

DAVID LEWIS ET LA RATIONALITÉ DES CONVENTIONS

Philippe BATIFOULIER, Jacques MERCHERS, Daniel URRUTIAGUER,

Revue de philosophie économique, 2002, n°6, pp 37-56, Vuibert

Résumé:

Lewis se propose de donner un sens à la notion de convention de langage fondée sur la coordination des activités d'agents rationnels. Mais il doit pour cela utiliser une version de la théorie des jeux de coordination qui exige l'emploi de concepts problématiques comme celui de savoir commun de la rationalité des agents. De ce fait, la liaison qu'il tente d'établir entre convention et rationalité semble à la fois trop forte et trop faible. Deux séries de difficultés opposées surgissent en effet. La première souligne les obstacles à faire de la conformité à une convention de Lewis un comportement purement rationnel. La seconde montre que la convention de Lewis exclut les jeux de coordination dans lesquels la rationalité individuelle conduit à un choix sous-optimal. **Mots clefs** : David Lewis ; convention ; rationalité ; jeux de coordination.

DAVID LEWIS ET LA RATIONALITÉ DES CONVENTIONS¹

Le philosophe David Lewis, récemment décédé, fit partie du courant de la philosophie analytique qui a rassemblé, dans les années 1960-70, linguistes, philosophes et logiciens dans la construction commune d'une sémantique pour les langues naturelles². David Lewis y a participé par des articles et des ouvrages sur la sémantique générale et les contrefactuels³ mais également par « *Convention* » (1969), un ouvrage bien connu des économistes français. Son idée était de rendre compte de la signification à partir de la notion de convention de langage. Une telle idée ne se ramène naturellement pas au lieu commun selon lequel le langage est conventionnel parce qu'un mot qui signifie telle chose aurait pu en signifier une autre. Lewis se proposait plutôt d'expliquer les propriétés sémantiques du langage naturel au moyen d'un concept particulier de convention qui s'écarte de son sens ordinaire. Ce concept pouvait permettre selon lui d'analyser la conventionnalité d'un grand nombre d'activités humaines auxquelles s'intéressent les sciences sociales. Cela explique la diversité des exemples choisis pour illustrer ses analyses.

1. Nous remercions G. de Larquier et F. Gannon (Forum, Paris X) ainsi que les rapporteurs anonymes

pour leurs remarques critiques qui ont permis d'améliorer l'article.

2. Voir sa nécrologie dans *Mind*, Vol. 111, n° 441, January 2002.

3. Respectivement « *General Semantics* » (1972) et « *Counterfactuals* » (1973).

Sa démarche a été très critiquée et d'abord par Quine⁴ dans la préface même de l'ouvrage avec l'argument célèbre sur l'impossibilité d'expliquer l'apparition du langage au moyen d'une *Convention*, qui requiert un langage pour pouvoir être exprimée⁵. Depuis la parution de l'ouvrage, beaucoup d'autres auteurs ont contesté la démarche de Lewis en prétendant par exemple, qu'il existe bien des conventions de langage mais plus simples que celles de Lewis (Millikan, 1998), que nous n'avons pas besoin de conventions pour expliquer le phénomène du langage (Winch, 1988), que si la convention explique l'usage du langage, elle ne permet pas de fonder la signification (Davidson, 1985) ou encore, que les sémantiques basées sur la convention manquent de la généralité nécessaire pour expliquer les propriétés des langues naturelles (Laurence, 1996).

4. Qui a encadré son Ph. D. obtenu en 1967.

5. Quine avait déjà énoncé un paradoxe analogue en logique dans son article « Truth by Convention »: "In a word, the difficulty is that if logic is to proceed 'mediately' from conventions, logic is needed for inferring logic from the conventions" (1936, p.104).

Cependant, notre intérêt dans cet article, ne vise pas l'analyse du langage mais certaines notions que Lewis utilise pour construire son concept de convention et qui, selon nous, en limitent l'intérêt. Il en va notamment du concept de *common knowledge* (savoir commun ou **CK**), que Lewis a défini le premier, et qu'utilisent de nombreux économistes pour justifier la conception de la rationalité des agents économiques où chacun agit compte tenu de la rationalité de l'autre. Or, il est insuffisant de postuler la connaissance par chacun de la rationalité de l'autre. « *En réalité, pour construire une théorie du comportement économique fondée sur la notion de rationalité, il faut même supposer davantage, à savoir que la rationalité de tous les agents doit appartenir au fond commun de la société, pour utiliser la terminologie introduite par le philosophe David Lewis. Chaque agent doit en effet savoir que les autres... sont rationnels mais encore que chacun des autres sait que tous les autres sont rationnels, que tous savent que tous sont rationnels, etc.* » (Arrow, 1987, p. 34).

Ce concept de **CK**, devenu ressource patrimoniale de l'économie, n'est cependant pas utilisé par Lewis pour nourrir une théorie du comportement écono-

mique rationnel mais pour construire une théorie des conventions. La convention est pensée comme un moyen de coordination offert à des individus parfaitement rationnels et sachant que les autres le sont également. Elle est sous-tendue par le CK de la rationalité. Le succès en économie de l'approche lewisienne doit beaucoup à cet ancrage de la notion de convention dans le champ de la rationalité.

Nous nous proposons dans cet article d'étudier le bien fondé de la liaison entre convention et rationalité (au sens économique du terme). La première partie explicite le projet de Lewis, souvent méconnu des économistes. La deuxième partie s'intéresse au moyen utilisé par Lewis pour réaliser son projet : la théorie des jeux de coordination. Ce moyen illustre le lien intime entre la convention à la Lewis et l'hypothèse de rationalité. Or, ce lien mérite d'être explicité dans la mesure où il donne une définition particulière de la convention. La dernière partie s'interroge sur le bien fondé de l'arrimage de la convention à la rationalité. Elle souligne les difficultés à faire du suivi de la convention un comportement purement rationnel et présente, dans cette perspective, les critiques tant internes qu'externes à la thèse de Lewis. Elle poursuit ce questionnement en s'interrogeant sur la possibilité de suivre rationnellement des mauvaises conventions.

