, 2008), dans un cadre d'action qui, bien qu'il soit informel, reste très codifié et intrinsèquement lié au fait culturel et aux pratiques ancestrales. Nous voyons donc bien, à travers cet exemple précis, que le problème du manque d'institutions et de règles formelles se complexifie si, au-delà de la simple question économique, les questions culturelles et historiques - qui entrent pleinement en compte dans les représentations mentales des individus, et donc dans leurs choix - sont prises en compte, comme le recommande North.

Par ailleurs, cette question du poids important du secteur informel dans les sociétés latino-américaines nous montre que, dans ces sociétés en développement, un élément doit être absolument pris en compte dans l'analyse des règles de décision qui orientent les individus : le poids du **réseau**.

Le réseau, concept fécond en économie, y compris en économie de la santé, a pour caractéristique, selon nous, de s'intercaler entre l'individu et l'offre d'un bien. Il est défini, dans la littérature économique, de deux façons distinctes (Béjean, Gadreau, 1997) :

- 1) il peut être défini comme étant une structure, un « *réseau d'infrastructure* » ou,
- 2) il peut être considéré comme un mode de coordination des activités économiques.

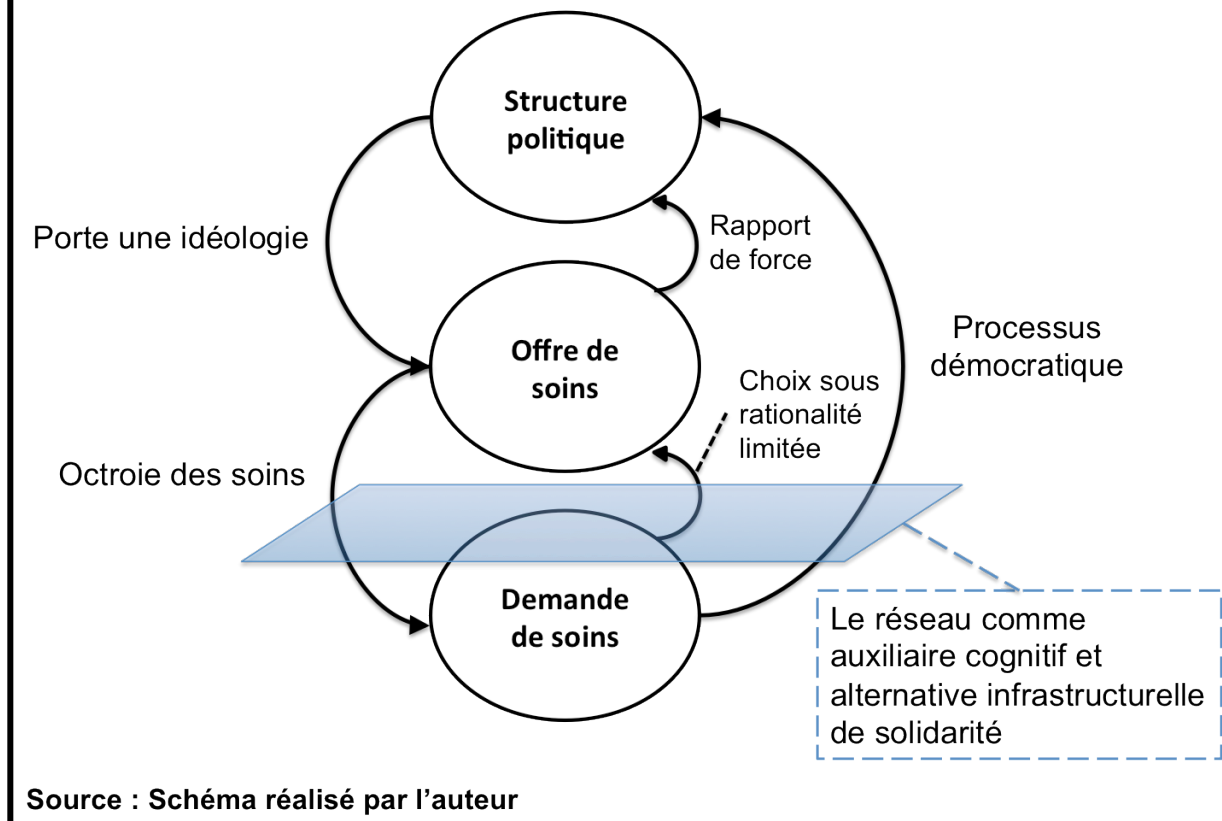
Nous allons retenir les deux définitions comme étant deux dimensions complémentaires de la notion de réseau. Nous allons, tout d'abord, expliciter la seconde définition : le réseau serait, dans cette optique « *un mode de coordination alternatif au marché, aux contrats ou aux conventions* ». Selon cette vision, le réseau interviendrait dans la décision des acteurs d'un système dans la mesure où, dans un contexte de rationalité limitée et d'information imparfaite, il « pousserait » les individus à faire des choix se situant en dehors du cadre institutionnel prévu initialement : en d'autres termes, dans un système, le réseau pourrait avoir comme effet d'éloigner le comportement « effectif » d'un individu du comportement « théorique » prévu par l'idéologie que sous-tend tout système. Par exemple, un individu fait partie d'une communauté globalement hostile à un certain type de soins, par exemple la médecine conventionnelle, lui préférant une médecine non-conventionnelle issue des traditions séculaires, bien que cette dernière se situe « hors-système », c'est à dire qu'elle ne bénéficie pas d'un remboursement des soins. Dans ce cas précis, on observe bien que l'individu est au départ placé dans un système d'incitations prévoyant un choix rationnel : choisir la médecine conventionnelle. Mais un réseau, un « mode de coordination alternatif », s'intercale entre lui et le choix rationnel et le pousse à choisir la médecine non-conventionnelle : il sort du circuit institutionnel.

