Chapitre 4

L'appariement

Les problèmes d'affectation des élèves dans les collèges, des collégiens dans les lycées ou des lycéens dans les filières d'enseignement supérieur ont été étudiés par les économistes et les théoriciens des jeux, qui ont proposé pour les résoudre des modèles dit « d'appariement » (*matching* en anglais). Toute une littérature s'est developpée depuis les années 60, depuis l'article pionnier [GS-62] de David Gale (1921–2008) et Lloyd Shapley (1923–2016), qui a valu le prix « Nobel » d'économie à ce dernier en 2012[1]. Dans cet article, sur lequel nous reviendrons amplement, le principal exemple est celui de l'appariement matrimonial où une agence dispose des préférences d'un ensemble d'hommes vis-à-vis d'un ensemble de femmes et, réciproquement, de celles des femmes vis-à-vis des hommes, et cherche un « bon » appariement entre hommes et femmes compte tenu de leurs préférences. Les problèmes d'affectation scolaires peuvent être vus comme des variations (non-triviales) de ce problème de l'appariement matrimonial.

Les problèmes d'appariement soulèvent plusieurs questions fondamentales. D'abord, il y a celle de savoir ce qu'est un « bon » appariement. Si on se concentre sur un seul individu et que l'on connaît ses préférences, il n'est pas difficile de déterminer ce qu'est un bon appariement : c'est celui qui satisfait au mieux ses préférences. La question devient autrement plus épineuse quand on s'intéresse à un ensemble d'individus, avec des préférences potentiellement divergentes. La première tâche qui incombe à l'analyse est donc celle de dégager les « bonnes » propriétés d'un mécanisme d'appariement, c'est-à-dire les propriétés dont on pense qu'il est souhaitable qu'un mécanisme les satisfasse. La seconde question est alors celle de savoir s'il est *possible* de satisfaire les bonnes propriétés que l'on a identifiées. Si c'est le cas, se pose ensuite celle de savoir quelle méthode ou *algorithme* peut construire de manière systématique ces bons appariements.

L'objectif central de ce chapitre est de donner un aperçu des principales réponses qui ont

1. Pour une introduction à cette littérature, voir [FHI-13] et [Rot-08], qui ont inspiré notre présentation.

été données à ces questions par une littérature qui s'est surtout développée dans le champ de l'économie théorique. Nous le faisons dans la section centrale du chapitre, la section 4.2, en partant du problème de l'appariement matrimonial qui est un appariement dit « 1-1 » au sens où il s'agit d'apparier *un* membre d'une première catégorie (les hommes) à *un* membre d'une seconde catégorie (les femmes). Nous nous rapprochons ensuite de l'affectation scolaire en passant aux appariements « plusieurs-1 » où il s'agit d'apparier ***plusieurs*** membres d'une catégorie (les élèves) à *un* membre d'une seconde catégorie (les établissements). Enfin, si la théorie semble apporter un cadre solide et des résultats qui peuvent donner confiance dans la capacité des modèles proposés à nous aider a résoudre les problèmes d'affectation, nous verrons que certaines difficultés spécifiques importantes apparaissent quand on veut appliquer ce cadre aux affectations scolaires.

Quand on le considère abstraitement, un problème d'appariement peut être vu comme consistant à trouver un mécanisme qui permette de satisfaire de la meilleure manière possible les préférences d'un certain ensemble d'individus. Les problèmes de cette forme sont traités à un niveau supérieur de généralité par la *théorie du choix social*, qui est une autre branche de l'économie théorique. Nous faisons donc précéder notre présentation de la théorie de l'appariement d'une brève introduction à la théorie du choix social, dans la section 4.1. Si nous procédons ainsi, c'est parce que la théorie du choix social s'est fait une spécialité de concevoir et d'analyser les « bonnes » propriétés que l'on doit attendre d'un mécanisme qui « agrège » (c'est le terme consacré par la littérature) des préférences. Et aussi parce qu'elle a montré, au travers de « théorèmes d'impossibilité » qui l'ont rendue célèbre, qu'il était tout sauf évident qu'existent des mécanismes qui satisfassent simultanément les bonnes propriétés qu'on pouvait attendre d'eux.

4.1 La théorie du choix social

4.1.1 Des paradoxes du vote à la théorie du choix social

La théorie du choix social (TCS) est une branche de l'*économie normative*, la partie de l'économie qui s'intéresse à l'évaluation des institutions et des politiques économiques. Elle est également étroitement liée à la philosophie politique et à la partie des sciences politiques qui s'intéresse aux mécanismes de vote. La TCS s'intéresse aux situations dans lesquelles des individus ont à choisir entre plusieurs options qui les concernent tous, des options dites *sociales*. C'est le cas par exemple lorsque, dans une démocratie, les citoyens doivent choisir un candidat à la présidentielle ou quand le gouvernement représentant les citoyens doit choisir comment affecter les lycéens dans le supérieur. Imaginons que nous connaissons les préférences des individus sur les différentes options possibles. La TCS étudie alors les opérations par lesquelles ces préférences individuelles des membres d'un groupe (par exemple, les citoyens d'un pays, les membres d'une association ou encore les lycéens de Terminale) sont « agrégées » ou synthétisées dans une préférence collective sur les différentes options ou dans une décision collective, c'est-à-dire le choix d'une des options.

Le vote constitue l'un des exemples les plus étudiés d'une telle opération. Les électeurs ont en effet des préférences entre les candidats, et la procédure de vote aboutit en principe à une décision collective (la sélection d'un des candidats), parfois (mais plus rarement) à une préférence collective (un classement des candidats en lice).