1 Le projet de Lewis : répondre au scepticisme de Quine

Lewis dévoile le sens de sa démarche dans la conclusion de son livre. Quine, nous dit-il, ne conteste pas seulement la notion de convention de langage mais aussi les idées d'analyticité, de synthéticité, de contradiction, d'implication ou de synonymie. Or, toutes ces notions sont interdéfinissables si bien que déterminer la signification de l'une d'entre elles aboutit à leur donner à toutes une signification⁶. Lewis a choisi de définir l'analyticité pour donner un sens à la notion de convention de langage (1.1) en tant que convention particulière fondée sur la coordination des activités d'agents rationnels (1.2).

⁶ C'est l'idée de l'analyse non réductive qui ne cherche pas à réduire un concept à un autre mais à éclairer leurs relations, cf. Avramides (1997)

1.1. La convention comme fondement du langage et le problème de l'analyticité

Lewis veut croire que Quine (1951) n'a pas dressé d'obstacles insurmontables à l'analyse des propriétés sémantiques du langage dans son célèbre article sur les deux dogmes de l'empirisme mais posé deux exigences rédhibitoires. La première est de respecter le holisme sémantique et la seconde de fournir une

explication qui ne soit pas fondée sur les intentions des locuteurs mais sur les comportements des utilisateurs du langage.

Quine (1951, "Two Dogmas of Empiricism") affirme que la vérité d'un énoncé dépend à la fois du langage et du monde mais il se refuse à en déduire qu'il serait possible de réduire la composante factuelle à zéro de sorte que certains énoncés, dits analytiques, ne dépendraient que du langage et seraient d'une autre nature que les énoncés dont la vérité dépend de ce que nous observons dans le monde. Il refuse également de parler du contenu factuel dans la vérité d'un énoncé individuel car il considère que l'unité minimale de signification n'est ni le mot ni la phrase mais l'ensemble de la science. C'est ce qu'on appelle le holisme de la signification⁷. Ce holisme sémantique se double d'un holisme épistémologique selon lequel aucune expérience particulière n'est liée directement à un énoncé particulier mais peut affecter notre corps de connaissance à n'importe quel endroit. Aucun énoncé ne peut être considéré comme définitif. Il en résulte immédiatement qu'aucun énoncé ne peut être tenu pour analytique. Ce « holisme radical » a été par la suite amendé par Quine (1960, *Word and object*) qui a admis que l'unité de signification pouvait consister en segments individuels de science et que des énoncés observationnels (du type « ce drapeau est rouge ») pouvaient être confirmés ou infirmés directement et indépendamment (Stroll, 2000, p. 200).

⁷ Dans cette perspective, estime Quine qui cite Meyerson, « l'ontologie fait corps avec la science » (*ibid.*, p.45) et il faut rejeter la neutralité ontologique de Carnap qui pensait qu'un langage de description pourrait être compatible avec plusieurs ontologies.

Il semble alors légitime de se demander s'il n'est pas possible de donner un sens à l'analyticité en maintenant une forme de holisme sémantique. On peut douter qu'il soit possible d'expliquer des concepts sémantiques en termes purement comportementaux et considérer par exemple qu'il est impossible pour un énoncé analytique d'être faux (*ibid.*, p. 205). Si on convient de considérer les différentes manières dont le monde aurait pu être comme des mondes possibles, nous obtenons la définition de Lewis de l'analyticité comme vérité dans tous les mondes possibles pour un langage donné. Ce n'est toutefois pas la fin de l'histoire car le second réquisit quinién exige de "sortir du cercle intentionnel" (Mac Dermott, *ibid.*, p. 1983). Quine (*ibid.*, p. 24) entend par là qu'une explication claire de la synonymie par exemple doit se référer à un comportement, verbal en l'occurrence, et non aux intentions des locuteurs. De même, "clarifier" l'analyticité exige de faire intervenir dans l'explication des

"*mental or behavioral or cultural factors*" (*ibid.*, p. 36) qu'il faut comprendre, assure Mac Dermott comme l'exigence de relier toute explication des propriétés sémantiques à des comportements linguistiques observables.

Lewis interprète cette exigence comme consistant à définir l'analyticité dans la langue d'une population déterminée (1969, *Convention*, p. 205). Il rejette donc l'explication de la signification fondée sur la reconnaissance des intentions du locuteur soutenue par les philosophes dans la ligne de Grice ou Searle⁸. Il montre qu'une explication fondée sur les conventions capte également les aspects intentionnels de la signification (p. 152-159) à condition d'utiliser un concept particulier de *Convention*, la convention de coordination. Ce concept lui permet de compléter sa justification de l'analyticité en donnant un sens à l'analyticité non seulement dans un langage mais dans une population reliée à ce langage par des conventions de coordination.

8. Comme le montre Avramides (*ibid.*), l'opposition entre Grice et Lewis n'est cependant pas aussi tranchée que ce dernier l'affirme.

Pour arriver à ce résultat, il développe une stratégie consistant à construire un modèle simple de communication qu'il étend progressivement jusqu'à englober les traits fondamentaux du langage ordinaire. Un agent communique avec un autre agent lorsqu'il cherche à lui transmettre une information sur l'état du monde en se servant de signaux conventionnels. Selon ce modèle, réussir à communiquer c'est atteindre un équilibre de coordination, quand la situation des deux agents ne peut être améliorée si n'importe lequel des deux mais pas les deux avait agi différemment (Coleman, 1989).