Cette interprétation du réseau comme « auxiliaire » des mécanismes de décision individuels se situe dans la droite ligne des échanges personnels basés sur la confiance qui caractérisent, selon North, la structure institutionnelle latino-américaine et la différence du modèle nord-américain (et occidental). C'est à travers ce concept de réseau comme mode de coordination alternatif que peut également s'interpréter la notion de **solidarité**, dont la société latino-américaine, à l'instar des autres sociétés en développement, est imprégnée.

Selon la première définition que nous avons mentionnée plus haut, un réseau peut être également perçu comme une structure, un « réseau d'infrastructure ». Cette définition constitue la seconde dimension du concept de réseau que nous faisons le choix d'associer, dans le cas des pays en développement, à la notion de solidarité. Le réseau peut-être en effet perçu par un individu comme une alternative lui permettant d'échapper aux contraintes du circuit institutionnel (coûts, démarches administratives, distance, etc.). En ce sens, dans le domaine de la santé, le réseau est surtout observable à échelle restreinte : en milieu familial ou communautaire, comme expression d'une solidarité tribale. Les cas abondent d'individus qui, en vertu de ce puissant réseau de solidarité, choisissent de se faire soigner par un membre de la famille ou du quartier, plutôt que de se rendre au centre de soin tout proche.

Dans notre cas très spécifique des deux pays en développement que sont la Colombie et le Venezuela, nous avons donc appliqué le concept de réseau qui se scinde en deux volets : 1) un volet cognitif par lequel l'individu fonde ses choix individuels, potentiellement en dehors des processus incitatifs du système et 2) un volet infrastructurel, ou « réel », identifié au réseau de solidarité dans notre cas et perçu comme alternative par l'individu dans les choix qu'il réalise. Le concept de réseau, issu initialement de l'analyse du secteur informel, peut désormais s'intégrer au circuit institutionnel comme particularité latino-américaine [Cf. Schéma 1.5 : Circuit institutionnel relatif à un système de santé incluant le concept de réseau].

**Schéma 1.5 : Circuit institutionnel relatif à un système de santé
incluant le concept de réseau**



Ainsi, les individus insérés dans un système possédant un ensemble d'interactions propres aux pays en développement rencontrent cette entité qu'est le réseau s'interposant entre eux et l'offre d'un bien, en l'occurrence le bien santé, comme un prisme dans leur cadre décisionnel.

Mais un autre facteur essentiel intervenant dans les comportements individuels au sein des sociétés latino-américaines est à relever, nous permettant ainsi de conclure : le poids des schémas de pensée importés par la colonisation et l'impérialisme. Nous avons effleuré en début de section ce thème crucial, caractéristique de l'Amérique latine, il convient ici de le développer un peu plus.