Dès le XVIII^{e} siècle, les travaux pionniers de CONDORCET (1743-1794) et de BORDA (1733-1799) ont mis en évidence le fait que, dès qu'il y a plus de deux candidats, il est tout sauf trivial de trouver une bonne procédure de vote. Considérons la *règle majoritaire binaire*, selon laquelle si une majorité d'électeurs préfère strictement le candidat x au candidat y, alors x doit être collectivement préféré à y. Soient trois électeurs a, b et c dont les préférences entre les trois candidats x, y et z (on notera par la suite X l'ensemble des candidats) se résument comme suit[2] :

a	b	c
x	y	z
y	z	x
z	x	y

Ce tableau représente ainsi ce que l'on appelle un « profil » de préférences individuelles, c'est-à-dire la donnée d'une relation de préférence par individu du groupe. Si on applique la règle majoritaire binaire pour comparer les options x et y avec le profil de préférences représenté ci-dessus, on doit en conclure que $x \succ_C y$ (ce qui signifie « x est strictement préféré à y du point de vue collectif »), mais aussi, en comparant y et z, que $y \succ_C z$, et en comparant z et x, que $z \succ_C x$. Autrement dit, la procédure d'agrégation en question engendre des préférences collectives « cycliques ». De telles préférences collectives posent problème. Premièrement, on peut considérer qu'elles sont pathologiques parce qu'on peut douter qu'elles soient intrinsèquement cohérentes. Le problème est qu'en conséquence on ne voit pas comment prendre une décision satisfaisante sur la base de ces préférences : quel que soit le candidat que l'on considère, il y en aura toujours un autre qui lui sera préféré du point de vue collectif ! Cette émergence de préférences collectives cycliques à partir de la règle majoritaire binaire est appelée le « paradoxe de Condorcet ».

Deux points méritent d'être soulignés.

Tout d'abord, nous avons considéré une seule propriété de la procédure de vote : sa faculté à engendrer des préférences collectives acycliques. Mais il existe manifestement de nombreuses propriétés dont on souhaiterait qu'une procédure de vote les possède. Par exemple, on pourrait vouloir qu'une procédure de vote traite tous les candidats (« neutralité ») ou tous les électeurs (« anonymat ») de la même façon. On pourrait aussi vouloir que lorsque tous les électeurs préfèrent x à y, la préférence collective s'aligne sur ces préférences unanimes (« principe d'unanimité »). Envisager les procédures de vote du point de vue des propriétés qu'elles satisfont (ou ne satisfont pas) est extrêmement utile. Cela permet, d'abord, d'évaluer une procédure donnée. Si une procédure de vote ne satisfait pas une propriété qui nous semble particulièrement attractive (par exemple, si elle ne satisfait pas le principe d'unanimité), alors cela donne une forte raison de ne pas l'employer. Dans le meilleur des cas, on peut arriver à *caractériser* une procédure de vote : énoncer une liste de propriétés qui sont telles que si la procédure est suivie, les propriétés sont satisfaites et, réciproquement, si les propriétés sont satisfaites, c'est que la procédure en question a été suivie. Inversement, plutôt que de partir d'une procédure de vote particulière, on peut se donner une liste de propriétés jugées désirables, une sorte de « cahier des charges », et se demander s'il existe des procédures de vote qui la satisfont. C'est typiquement ce genre de question dont traite la TCS.

Le second point à souligner est le suivant. Nous avons considéré des situations de vote. Mais bien d'autres situations ont une structure analogue. C'est le cas des situations d'allocations de biens. Supposons par exemple qu'il y ait une certaine quantité d'un bien à partager entre plusieurs individus. Au lieu d'avoir comme options sociales un ensemble de candidats X, on a désormais des partages possibles du bien entre les individus. Et au lieu d'avoir des préférences individuelles entre les candidats, on part de préférences individuelles entre ces partages possibles. C'est aussi le cas des situations d'appariements (qui peuvent se voir comme des formes d'allocations). Supposons par exemple qu'une agence matrimoniale ait une liste d'hommes et de femmes à apparier. X désigne alors l'ensemble des appariements possibles. Les clients de l'agence ont des préférences entre ces appariements possibles, et normalement l'agence propose un appariement qui satisfait au mieux les préférences de ses clients.

2. La convention de lecture est la suivante. Chaque colonne décrit les préférences d'un électeur. Si une option x est située plus haut qu'une option y, elle lui est préférée strictement par l'électeur. Si deux options se situent sur la même ligne, dans la même colonne, l'électeur est indifférent entre les deux options.

Dans toutes ces situations, il est question d'agréger des préférences individuelles définies sur un certain ensemble d'options X (pouvant être interprété comme désignant des candidats, des partages possibles, des appariements possibles, etc.) en une préférence collective ou en une décision collective. La TCS fait précisément le pas vers l'abstraction que ces exemples suggèrent : elle considère, de manière générale, les manières d'associer une préférence collective à partir de préférences individuelles définies sur un ensemble X.