Pour l'illustrer, Lewis (1969, *Convention*, p.125-126) s'appuie sur le code maritime international. Par exemple le choix d'un drapeau de couleur C pour signaler le transport de matières explosives permet aux bateaux d'identifier la situation et d'adopter la meilleure stratégie mutuelle pour éviter un accident en se croisant. Signaux et état de choses sont reliés par une convention reposant dans ce cas sur un accord explicite. Les deux parties se conforment à cette convention de coordination car chacune s'attend à ce que la rationalité de l'autre lui permette de tenir le même raisonnement.

Ce type de coordination ne met en jeu que des comportements rationnels et réguliers de conformité à une convention préalable dans une situation caractérisée par un problème de coordination. Un tel problème peut recevoir plusieurs solutions puisqu'il est possible d'imaginer un autre code maritime employant des

drapeaux d'autres couleurs pour signaler les mêmes états de choses. Lewis montre alors que ce modèle peut être généralisé de façon à rendre compte de l'usage de fragments significatifs du langage.

Tout d'abord, une convention peut s'établir par accord préalable comme dans le cas des drapeaux maritimes mais aussi du fait de comportements conformes à des précédents ou s'inspirant de traits particulièrement saillants de la situation. Il peut ensuite exister dans une population un grand nombre de conventions établies pour résoudre une multitude de problèmes de coordination. Enfin, un signal peut être purement verbal tout comme sa réponse. Si la population échange des signaux verbaux selon un grand nombre de conventions et face à de multiples problèmes de coordination, cela ne commence-t-il pas à ressembler à l'usage d'un langage, se demande Lewis (1969, *Convention*, p. 142) ?

On peut enfin imaginer un usage opportuniste de la convention avec un bateau sans explosifs qui arbore le drapeau C pour dissimuler un trafic illégal. Pour empêcher un usage dévoyé de la *Convention*, Lewis imagine une autre convention réglant l'usage des systèmes de communication. L'émetteur doit être sincère en envoyant le signal qui indique un état des choses et son destinataire doit lui faire confiance.

C'est la convention de sincérité (*truthfulness*) (1969, *Convention*, p. 148)⁹. En indiquant qu'elle est une régularité d'action dans un langage particulier, Lewis montre qu'elle ne correspond pas à une exigence générale de dire la vérité qui n'admettrait pas d'alternative conventionnelle.

9. Cette vision a été remaniée dans différents articles dont le dernier a été publié dans ses *Philosophical Papers* de 1983 et traduit dans l'article de 1993. Les conventions qui relient les langages aux populations deviennent dans cet article des conventions de vérité et, ajoute Lewis, de confiance. La coordination par l'usage du même langage ne repose plus seulement sur des raisons pratiques en suivant les autres, mais aussi épistémiques avec la confiance accordée par l'auditeur à son locuteur. Elle s'établit "*entre l'action d'un locuteur véridique et la croyance par laquelle répond son auditeur en se fiant à lui*" (1993, p. 15-17, "*Langages et langage*", *Réseaux*, n° 62, p. 11-18.).

La conception de l'analyticité découle des liens entre la vérité et le monde. Si j'indique dans ma langue que Dupont est célibataire, alors, je veux dire que dans le monde où je me situe, l'état de choses est tel que Dupont n'est pas marié¹⁰. Dans un monde un peu différent, il pourrait y avoir un Dupont légèrement différent¹¹ et la même phrase indiquerait toujours qu'il est

non marié. Les vérités nécessaires (comme il est nécessaire qu'un célibataire soit non marié) ne sont pas créées par des conventions, même si elles ont besoin pour être énoncées d'un langage qui pourrait être différent. Tout ce qu'une convention humaine peut faire est de sélectionner une expression verbale plutôt qu'une autre, pour lui octroyer le privilège d'être vraie en vertu de tous les faits possibles. L'analyticité est alors la vérité dans tous les mondes possibles dans une langue reliée à une population par des conventions de langage. Quine reprochera sans doute à Lewis de faire reposer sa théorie sur la notion de monde possible. On peut aussi se demander, lorsqu'on s'aperçoit que sa définition formelle de la convention repose sur des hypothèses de connaissance et de rationalité très discutées, si le prix à payer pour la conformité aux exigences de Quine n'est pas trop élevé.

10. Notons toutefois que le terme bachelor (non marié et en âge de l'être selon l'O.E.D) est plus restrictif que célibataire.

11. Dans l'ouvrage *Convention*, Lewis reconnaît avoir utilisé librement la notion de monde possible. Il règle, par exemple, de façon « *drastique* » la question de l'individuation dans les mondes possibles en supposant qu'aucun individu ne peut être dans deux mondes possibles. Il remarque toutefois que la notion de contrepartie (un individu du monde actuel possède des contreparties dans les mondes possibles) serait peut-être préférable (Lewis, 1969., p. 208). Sur ce point, voir Lewis (1968, "Counterpart theory and quantified modal logic", *Journal of philosophy*).

1.2 La convention de coordination : définition formelle

Lewis formalise sa notion de convention dans le cadre des jeux de pure coordination quand les intérêts des agents en interaction coïncident parfaitement ou presque, à la différence des jeux de pur conflit. Avant d'analyser en détail ce type de jeux (cf. infra 2ème partie), il faut dégager deux conditions découlant du choix théorique de Lewis qui ont été très contestées.