Depuis environ 500 ans, l'Amérique latine est un point de passage des différents empires qui ont dominé successivement la planète : l'Europe, tout d'abord, puis les Etats-Unis depuis deux siècles. Au-delà du génocide des populations autochtones, cette histoire coloniale laissa de multiples empreintes parmi lesquelles [Galeano, 1971] :

Premièrement, d'un point de vue strictement économique, une société entièrement tournée vers l'exportation des matières premières, laquelle, fruit d'un pillage méthodique du continent, ne bénéficia qu'à une petite élite compradore. Ainsi, le mode d'extraction et d'acheminement des richesses tirées du sol d'Amérique latine pendant des siècles (Eduardo Galeano parle des « *veines ouvertes de l'Amérique latine* ») modela un système économique dit « excluant », c'est à dire intrinsèquement inégalitaire, n'accumulant que pauvreté pour l'immense majorité de la population.

Deuxièmement, l'élément économique précité explique que d'un point de vue institutionnel, la démocratie étant par ailleurs un concept récent sur le « continent des dictatures militaires », les individus s'inscrivent dans des schémas de pensée où, généralement, aucune option n'est écartée. Sur le plan des choix des individus relatif à un marché particulier, comme par exemple les soins de santé, cela se traduit par un plus grand usage des marchés parallèles et de l'informalité, considérés comme des options parfaitement viables et notamment par l'intermédiaire du *réseau*, comme nous l'avons analysé précédemment. Mais sur le plan des structures politiques en général, cette idée que les circuits institutionnels classiques n'ont pas forcément vocation à être respectés conduit *ipso facto* à une très grande instabilité politique et donc, par extension, à une très grande fragilité institutionnelle. En effet, dans les deux pays qui font l'objet de ce travail de thèse, l'un, la Colombie, est frappé par une guerre civile depuis plus de cinquante ans et l'autre le Venezuela, n'est sorti de troubles politiques très intenses (le « *caracazo* » en 1989, entre autre) que pour entrer dans une période où se succédèrent les coups d'Etats et les tentatives de déstabilisation, la dernière datant de 2003 avec « le coup d'Etat pétrolier ».

Ainsi et troisièmement, cette instabilité institutionnelle, dont la racine est incontestablement à chercher dans la pauvreté généralisée laissée dans le sillage de la colonisation, a produit au cœur de ces sociétés deux mouvements totalement contradictoires : d'abord, un attachement profond et une fascination pour les valeurs héritées de la période coloniale et néocoloniale, en particulier les valeurs nord-américaines qui imprègnent les sociétés colombiennes et vénézuéliennes. Et parallèlement, ce qui pourtant en est l'exact contraire : un profond rejet de l'hégémonie

américaine et de l'ingérence politique de cette nation à leur endroit, traduite par des mouvements de contestation anti-impérialistes dans les deux pays, à des échelles différentes. Ces deux mouvements, fascination-rejet, *a priori* opposés, se retrouvent de manière paroxystique dans l'analyse des architectures institutionnelles des systèmes de santé des deux pays : l'un, le système de santé colombien, est pour ainsi dire une copie du système de santé américain, laissant une place centrale aux assureurs privés et aux schémas de décision « néoclassiques » et l'autre, le système de santé vénézuélien, est la tentative de construire un modèle absolument opposé au modèle néolibéral : un système de santé gratuit et universel, érigé selon les principes du « *Socialisme du XXI^{ème} siècle* ». Le développement et l'explicitation de ces deux macro-systèmes de santé est précisément l'objet du second chapitre.

Conclusion du Chapitre 1

Ainsi, l'analyse des soubassements théoriques de l'économie des institutions, d'une part, puis de l'économie de la santé d'autre part, nous a démontré, au-delà des limites des modèles néoclassiques que nous avons soulignées, la nécessité d'interpénétration de ces deux disciplines afin de comprendre pleinement et de façon réaliste le comportement des acteurs d'un système de santé et l'ensemble des facteurs psychologiques qui s'exercent sur eux. Nous avons, dans cet esprit, tenté d'ériger un cadre théorique se voulant « synthétique », en construisant le concept de « circuit institutionnel » basé sur des concepts-clés qui structurent les différents courants institutionnalistes : la rétroaction, le *path dependance*, la rationalité limitée ou encore le rôle de l'Histoire sur les architectures institutionnelles. Cette démarche nous a aidé à identifier quelques spécificités fondamentales aux sociétés latino-américaines parmi lesquelles le rôle multidimensionnel des réseaux et le lourd héritage institutionnel de la colonisation. Tous ces jalons théoriques nous permettent désormais de nous lancer dans l'analyse de deux systèmes de santé qu'à priori tout oppose : les systèmes de santé colombien et vénézuélien.