4.1.2 Le choix social non dictatorial : évidence des principes et impossibilité théorique

On considère généralement que l'acte de naissance de la TCS remonte à l'ouvrage *Social Choice and Individual Values* (1951/1963) [Arr-51] de l'économiste Kenneth ARROW (1921 - 2017). ARROW élabore un formalisme mathématique qui effectue précisément les deux mouvements d'abstraction que nous venons de décrire : il considère les procédures d'agrégation des préférences en général, et il s'intéresse aux propriétés (jugées désirables ou non) de ces procédures d'agrégation. Surtout, il établit mathématiquement un résultat, appelé couramment le « Théorème d'Arrow », qui montre, en substance, que le paradoxe de Condorcet n'a malheureusement rien d'accidentel. Revenons en effet au paradoxe : comment peut-on réagir quand on s'en rend compte ? On peut procéder par tâtonnement : changer la procédure de vote, déterminer certaines des propriétés qu'elle satisfait (ou pas), et l'accepter ou la rejeter sur cette base. La tâche est manifestement colossale. Elle a occupé les spécialistes de la théorie du vote pendant près de deux siècles. L'approche d'ARROW est différente : il se donne un ensemble minimal de propriétés supposées désirables et cherche à déterminer s'il existe (au moins) une procédure d'agrégation qui les satisfait. Si l'on est d'accord sur cet ensemble minimal de propriétés alors une réponse négative signifierait qu'il est vain de chercher, en tâtonnant, une bonne procédure d'agrégation. Or, c'est précisément ce qu'affirme le Théorème d'Arrow, qu'on qualifie souvent, pour cette raison, de « théorème d'impossibilité » : il n'existe pas de procédure d'agrégation qui soit capable de satisfaire simultanément l'ensemble minimal de propriétés qu'ARROW se donne.

L'intérêt du résultat dépend évidemment des propriétés retenues. Elles sont au nombre de cinq.

(i). La *condition d'Universalité* (U) exige d'une procédure d'agrégation qu'elle s'applique à n'importe quel profil de préférences individuelles, pour autant que celles-ci soient transitives et complètes [3]. La règle majoritaire binaire et bien d'autres procédures satisfont cette condition.

(ii). La *condition de Rationalité* (R) exige que les préférences collectives engendrées par la procédure soient toujours transitives et complètes. Le paradoxe de Condorcet montre que la règle majoritaire binaire ne satisfait pas cette condition. En revanche, les procédures qui associent des *scores* à chaque option en fonction des préférences individuelles (par exemple, une célèbre règle dite de Borda) satisfont la condition (R) : en supposant que les préférences collectives s'alignent sur ces scores, on n'obtient pas forcément des préférences transitives et complètes.

(iii). La *condition d'unanimité* (appelée aussi *principe de Pareto faible* (PPf)) exige que si tous les individus préfèrent (strictement) une option x à une option y, alors x est préféré collectivement à y. La condition d'unanimité est l'une des propriétés qui fait le plus consensus parmi les spécialistes de l'économie normative. Si une procédure n'y obéit pas, cela signifie qu'il y a une option x que tout le monde préfère strictement à y et que pourtant la procédure classe derrière y.

3. Désignons par $\succeq_i$ les préférences dites « larges » de l'individu i. Alors $x \succeq_i y$ signifie que x est au moins aussi bon que y du point de vue de i. Les préférences de i sont transitives si, dès que $x \succeq_i y$ et $y \succeq_i z$, alors $x \succeq_i z$. Des préférences sont *complètes* si, dès qu'on considère deux options x et y, il y en a au moins une des deux qui est jugée au moins aussi bonne que l'autre. La complétude n'exclut pas que l'individu soit indifférent, c'est-à-dire qu'il juge x et y de même valeur. Par contre, elle exclut que x et x soient jugés incomparables.

(iv). La *condition d'Indépendance* (I) exige que les préférences collectives entre deux options x et y ne dépendent que des préférences individuelles entre x et y (et pas des préférences individuelles entre les autres options).

(v). La *condition d'absence de Dictature* (D) interdit qu'il existe un individu (le « dictateur ») qui impose ses préférences au groupe dans toutes les situations possibles [4].

Le théorème d'Arrow affirme qu'aucune procédure d'agrégation ne peut satisfaire ces propriétés simultanément.

4.1.3 La théorie du choix social et la question de la sincérité des individus

Jusqu'à présent, nous avons tacitement supposé que la procédure d'agrégation était informée des « vraies » préférences des individus. Dans de nombreuses applications, ce sont les individus eux-mêmes qui déclarent leurs préférences au mécanisme d'agrégation. Dans les scrutins électoraux, les électeurs indiquent typiquement le nom d'un candidat. Dans les procédures d'appariement, les individus formulent souvent des vœux. Or, rien ne garantit *a priori* qu'une procédure d'agrégation incite les individus à révéler sincèrement leurs préférences. Par exemple, la procédure de vote improbable qui désigne comme vainqueur le candidat qui est apparu le plus de fois en bas des classements individuels ***déclarés*** aurait exactement l'effet inverse. Si l'électeur a préfère le candidat x, il aura intérêt à déclarer que x est en bas de son classement et donc à mentir sur ses préférences.

Pour aborder la question de l'incitation à la déclaration sincère des préférences individuelles, la TCS a défini la notion de « manipulabilité ». On dit qu'une procédure d'agrégation est *manipulable* s'il existe un individu i^* et un profil de préférences individuelles déclarées p^* tels que, compte tenu de p^* et de la procédure d'agrégation, l'issue de l'agrégation pourrait être meilleure du point de vue de i^* s'il mentait sur ses préférences. Il suffit qu'il y ait une seule situation où un seul individu ait intérêt à mentir pour qu'une procédure soit considérée comme manipulable. La manipulabilité définit l'incitation au mensonge en supposant, en quelque sorte, que l'individu i^* connaît les préférences que les autres individus vont déclarer. Cette dernière hypothèse est manifestement erronée dans beaucoup d'applications, même si des informations peuvent parfois donner une idée assez précise de ce à quoi ressembleront les déclarations des autres individus dans le cas des sondages. Néanmoins, il peut tout à fait suffire que i^* juge p^* assez probable pour qu'il ait intérêt à mentir sur ses préférences. Or, lorsque le risque est sérieux que les individus mentent sur leurs préférences, ce sont toutes les « bonnes propriétés » de la procédure qui sont menacées. En effet, comme nous l'avons vu dans la section précédente, la TCS évalue les procédures en considérant que les profils de préférences sont sincères. Par exemple, cela n'a de sens d'exiger d'une procédure de vote qu'elle satisfasse la condition d'unanimité que si cette condition est définie à partir des vraies préférences des individus.