Tout d'abord, la notion de convention est construite dans le cadre de la théorie de la décision appliquée à des agents rationnels dont l'action résulte du désir de se conformer à une régularité R et de la croyance ou attente que les autres se conformeront à R, dans une situation où existe au moins une alternative R'. Ensuite, le maintien de la convention dans une population P exige que les membres de cette population sachent que chacun se conforme à la régularité R et non à une alternative R' et que chacun sache que les autres savent qu'il sait, etc. C'est la célèbre condition de savoir commun¹². Lewis (1993, "*Langages et langage*") résume les conditions pour qu'une régularité R dans l'action, ou dans

l'action et la croyance, constitue une convention dans une population P par six propositions¹³ :

- (1) chacun se conforme à R ;
- (2) chacun croit que les autres se conforment à R ;
- (3) cette croyance que les autres se conforment à R donne à chacun une bonne et décisive raison pour se conformer lui-même à R ;
- (4) tous préfèrent une conformité générale à R plutôt qu'une conformité légèrement moindre que générale ;
- (5) il existe au moins une alternative R' telle que la croyance que les autres s'y sont conformés donnerait à tout le monde une bonne et décisive raison, pratique ou épistémique, de s'y conformer également ;
- (6) les différents faits énumérés dans les conditions (1) à (5) sont affaire de savoir commun.

12. La connaissance commune est d'ordre 1 quand chaque joueur est rationnel et sait que l'autre l'est aussi, mais ignore si l'autre sait qu'il est lui-même rationnel. Elle est d'ordre 2 si chaque joueur sait aussi que l'autre sait qu'il est rationnel sans être assuré que l'autre sait qu'il le sait. Une connaissance commune d'ordre fini est appelée savoir mutuel. La notion de savoir commun (CK) est réservée à une connaissance commune d'ordre infini. Il définit un système d'anticipations du type « je sais que tu sais que je sais, etc. » jusqu'à l'infini.

13. Nous présentons ici sa dernière version des conditions pour définir une *Convention*, reproduite dans l'article de 1993 (cf. note 9).

La convention de Lewis est donc une régularité de comportement fondée sur le savoir commun relatif aux préférences individuelles pour la conformité. Cette focalisation réfléchie sur un choix arbitraire constitue la base d'un ordre social spontané. Sa vision rejoint ainsi les deux hypothèses de la théorie économique néo-classique¹⁴.

14. Selon Arrow (1974), les deux piliers de la science économique néoclassique sont la coordination des comportements par le marché et la rationalité des individus exprimée par un mécanisme d'optimisation sous contrainte.

2 Convention et théorie des jeux de coordination

La reconnaissance du travail de Lewis au-delà de son propre champ de recherche, et en particulier en économie, est imputable, en grande partie aux outils qu'il a mobilisés : les jeux de coordination de Schelling

basés sur la notion de point focal (2.1). La compréhension du concept de convention de Lewis suppose de le comparer avec la vision de Schelling (2.2).

2.1 La coordination par point focal

Comme Lewis s'intéresse aux problèmes de coordination où les individus ont un avantage mutuel à atteindre un résultat donné, il recourt à la théorie des jeux de coordination définie par Schelling (1960) tout en explicitant les mécanismes de coordination mis en jeu. « *My theory of convention had its source in the theory of games of pure coordination – a neglected branch of the general theory of games of von Neumann and Morgenstern, very different in method and content from their successful and better known theory of games of pure conflict. Coordination games have been studied by Thomas C. Schelling, and it is who supplied me with the makings of an answer to Quine* » (Lewis, 1969, *Convention*, p. 3).

Les situations de conflit armé constituent le point de départ de Schelling (1960). Dans une guerre, les belligérants peuvent avoir intérêt à une « *accommodation mutuelle* » en refusant certains moyens comme l'emploi de l'arme chimique pendant la seconde guerre mondiale. Ces situations de « guerre limitée » (*limited war*) relèvent d'une « *stratégie du conflit* » qui nécessite un « *marchandage tacite* » dans lequel la communication est incomplète ou impossible. On est bien là au cœur du thème de recherche de Lewis : se coordonner sans communiquer.

Ces situations, souligne Schelling, peuvent se rencontrer dans des configurations diverses où le problème fondamental est de se coordonner. Ainsi pour rester dans le domaine militaire, deux parachutistes tombés à des points différents d'une contrée inconnue chercheront à se rejoindre sur le pont le plus proche. Ce pont présente sur leur carte une prégnance particulière et dicte la direction à suivre.

Les situations de rendez-vous constituent l'archétype des problèmes de coordination analysés par Schelling. Quand deux individus perdus cherchent à se retrouver sans pouvoir communiquer, ils choisissent l'endroit qui va focaliser l'attention. Ce repère de coordination est qualifié de point focal. Il met fin aux spéculations sur l'intention de l'autre à condition d'être persuadé que l'autre opère le même choix. Mais « *the solutions are, of course, arbitrary to this extent any solution is "correct" if enough people think so* » (Schelling, 1960, p. 55). Un autre endroit sur la carte aurait pu réunir les parachutistes et plus généralement nos acteurs auraient pu choisir un autre lieu¹⁵. Malgré son caractère arbitraire, le point focal coordonne les anticipations et fournit la clef de la coordination réussie.

15 L'exemple le plus célèbre, mais que Schelling (1960) présente très rapidement (la solution faisant l'objet d'une note au bas de la page 55), est celui de deux individus perdus dans New York qui se retrouvent à la gare centrale.

2.2 La coordination par convention

Le cas de la coupure téléphonique¹⁶ permet d'explicitier et de développer la nature arbitraire de la solution. La communication entre Franck et Guillemette est interrompue. Qui doit rappeler le premier ? Le jeu suivant décrit les 4 cas possibles. Deux situations assurent la coordination (cases B et C avec un gain de « 1 » pour chacun), les deux autres sont des configurations où il y a échec de coordination¹⁷ (A et D, avec un gain de « 0 » pour chacun).