La non-manipulabilité est donc une propriété désirable pour les procédures qui reposent sur des préférences déclarées. Or, dans deux contributions indépendantes, GIBBARD (1973) et SATTERWHAITE (1975) ont montré que, sous des conditions assez générales, il n'existe pas de procédures d'agrégation qui soit non-manipulable. Il s'agit là d'un théorème d'impossibilité, comme celui d'ARROW. (Précisons, néanmoins, que le présent théorème (dit « de Gibbard-Satterwhaite ») s'intéresse non pas aux procédures qui engendrent une préférence collective sur l'ensemble des options (comme chez ARROW) mais à celles qui engendrent une décision collective, c'est-à-dire le choix d'une unique option.)

4. Notons que le principe de Pareto n'est pas du tout incompatible avec l'existence d'un dictateur : quand il y a unanimité, cela signifie que le dictateur et le reste du groupe sont d'accord. En suivant l'opinion unanime, on suit donc toujours le dictateur.

Quelle conclusion peut-on tirer de ces deux théorèmes d'impossibilité ? Que l'on ne peut pas exiger grand chose à un niveau général. Il n'est pas désespéré de trouver des procédures d'agrégation qui satisfassent de bonnes propriétés, mais cela ne se fera qu'à un niveau spécifique, par exemple pour des procédures spécifiques ou des profils de préférences spécifiques. Ainsi nous allons voir maintenant ce que dit la théorie dans le cas particulier des procédures d'appariement.

4.2 Les modèles d'appariement

4.2.1 Une décision centralisée

Lloyd Shapley et Alvin Roth ont obtenu en 2012 le prix de sciences économiques de la Banque Royale de Suède à la mémoire d'Alfred Nobel, pour leurs travaux sur « *un problème économique central : comment associer différents agents le mieux possible* ». Cette déclaration de l'Académie royale suédoise des sciences donne à une littérature, considérée jusque-là comme très spécialisée et technique, un statut central dans la discipline des sciences économiques. Elle rappelle que si les économistes s'intéressent aux marchés, ils s'intéressent plus généralement à la question de l'allocation et de la coordination ; que le marché et les prix ne sont pas toujours la bonne réponse à apporter ; enfin que la question de l'organisation d'une procédure décentralisée ou centralisée dépend des caractéristiques du problème économique.

Considérons le problème que nous avons pris comme étude de cas, celui de l'affectation des lycéens dans les établissements du supérieur. On peut très bien imaginer une procédure décentralisée : aux candidats de chercher l'information et de présenter des dossiers de candidature ; aux établissements de faire une offre de formation et d'étudier les dossiers. C'est un peu comme si on ouvrait un marché comparable au marché de l'emploi, mais avec des contraintes strictes en termes de dates [5]. On imagine bien les problèmes que cela peut poser. Par exemple, un candidat, qui a envoyé plusieurs dossiers et reçoit une première réponse favorable venant d'un établissement qui ne fait pas parti de ses favoris, va devoir décider d'accepter avec quelques regrets ou de refuser au risque de ne pas obtenir mieux par la suite. Alvin Roth et Xing (1994) ont étudié différents problèmes d'affectation (par exemple, dans les championnats universitaires de football américain) et ont montré les inefficacités que peut générer une procédure décentralisée dès lors qu'existent des contraintes temporelles. On conçoit bien, en effet, que l'on a beaucoup à gagner, du point de vue de la diffusion de l'information et de la coordination, à trouver une procédure centralisée. Par ailleurs, une procédure centralisée se recommande pour la gouvernance d'un domaine public. Elle doit permettre de traduire un certain nombre de décisions politiques dans les caractéristiques générales de la procédure d'affectation, tout en respectant au mieux les préférences des candidats. Il s'agit bien d'un problème concret de choix social.

4.2.2 Les bonnes propriétés d'un modèle d'appariement

La littérature sur les modèles d'appariement trouve son origine dans l'article fondateur [GS-62] publié en 1962 et intitulé « *College admissions and the stability of marriage* ». David Gale et Loyd Shapley y présentent le problème de l'affectation à travers la métaphore du mariage. Ils proposent un modèle d'appariement qui pourrait être utilisé par une agence matrimoniale sophistiquée à laquelle des hommes et des femmes fourniraient leurs préférences sur les candidats de l'autre sexe. Plus exactement, chaque individu qui fait appel à l'agence, homme ou femme, doit fournir un classement exhaustif de l'ensemble des candidats de l'autre sexe. L'agence peut alors utiliser le modèle d'appariement, autrement dit un algorithme, pour constituer des couples

5. Les formations nécessitent en effet qu'un groupe d'élèves ou d'étudiants soient réunis de manière stable pendant une période relativement longue.

à partir de ces deux classements. Bien sûr, chaque individu a la possibilité de refuser le mariage avec un certain nombre de candidats du sexe opposé, ce qu'il spécifie en s'insérant lui-même dans le classement. Ce qui signifie qu'il préfère rester célibataire qu'être marié aux individus situés en deçà de lui-même dans son classement.