16 Ce jeu tombé en désuétude retrouve une certaine actualité avec l'arrivée des téléphones portables.

17 Nous définissons l'« échec de coordination » comme une situation hors équilibre (s).

Tableau 1 : La coupure téléphonique : un jeu de pure coordination

		<i>Guillemette</i>	
		Rappelle	Ne rappelle pas
<i>Franck</i>	Rappelle	(0,0) (A)	(1,1) (B)
	Ne rappelle pas	(1,1) (C)	(0,0) (D)

Franck et Guillemette cherchent à reprendre leur conversation. Leur intérêt identique les situe dans un jeu de coordination. Les solutions sont parfaitement symétriques avec les mêmes résultats pour chacun : 1 si la coordination est réussie, 0 si elle échoue. On dira qu'il s'agit d'un jeu de pure coordination.

Ce jeu admet deux solutions où l'un rappelle et l'autre pas (B et C) que Lewis qualifie « d'équilibre propre de coordination ». Ces équilibres sont des équilibres de Nash du jeu de coordination. Ils présentent – en supplément – la particularité d'être préférés par les agents dans la mesure où, ici, il est avantageux pour l'un de ne pas rappeler quand l'autre rappelle et inversement. Nous les qualifierons d'équilibre de « Nash-propre »¹⁸.

18. Lewis n'emploie jamais l'appellation « équilibre de Nash ». Son concept d'équilibre « propre » est plus exigeant car il ajoute aux conditions de Nash la nécessité que le choix d'un joueur, étant donné le

choix de l'autre, soit le meilleur pour le premier. Un équilibre propre apparaît en effet « *if each agent likes it better than any other combination he could have reached, given the other's choices* » (Lewis, 1969, p. 22). Pour l'identifier sur une matrice de paiements 2x2, on cherche pour chaque cellule (ligne, colonne) s'il existe sur la même ligne un paiement supérieur pour le « joueur colonne » ou sur la même colonne un paiement supérieur pour le « joueur ligne ». En cas d'égalité, on obtient un équilibre de Nash au sens large. Comme la notion d'équilibre propre de Lewis est liée à celle d'équilibre de Nash, on peut la qualifier d'équilibre « Nash-propre ». On peut se reporter à Klein (1997) ou Coleman (1989) pour une formulation plus précise des différents concepts d'équilibre utilisés par Lewis.

Mais on ne peut pas départager ces deux solutions sans recourir à un repère extérieur au jeu, provenant de l'histoire personnelle des individus ou des règles partagées. Ainsi, la règle « celui qui a appelé le premier, rappelle » offre un moyen de résoudre le problème de coordination. Cette règle, que Lewis qualifie de *Convention*, fournit la clef de la coordination réussie mais reste arbitraire. « *Any convention is arbitrary because there is an alternative regularity that could be our convention instead* » (Lewis, 1969, *Convention*, p. 70).

La convention de Lewis présente alors de fortes analogies avec le point focal de Schelling : les solutions sont des équilibres de Nash en stratégie pure, qui résolvent un problème de coordination. La sélection d'une solution (un équilibre de « Nash-propre ») est arbitraire car une autre solution aurait pu convenir.

Cependant, l'approche de Lewis va plus loin que celle de Schelling car il analyse les situations récurrentes. Le point focal, au sens de Schelling, élabore une « théorie de la première fois ». Il constitue pour Lewis une « saillance », utile à un problème de coordination inédit. Il ne peut devenir une convention que si son usage est répété. Lewis distingue donc bien point focal (ou saillance) et convention.

Cette extension ne va pas sans bouleverser les outils nécessaires à l'analyse. En effet, dans la théorie de Schelling, se coordonner nécessite uniquement de faire travailler son imagination. Il le déclare clairement en affirmant (p. 57) : « *Finding the key, or rather finding a key – any key that is mutually recognized as the key becomes the key – may depend on imagination than on logic* » ; et dans ce cas « *Poets may do better than logicians* » (p. 58)¹⁹.

¹⁹ Schelling réitère souvent ces propos dans son livre, par exemple pour choisir une carte dans un

jeu : « *how to use it may depend more on imagination than on logic, more on poetry or humor than on mathematics* » (p.97). Cet aspect de l'approche de Schelling est également souligné par Boyer et Orléan (1994, p. 220).

Lewis ne se satisfait pas du recours à l'imagination qui semble *ad hoc*. Se coordonner sur un point focal nécessite – sauf à supposer une large part d'inné – l'acquisition d'une gymnastique d'esprit forgée au cours du temps. Il faut donc mobiliser la notion d'acoutumance pour justifier la coordination par convention. Dans ces conditions, le problème se déplace. Il ne s'agit pas d'étudier comment une convention peut s'établir mais comment elle se perpétue : « *a convention is so called because of the way it persists, not because of the way it originated* » (Lewis, 1983, p. 81). Il a besoin, dans cette entreprise, d'un concept qui dispense chacun de faire travailler son imagination tout en ayant l'assurance de ce que va faire l'autre : le savoir commun de la rationalité. Le savoir commun opère ainsi un transfert du pouvoir de coordination de l'imagination à la rationalité. Il constitue une condition essentielle pour la stabilité d'un équilibre de coordination. Le choix privé des acteurs de suivre R doit en effet être rendu public pour que chacun infère la rationalité de se conformer à R. Le savoir commun conduit à une spéculativité²⁰ itérative, comme dans un jeu de miroir qui empile des représentations croisées.

²⁰ Le terme est utilisé par Dupuy (1989, p. 366) qui le définit comme l'acte mental par lequel un esprit humain se met à la place d'un autre.