Munie de tous ces classements, de chaque homme sur l'ensemble des femmes et de chaque femme sur l'ensemble des hommes, l'agence tente donc d'apparier au mieux les candidats, ce qui soulève une première question importante : Quel est le sens du qualificatif « au mieux » ? Notons tout d'abord qu'il est peu probable que l'agence parvienne à satisfaire au mieux chaque individu. Sauf configuration exceptionnelle, il y aura bien un homme ou une femme qui va se retrouver avec un *alter ego* qui n'était pas son premier choix. Nous retrouvons ici l'une des questions centrales de la TCS : celle de savoir quelles propriétés attendre d'un « bon » mécanisme d'agrégation des préférences.

Gale et Shapley proposent d'utiliser comme critère de satisfaction collective ce qu'ils appellent *la stabilité* de l'appariement. Il s'agit d'éviter que deux individus de sexe opposé, qui ne constituent pas un couple dans l'appariement, soient prêts à *bloquer* celui-ci, ce qui signifie que ces deux individus préfèreraient laisser tomber le partenaire que l'agence leur a trouvé pour se marier ensemble.

Exemple 1.- Soient trois femmes $\{a, b, c\}$ et trois hommes $\{x, y, z\}$ dont les préférences sont les suivantes :

- préférences des femmes :

a	b	c
y	z	y
x	y	z
z	x	x

- préférences des hommes :

x	y	z
a	b	c
b	a	a
c	c	b

Considérons l'appariement $\mu = \{(x : a), (y : b), (z : c)\}$. On peut vérifier que μ est stable : il n'existe aucun couple femme-homme non appariés ensemble où la femme et l'homme préfèreraient être appariés ensemble.

Il est naturel de se poser la question suivante : un appariement stable existe-t-il toujours, quelles que soient les préférences des hommes et des femmes concernés par l'appariement ? À cette question, Gale et Shapley répondent qu'il existe toujours au moins un appariement stable et ils proposent une méthode d'appariement qui, dans toute situation, génère un appariement stable.

Mais si plusieurs appariements stables existent, comment choisir entre eux ? Gale et Shapley ont aussi une réponse à cette question en proposant un second critère, *l'optimalité*. Imaginons plusieurs appariements, tous stables. Pour quelle raison pourrait-on collectivement préférer un de ces appariements aux autres ? Évidemment, si certains préfèrent leur partenaire dans cet appariement plutôt que dans tous les autres appariements stables, alors que tous les autres agents sont indifférents, cet appariement est collectivement préférable et les autres appariements stables peuvent être éliminés. Malheureusement, en général ce n'est pas le cas et tout le monde n'est pas d'accord sur ce qui constitue le « meilleur » appariement stable. Aussi Gale et Shapley proposent-ils de s'intéresser à un critère d'optimalité un peu moins exigeant qui concerne une seule des deux catégories de protagonistes : les hommes ou les femmes. Ainsi, si certains hommes préfèrent leur partenaire dans cet appariement stable plutôt que dans tous les autres appariements stables, alors que tous les autres hommes sont indifférents, cet appariement est *optimal pour les* *hommes* [6].

4.2.3 L'algorithme d'acceptation différée

Après avoir défini ces deux critères de choix collectifs, la stabilité et l'optimalité pour les hommes ou pour les femmes, Gale et Shapley montrent qu'il existe un modèle d'appariement qui permet toujours d'obtenir un appariement qui les satisfait. Ils l'appellent l'*algorithme d'acceptation différée* (AAD). Cet algorithme fonctionne de la façon suivante. Imaginons que l'agence dispose donc d'un groupe d'hommes et d'un groupe de femmes et qu'elle a pour chaque individu le classement exhaustif de chaque individu du sexe opposé. Peu importe que le nombre d'hommes et de femmes soit le même. Imaginons que l'agence veuille donner la priorité aux femmes, c'est-à-dire qu'elle va d'abord considérer les classements des femmes sur les hommes. Comment va-t-elle procéder ?

Première étape :

(a) L'agence associe à chaque femme l'homme qu'elle préfère. (Pour celles qui préfèrent le célibat, la procédure s'arrête ici.)

(b) L'agence attribue *temporairement* à chaque homme, parmi les femmes qui l'ont classé en premier, celle qu'il préfère. Les autres offres qui lui ont été faites sont rejetées.

Aux étapes suivantes :

(a) Pour chaque femme dont la proposition a été rejetée à l'étape précédente, l'agence l'associe à son partenaire préféré parmi ceux qui ne l'ont pas encore rejetée. (À moins qu'elle préfère le célibat à cette étape, auquel cas la procédure s'arrête ici.)

(b) L'agence attribue *temporairement* à chaque homme la partenaire qu'il préfère, en considérant toutes celles qui viennent de lui être associées, ainsi que la partenaire à laquelle il était apparié à l'étape précédente. Les autres offres sont rejetées.

La procédure s'arrête quand chaque femme est appariée à un homme (ou reste célibataire). Les appariements sont alors définitifs.

Gale et Shapley montrent que cet algorithme, qui donne la priorité aux femmes, engendre toujours un appariement stable et optimal pour les femmes. (Bien entendu, le résultat symétrique vaut pour l'algorithme qui donne la priorité aux hommes). Comprenons bien ce que cela signifie : Quel que soit le groupe d'hommes, le groupe de femmes et leurs classements, l'agence est sûre de pouvoir au final proposer une liste de mariages qui, non seulement est stable et ne pourra pas être bloquée par un couple non satisfait, mais qui donne en plus les meilleurs mariages stables que les femmes (ou les hommes) pourraient espérer. Notons, en outre, que l'algorithme qui donne la priorité aux femmes aboutit à une affectation qui est *faiblement Pareto-optimale* pour les femmes : il n'existe aucune autre affectation (stable ou pas), qui soit strictement préférée par toutes les femmes [7].

6. Évidemment, on définit de la même façon l'optimalité pour les femmes, nous avons seulement ici gardé les choix de présentation de Gale et Shapley.

7. Roth & Sotomayor (1990), Théorème 2.27. En revanche, il se peut que cette affectation ne soit pas *fortement* Pareto-optimale : il peut exister une autre affectation que toutes les femmes trouvent au moins aussi bonne et certaines femmes strictement meilleures.

4.2.4 De l'appariement matrimonial à l'affectation scolaire

L'appariement matrimonial peut sembler bien éloigné du problème de l'affectation des lycéens aux établissements de l'enseignement supérieur. En fait GALE et SHAPLEY montrent dans leur article que l'algorithme se généralise sans difficulté à des problèmes d'affectation de plusieurs candidats d'une catégorie à un seul candidat de l'autre catégorie (appariement « plusieurs-1 »). Ses bonnes propriétés sont conservées si les femmes (ou les hommes) sont polygames et l'algorithme peut donc être utilisé pour affecter des collégiens à des lycées, ou des lycéens à des établissements de l'enseignement supérieur pour les filières sélectives.

Néanmoins, si l'on s'intéresse à de tels problèmes d'affectation, certaines simplifications du modèle de GALE et SHAPLEY doivent être considérées, et elles l'ont été plus tard, par d'autres auteurs.

La première de ces simplifications concerne les informations fournies au mécanisme d'affectation. Dans l'appariement matrimonial, nous avons supposé que l'agence connaît les vraies préférences des hommes et des femmes, et que ceux-ci sont parfaitement sincères quand ils déclarent leurs classements. Or, dans la réalité, les choses ne sont pas si simples. Les lycéens peuvent penser avoir intérêt à ne pas révéler leurs vraies préférences à APB. Ils peuvent penser utile d'essayer de manipuler le résultat en fournissant à l'algorithme un classement non sincère. Peu importe qu'ils soient ou non en mesure d'améliorer ainsi leur situation, s'ils ne révèlent pas leurs vraies préférences, les bonnes propriétés de l'algorithme de GALE et SHAPLEY ne sont plus nécessairement satisfaites. Nous avons déjà expliqué pourquoi en abordant la TCS : les résultats concernant la stabilité et l'optimalité de l'affectation finale ne sont valables que si ce sont les mêmes préférences qui servent à informer l'algorithme et à évaluer la « qualité » de l'affectation finale. Ce n'est pas le cas quand les individus ne sont pas sincères. Par ailleurs, dans le contexte spécifique de l'affectation scolaire, un argument supplémentaire en faveur de l'importance de la non-manipulabilité a été avancé : s'il est possible de manipuler la procédure d'affectation, alors les familles mieux informées (sur les vœux probables des autres familles ainsi que sur les propriétés de la procédure) risquent d'être avantagées par rapport aux familles qui le sont moins.

Le théorème d'impossibilité de Gibbard-Satterwhaite laisse présager que les algorithmes d'appariement seront généralement manipulables. Dans un article publié en 1982 [Rot-82], Alvin ROTH montre en effet que, conformément au résultat général de la théorie du choix social, il n'existe pas de modèle d'appariement stable et non manipulable par les deux types. Mais il montre aussi, et ce résultat est beaucoup plus intéressant pour le problème pratique de l'affectation des collégiens et des lycéens, que lorsqu'on se sert de l'algorithme (AAD) de GALE et SHAPLEY, l'appariement qui en résulte est stable et que les candidats de la catégorie qui est prioritaire ont intérêt à être sincères. Bien sûr, rien ne dit que, lors d'une application de l'algorithme, les membres de la catégorie prioritaire seront pour autant sincères ; et s'ils ne le sont pas, on risque de perdre les bonnes propriétés de la procédure. Cependant le résultat d'Alvin ROTH est encourageant, car il signifie que l'organisateur de l'affectation (par exemple, le Ministère de l'Éducation Nationale) peut sincèrement dire aux membres de la catégorie prioritaire qu'il est dans leur intérêt d'être sincère.

Par contre, les candidats de la catégorie non-prioritaire n'ont pas toujours intérêt à être sincères. Ce n'est pas forcément grave si on a des raisons de penser que les membres de cette catégorie sont sincères. Dans le cas des algorithmes d'affectation scolaire, on peut justement s'attendre à ce que les deux catégories en question (les collégiens ou lycéens d'un côté, les établissements de l'autre) diffèrent du point de vue stratégique. Si un lycéen pense avoir intérêt à mentir sur ses préférences, rien ne l'empêche de le faire. Du côté des établissements, la situation est plus complexe. Il peut y avoir des cas où les établissements forment leurs préférences en toute autonomie de l'organisateur de la procédure d'affectation, et où ils n'ont pas d'obligation particulière d'être sincères. Dans d'autre cas, ce peut être l'organisateur lui-même qui détermine leurs préférences, et alors le problème de manipulation disparaît. Dans ce cas, la littérature sur l'appariement parle d'ailleurs parfois plutôt des « priorités » des établissement que de leurs « préférences », pour indiquer le fait que ces priorités sont déterminées par l'organisateur. Un exemple est fourni par la procédure AFFELNET pour les collégiens, où les rectorats organisent l'affectation et déterminent les préférences des établissements (qui se réduisent aux « points » attribués aux collégiens selon un barême pré-établi). Notons que les propriétés d'optimalité et de non-manipulabilité sont liées : c'est la même catégorie (la catégorie prioritaire) pour laquelle l'algorithme est optimal et non-manipulable.