Grâce à la rationalité et à sa connaissance commune, les individus savent tacitement ce que va faire l'autre. Le savoir commun de la rationalité agit comme un opérateur de « taciticité ». Pour se coordonner, les individus ont juste besoin de savoir ce qu'ils ont fait dans le passé (le précédent) même s'ils ont oublié la genèse de ce dispositif de coordination (cf. tableau 2). Comme le point focal, la convention permet un accord sans communication préalable. En ce sens, tous deux coordonnent les anticipations (Mc Adams, 2001). Mais, alors que l'imagination d'un point focal fournissait un signal unique pour éviter les anticipations sans fin, la convention repose sur un jeu d'anticipations croisées infinies.

L'originalité de l'approche de Lewis est donc de montrer que les conventions résolvent un problème de coordination entre individus qui restent pleinement rationnels. Grâce au savoir commun de la rationalité, les individus n'ont pas besoin de s'en remettre à la seule imagination des autres.

Tableau 2 : du point focal à la convention

Point focal (Schelling)	Convention (Lewis)
Coordination des anticipations	
Problème de coordination singulier	Problème de coordination récurrent
Coordination par prégnance	Coordination par précédent
Dispositif non oublié	Dispositif oublié
Usage pour une première fois	Usage répétitif (régularité)
Imagination	Connaissance commune de la rationalité

3 La convention fait-elle bon ménage avec la rationalité?

La théorie de Lewis permet de concilier l'arbitraire et le rationnel dans la résolution des problèmes de coordination. A ce titre, sa notion de convention se réfère à la théorie du choix rationnel, appliquée notamment aux jeux de coordination, ce qui la rend pertinente pour les économistes²¹ et pas seulement pour les philosophes.

21. De plus en plus de travaux de théorie des jeux modélisent la notion de convention. Voir Batifoulier et de Larquier (2001) pour une revue de la littérature.

Cependant la soumission de la convention aux contraintes du choix rationnel permet-elle de lui conserver sa cohérence? L'analyse sera menée en deux temps. Dans un premier temps (3.1), on insistera sur les critiques internes et externes qui font de la notion de convention un objet qui ne fait pas toujours bon ménage avec la notion de rationalité. Dans un second temps (3.2), on examinera le cas du défaut de coordination où les individus adoptent une convention qui correspond à un équilibre de « Nash-propre » sous-optimal au sens de Pareto.

3.1 La rationalité de la convention : problèmes logiques

Sous sa forme itérative infinie sur les anticipations croisées des intentions de soi et des autres, le **CK** se heurte à des limites importantes. L'attribution du **CK** aux agents par le modélisateur revient à les doter de compétences cognitives irréalistes. Elle annule l'asymétrie d'information et conduit les individus à avoir une opinion identique (Geanakoplos, 1992). Ceci découle du théorème de l'accord d'Aumann (1976)

selon lequel on ne peut trouver un accord sur un désaccord. Si deux individus ont un *a priori* commun et que leurs croyances sur un événement sont **CK**, alors leurs croyances doivent être identiques. Le problème de coordination est ainsi contourné en rendant les individus semblables par la pensée.

Dupuy (1989, 1994) a également souligné les difficultés du **CK** en dégageant les paradoxes des inférences liées au raisonnement rétrograde ("*backward induction*") dans les jeux séquentiels du type du mille-pattes. En remontant le jeu à partir de la dernière étape par éliminations successives pour trouver l'équilibre, la prédiction rationnelle est une défection dès le premier tour. Elle laisse le joueur désarmé si son partenaire a laissé passer son tour. Cet écart de conduite pose un problème d'interprétation sur l'intention de l'autre : est-il irrationnel ou fait-il semblant de l'être pour accroître ses gains ?

Dès qu'un doute s'introduit sur le comportement des autres joueurs, la compréhension des intentions individuelles nécessite de s'écarter du **CK**. Cette attitude est mieux adaptée aux inférences que développent les joueurs dans les expériences car ils laissent plusieurs étapes du jeu se dérouler (Kreps et alii, 1982 ; Mac Kalfey et Palfrey, 1992 ; Aumann, 1994).

Cependant Lewis prend soin de se démarquer de cette version du **CK** de la rationalité sous une forme itérative infinie. Il estime que les capacités cognitives limitées des individus réduisent de fait les imputations sur les croyances mutuelles à quelques ordres, dont le nombre ne peut guère excéder quatre dans la vie quotidienne. "*Of course I do not imagine that anyone will solve a coordination problem by first acquiring a seventeenth order expectation from somewhere and then sitting down to do his replications. For one thing we rarely do have expectations of higher order than, say, fourth.*" (1969, *Convention*, p. 32). Une raison de croire que l'autre a une certaine intention se transforme en croyance effective si chacun a un « *degré suffisant de rationalité* » (p. 55). Lewis estime donc qu'un savoir mutuel de quelques ordres suffit à assurer la stabilité d'une convention. Selon la typologie de Barwise (1988), Lewis choisit une version intermédiaire entre le savoir commun itératif infini et le savoir commun à point fixe ou d'ordre 1 avec un **CK** à situation partagée.

Mais quelles sont les conditions pour que le savoir commun se forme avec un jeu de croyances mutuelles limité à quelques ordres ? Lewis (1969, *Convention*, p. 57) envisage trois possibilités : l'accord, irrecevable car hors de son champ des conventions tacites-, la saillance, -inacceptable car jugée fragile-, et le précédent qu'il retient comme solution. La conformité

passée à une convention constitue une base solide pour que sa reconduction future soit un **CK**.