Mentionnons une dernière propriété intéressante de l'algorithme AAD dans le contexte des affectations scolaires. Nous avons insisté sur le fait que, souvent, les « préférences » des établissements sont l'expression des objectifs politiques des organisateurs de l'affectation. On s'attend donc à ce que, plus un élève progresse dans les préférences des établissements, plus il se trouve apparié à un établissement qu'il préfère. Par exemple, si les préférences des établissements sont déterminés par un barême, et que les points accordés par celui-ci croissent avec les résultats scolaires moyen d'un élève, alors on s'attend à ce que, toutes choses étant égales par ailleurs, si l'élève i obtient de meilleurs résultat, il sera apparié à un meilleur établissement (de son point de vue). On dit dans ce cas que la méthode « respecte les améliorations » (Balinski & Sonmez 1999) [BS-99] ou « respecte les priorités » (Hiller & Tercieux 2014) [HT-14]. Or, Balinski & Sonmez (1999) ont montré que l'algorithme AAD qui donne priorité aux élèves est la seule procédure stable qui respecte les améliorations.

Les économistes Hiller et Tercieux (2014) ont montré qu'AFFELNET est probablement conforme à l'algorithme de Gale et Shapley qui donne la priorité aux collèges. Ce n'est pas un choix qui va de soi, compte tenu du fait que, l'organisateur déterminant les préférences des établissement, le risque de manipulation vient plutôt de l'autre catégorie (les collégiens). Par ailleurs, il semble naturel de considérer que l'objectif de la procédure d'affectation est plutôt de favoriser les préférences des élèves que celles des établissements. C'est la raison pour laquelle l'algorithme qui donne priorité aux élèves est celui qui a été choisi pour déterminer l'affectation dan les écoles de New York (2003) et de Boston (2005) [8].

4.2.5 Classements incomplets, indifférences et tirage au sort : le cas des formations non sélectives

La seconde simplification importante effectuée par l'algorithme de Gale et Shapley consiste à supposer que chaque membre de chaque catégorie transmet un classement complet et strict des options. Cela signifie que le classement d'un membre fait figurer chaque membre de l'autre catégorie, et qu'il n'y a pas d'indifférence. Cela soulève deux problèmes. Le premier est qu'il n'est pas toujours réaliste d'exiger un classement complet. Par exemple, on n'exigera certainement pas que chaque élève ou étudiant fournisse un classement complet de tous les établissements : on demande généralement aux intéressés un « classement tronqué ». Cela peut affecter la qualité des affectations, et engendrer de nouvelles incitations à ne pas révéler sincèrement ses préférences [9] – l'idée étant alors de ne pas faire figurer dans son classement tronqué des établissements supposés très demandés.

Le second problème se situe plutôt du côté des établissements : typiquement, les classements des élèves par les établissements sont des classements à granularité large, avec de très larges zones d'indifférence. La raison pour laquelle il en est ainsi est que les critères utilisés pour effectuer ces classements sont eux-mêmes assez grossiers (demander un établissement dans la zone de rattachement du domicile ou non, être boursier ou pas, etc). Une manière courante de se

8. Voir les références dans Erdil & Ergin 2008 [EE-08].
9. Voir Hiller & Tercieux (2014) [HT-14], p. 638.

rapprocher de l'algorithme de GALE et SHAPLEY consiste à « casser les indifférences » en *tirant au sort* pour obtenir des préférences strictes. Par exemple, si l'établissement x classe au même niveau l'élève a et l'élève b, on tirera au sort pour déterminer si l'algorithme prendra comme hypothèse effective que x classe a devant b (ou b devant a). Cela ne va pas sans difficultés. Il se peut en effet que cette manière de casser l'indifférence résulte dans l'affectation de a mais pas de b dans l'établissement x. La décision peut ne pas être facile à accepter par le candidat b s'il se retrouve affecté dans un établissement x' auquel il préfère x. La déception peut facilement alimenter la méfiance à l'égard du processus par lequel l'indifférence a été cassée.

Toutes les bonnes propriétés de l'algorithme de GALE et SHAPLEY ne sont cependant pas perdues en tirant au sort. Par exemple, la stabilité est conservée : on ne peut pas trouver un élève a et un établissement x tels que si a n'est pas affecté à x alors a préfère strictement x à l'établissement auquel il est affecté et x préfère strictement a à l'un des élèves qui lui sont affectés. Il n'en va pas de même de l'optimalité pour les élèves. Non seulement l'affectation choisie par l'algorithme n'est plus nécessairement meilleure pour tous les élèves que toutes les autres affectations stables [10] ; mais il se peut même qu'une autre affectation stable soit au moins aussi bonne pour tous les élèves et strictement meilleure pour au moins un des élèves (on dit que cette autre affectation stable la *Pareto-domine*). C'est ce que montre l'exemple 2 [11].