Ce point central dans le raisonnement de Lewis est critiqué par Miller (1986, 1987, 1990) et Gilbert (1990, 1992). Le savoir commun de l'action passée ne peut constituer une base rationnelle aux attentes sur le comportement d'autrui. Un agent rationnel n'est pas lié par sa décision passée de suivre une convention mais par un raisonnement actuel. Or la conformité passée ne peut me donner une raison de croire que l'autre va se conformer car cette croyance est suspendue à celle de ma propre conformité actuelle et non passée. Par exemple dans le jeu de la coupure téléphonique, le fait que Guillemette ait hier adopté la règle « en cas de coupure, je rappelle si j'ai appelé la première » ne garantit pas à Franck que cette règle va prévaloir aujourd'hui comme demain. Il n'est pas sûr qu'un comportement passé deviendra un précédent qui s'imposera avec force au comportement présent pour des agents rationnels (voir Urrutiaguer et alii, 2001 pour un développement). La reproduction rationnelle de la conformité passée suppose l'existence d'une règle d'apprentissage, fondée par exemple sur une révision bayésienne des croyances, si on veut rester dans la théorie du choix rationnel, que Lewis n'a pas conceptualisée.

3.2 Rationalité de la convention et défaut de coordination

Accoler rationalité et convention peut donc conduire à certains problèmes logiques. C'est le cas aussi quand des individus suivent des conventions qui leur paraissent désagréables ou coûteuses à suivre (mettre une cravate ou échanger des cadeaux en fin d'année par exemple). Nous qualifierons ce type de « mauvaises conventions ». Leur existence découle du caractère arbitraire de la convention. Un autre choix aurait pu procurer des gains supérieurs aux acteurs. Cette caractéristique intrinsèque à la convention heurte les canons de la rationalité. Pourquoi des individus adoptent des conventions dont ils savent qu'elles affaiblissent leur intérêt personnel ?

Lewis ne s'intéresse pas au problème. Pour lui une convention est forcément « bonne » pour trois raisons. D'abord, elle est satisfaisante car elle permet la coordination. Lewis écarte ainsi le problème de la normativité des conventions²². Ensuite, la notion d'équilibre « propre » porte intrinsèquement un marqueur de normativité. Les joueurs ont intérêt dans un jeu de coordination à préférer un équilibre Nash propre à un équilibre de Nash impropre. Enfin, il va de soi que quand un équilibre de Nash propre de coordination amène pour les deux joueurs des gains supérieurs à ceux d'un autre équilibre, il doit être choisi et

c'est le **CK** de la rationalité qui doit conduire à ce choix. Le problème est ici en fait, pour Lewis, un problème trivial. Seul est intéressant le cas où deux équilibres de Nash propres ne peuvent être départagés comme dans le jeu de la coupure téléphonique²³.

22 Il reconnaît pourtant que les "*conventions may be a species of norms*" (1969, *Convention*, p. 97) et "*failure to conform tends to evoke unfavourable responses from others*" (p. 99). Mais sa reconnaissance d'une composante normative de la convention se limite à ces phrases isolées. En effet, supposer que l'approbation sociale puisse guider la conformité à la convention affaiblit considérablement sa composante arbitraire. Une convention pas ou peu arbitraire reste-t-elle toujours une convention ? Ce point a été particulièrement souligné et discuté par les philosophes du droit que sont Marmor (1996) et Mac Adams (2001). Nous discutons ici le traitement de cette composante normative de la convention par l'économie.

23. Un problème de coordination doit contenir au moins deux équilibres propres de coordination et « *the two proper equilibria are sufficient to keep the problem non trivial* » (Lewis, 1969, *Convention*, p. 2).

Un exemple peut utilement éclairer le lecteur. Considérons la matrice de gain du tableau 3 (Lewis, 1969, *Convention*, p. 22). Ce jeu admet 3 équilibres de Nash, 2 sont « Nash-propres » (A et B), C est « Nash-impropre ». Il est évident que A et B doivent être préférés à C. Mais un problème non trivial de sélection se pose si on élimine la dernière ligne et la dernière colonne.

Tableau 3 : équilibres de Nash propre et impropre

(2,2) A	(0,0)	(0,0)
(0,0)	(2,2) B	(0,0)
(0,0)	(1,1)	(1,1) C

Si l'on sort de ces jeux de pure coordination pour lever l'hypothèse d'équivalence des alternatives, alors la théorie de Lewis se heurte à un certain nombre de limites. Un jeu de coordination 2x2 symétrique admet deux équilibres de Nash propres, qui peuvent être ordonnés selon le critère de Pareto. On dira alors qu'il y a un défaut de coordination car les joueurs peuvent adopter la solution la plus défavorable, en d'autres termes, suivre une mauvaise convention.

Pour Lewis, ce défaut de coordination n'existe pas car les joueurs doivent se coordonner sur l'équilibre Pareto optimal²⁴. Tout équilibre de Nash propre qui est Pareto optimal règle le problème de la coordination pour Lewis. En croisant les deux critères, on peut obtenir un équilibre unique qui fournit une définition de la bonne convention dans la mesure où on ne voit pas pourquoi il ne serait pas adopté par des individus qui ont des intérêts communs.

24. Même s'il n'emploie jamais cette expression.

Toutefois, le critère de Pareto (qui constitue la normativité de base de l'économiste) semble insuffisant pour départager des conventions car rien n'interdit aux joueurs d'adopter un équilibre Pareto dominé. Quel que soit le type de jeu de coordination envisagé, il semble impossible de prédire, à coup sûr, que la convention va correspondre à la situation optimale :

1 Dans un jeu de pure coordination, comme dans le cas de la coupure téléphonique, les 2 équilibres de Nash propres sont identiques et tous deux Pareto-optimaux. La Pareto optimalité n'est ici d'aucun secours²⁵.

2 Dans un jeu du rendez-vous, un seul des deux équilibres de Nash propres est Pareto-optimal mais rien ne garantit son adoption par les joueurs.