Exemple 2.- Soient trois candidats a, b, c et trois établissements x, y, z dotés chacun d'une place, dont les préférences sont les suivantes :

- préférences des élèves :

a	b	c
y	z	y
x	y	z
z	x	x

- préférences (ou « priorités ») des établissements :

x	y	z
a b, c	a, b, c	a, b, c

Supposons maintenant que l'on casse les indifférences des établissements de la manière suivante :

x	y	z
a	b	c
b	a	a
c	c	b

Dans ce cas, l'affectation proposée par l'algorithme qui donne la priorité aux élèves est $\mu = \{(x : a), (y : b), (z : c)\}$. Or, l'affectation $\nu = \{(x : a), (y : c), (z : b)\}$ est stable elle aussi, et elle Pareto-domine μ : en effet a est indifférent entre μ et ν mais b et c préfèrent strictement les établissements que ν leur affecte à ceux que μ leur affecte. Dans un tel cas, les économistes disent qu'il y a alors une « perte d'efficacité ». ABDULKARIDOGLU, PATHAK & ROTH (2009) [APR-09] estiment que, lors des quatre premières années de fonctionnement de l'algorithme d'affectation des élèves mis en place à New York, autour de 2,5 % des élèves [12] auraient pu bénéficier d'une meilleure affectation (du point de vue de leurs propres préférences) sans nuire aux autres élèves. On peut modifier l'algorithme d'affectation de manière à retenir une affectation stable et Pareto-optimale parmi les affectations stables. Mais ERDIL & ERGIN (2008) [EE-08] ont montré qu'un tel algorithme est manipulable.

10. Au-delà de l'algorithme de GALE et SHAPLEY, il n'est plus vrai qu'une affectation stable et optimale pour les élèves existe nécessairement. C'est un résultat dèjà connu pour le problème des mariages.
11. On s'est inspiré ici d'un exemple de ERDIL & ERGIN 2008 [EE-08], p. 671.
12. Ce qui représente quand même environ 1 700 élèves par an.

Les difficultés soulevées par la présence d'indifférences sont donc profondes : elles interdisent l'heureuse convergence de l'optimalité et la non-manipulabilité, qui vaut quand les préférences sont strictes [13].

La situation décrite dans l'exemple précédent est d'autant plus regrettable que les élèves b et c pourraient échanger leurs affectations sans objection de la part des établissements, qui sont parfaitement indifférents entre ces deux élèves. On peut considérer ce problème comme une instabilité d'un type différent de ce qui a été utilisé jusque-là. Ici, l'appariement est instable parce que deux élèves et deux établissements préfèreraient une autre affectation. Plus exactement, les deux élèves préfèreraient strictement une autre affectation et les établissements seraient indifférents. On peut alors envisager un mécanisme qui permettrait aux élèves d'échanger leurs établissements d'affectation. Il faudrait pour cela que les choix et tirages au sort effectués par les établissements soient parfaitement publics. Ainsi les élèves pourraient savoir quand ils peuvent avoir recours au mécanisme d'échange.

4.3 Le cas des pseudo-classements

Les filières non sélectives n'ont pas le droit d'étudier les dossiers pour établir un ordre de préférence. Cependant nous avons vu que l'algorithme de Gale et Shappley repose sur deux tableaux de vœux : les vœux des candidats et les « vœux » des filières. On doit donc fournir un **pseudo-classement** de la part des filières pour que l'algorithme puisse se dérouler.

Comment effectuer le pseudo-classement ? Est-il possible de l'effectuer de façon à ce que la stabilité soit sauvegardée ?

Nous nous intéresserons, dans une première étape, à la méthode de pseudo-classement choisie pour insister sur une conséquence. On obtient bien ainsi un appariement mais celui-ci n'est pas stable : le candidat C_1 peut être affecté dans la filière f_1 et le candidat C_2 dans la filière f_2 alors que C_1 aurait préféré f_2 et C_2 la filière f_1. On a cependant une *pseudo-solution* puisqu'on peut faire l'échange sans dommage pour personne.

4.4 Conclusion

4.5 Bibliographie

[APR-09] Abdulkadiroglu, Atila & Pathak, Parag A. & Roth, Alvin E., *Strategy-Proofness versus Efficiency in Matching with Indifferences : Redesigning the NYC High School Match*, **American Economic Review**, vol. 99, n^o 5, December 2009, pp. 1954–78.

[Arr-51] Arrow, Kenneth J., **Social Choice and Individual Values**, J. Wiley / Chapman & Hall, 1951. Seconde édition, Yale University Press, 1963. Traduction française **Choix collectifs et préférences individuelles**, Calmann Lévy, 1974 ; réédition en poche, Diderot éditeur, 1997.

[BS-99] Balinksi, M. & Sonmez, T., *A Tale of Two Mechanisms : Student Placement*, **Journal of Economic Theory**, vol. 84, 1999, pp. 73–94.

13. Sur cette tension, voir aussi Abdulkadiroglu, Pathak & Roth (2009) [APR-09].

[EE-08] ERDIL, Aytek & ERGIN, Haluk, *What's the Matter with Tie-Breaking ? Improving Efficiency in School Choice*, **American Economic Review**, vol. 98, n^o 3, June 2008, pp. 669–89.

[FHI-13] FORGES, HAERINGER & IEHLÈ, *Appariement : des modèles de Lloyd Shapley à la conception de marchés d'Alvin Roth*, **Revue d'économie politique**, Dalloz, 2013/5, vol. 123, pp. 663–696.

[GS-62] GALE, D. and SHAPLEY, L. S., *College admissions and the stability of marriage*, **Amer. Math. Monthly**, vol. 69, 1962, pp. 9–15. Numérisé par *JStor*.

[HT-14] HILLER, V. & TERCIEUX, O., *Choix d'écoles en France : Une évaluation de la procédure Affelnet*, **Revue économique**, vol. 65(3), 2014, pp. 619–655.

[Rot-08] ROTH, Alvin E., *What Have We Learned from Market Design ?*, **The Economic Journal**, 2008, vol. 118, pp. 285–310.

[Rot-82] ROTH, Alvin E., *The Economics of Matching : Stability and Incentives*, **Mathematics of Operations Research**, vol. 7, n^o 4, Nov. 1982, pp. 617–628.