3 Dans un jeu de chasse au cerf²⁶, les conclusions sont identiques au jeu du rendez-vous à la différence que, cette fois, l'équilibre Pareto dominé présente un nouvel avantage en étant dominant en risque selon la terminologie introduite par Harsanyi et Selten (1988).

25. On peut tenir les mêmes conclusions pour un jeu de « bataille des sexes ».

26. Ce jeu est adapté de la célèbre parabole de J-J. Rousseau que l'on trouve dans son « Discours sur l'origine de l'inégalité parmi les Hommes » et reprise par Lewis (1969, *Convention*, p. 47). Pour se nourrir, deux individus totalement libres peuvent chasser le lièvre ou le cerf. Le cerf est plus nourrissant mais doit être chassé à deux, l'un devant faire le rabatteur. L'avantage du lièvre est qu'il peut être chassé seul tandis qu'un chasseur de cerf isolé n'a aucune chance de se nourrir et meurt. Dans ce jeu, la meilleure situation du point de vue de la Pareto-optimalité (chasser le cerf) n'est pas celle que nous enseigne la prudence la plus élémentaire (attraper « sans scrupule » le lièvre). Aussi, (des individus rationnels peuvent être conduits à choisir la solution que nous pourrions qualifier de plus raisonnable, au risque de négliger l'accroissement de leurs gains. Harsanyi et

Selten montrent qu'il s'agit d'un comportement individuellement rationnel face au risque.

La *Convention*, parce qu'elle est arbitraire, peut toujours être sous-optimale. En d'autres termes, il est toujours concevable que des individus, pourtant rationnels et cherchant en toutes circonstances à maximiser leurs gains, adoptent une solution qui ne corresponde pas à cette maximisation. La rationalité des conventions est alors prise en défaut.

Le concept de convention est une solution générale à un problème de coordination mais l'outillage de Lewis n'est pas conçu pour appréhender tous les jeux de coordination. Pour expliquer la rationalité du suivi d'une *Convention*, il faut s'écarter du critère de rationalité mobilisé par Lewis. C'est ce que font, par exemple, Harsanyi et Selten²⁷, dont les travaux sur le concept de domination en risque, permettent de justifier la rationalité du choix d'une convention sous-optimale. Le choix entre la convention Pareto optimale et celle sous-optimale mais dominante en risque dépend du type de rationalité des joueurs, collective dans le premier cas, individuelle dans l'autre. « *Risk dominance, is based on individual rationality : it is an extension of bayesian rationality [...] If equilibrium point E1 risk-dominates equilibrium point E2, this means that in a situation where the players are uncertain whether E1 or E2 will be the actual outcome, any player who tries to maximize his expected payoff in terms of rationally chosen subjective probabilities over the other players' strategies will opt for E1. In contrast, payoff dominance is based on collective rationality : it is based on the assumption that in the absence of special reasons to the contrary, rational players will choose an equilibrium point yielding all of them higher payoffs, rather than one yielding them lower payoffs. That is to say, it is based on the assumption that rational individuals will cooperate in pursuing their common interest if the conditions permit them to do so.* » (1988, p 356)

27. Dans leur ouvrage, Harsanyi et Selten n'utilisent jamais le terme de convention. Ils s'intéressent à des jeux qui admettent plusieurs équilibres et où un problème de sélection d'un équilibre se pose.

Lewis ne prend en considération que cette dimension collective de la rationalité qui amène à proscrire tout défaut de coordination. Plus généralement, sa théorie, malgré ses apports indéniables, est entièrement construite pour éviter le problème des « mauvaises » conventions. Or, c'est en explorant ce type de problèmes qu'on peut mettre l'accent sur les « bonnes » raisons qui poussent les individus à adopter une convention.

Références bibliographiques

- ARROW K. (1974), "Limited knowledge and economic analysis", *American Economic Review*, vol. LVIV, n° 1, p 1-10.
- ARROW K. (1987), "De la rationalité des individus et des autres dans un système économique", *Revue française d'économie*, vol. II, n° 1, p 22-47.
- AUMANN R. (1976), "Agreeing to disagree", *Annals of statistics*, n° 4, p. 1236-1239
- AUMANN R. (1994), "L'irrationalité dans la théorie des jeux" dans Orléan A. (dir.), *Analyse économique des conventions*, p. 43-60.
- AVRAMIDES A. (1997), "Intention and Convention" dans Hale B. et Wright C. (dir.), *A Companion to the Philosophy of Language*, Oxford, Blackwell.
- BARWISE J. (1988), "Three Views of Common Knowledge", dans Vardi (dir.), *Theoretical Aspects of Reasoning about Knowledge*, 2d Conference, Morgan Kaufmann, p. 365-378.
- BATIFOULIER P. et LARQUIER G. de (2001), "L'approche stratégique des conventions", *Communication au séminaire "Approches conventionalistes du travail"*, mai.
- BOYER R. et ORLÉAN A. (1994), "Persistance et changement des conventions. Deux modèles simples et quelques illustrations" dans Orléan A. (dir.) *Analyse économique des conventions*, p. 219-247.
- COLEMAN K. (1989), "Lewis's notion of a convention", *Auslegung*, vol. 15, n° 2, p. 173-184.
- DAVIDSON D. (1982), "Communication et convention", rééd. dans Davidson D. (1993), *Enquêtes sur la vérité et l'interprétation*, éditions Jacqueline Chambon, p. 377-398.
- DUPUY J-P. (1989), "Common knowledge et convention", *Revue économique*, vol. 40, n° 2, mars, p. 361-400
- DUPUY J-P. (1994), "Temps et rationalité", *Cahiers d'économie politique*, n° 24-25, p. 69-104.
- GEANAKOPOLOS J. (1992), "Common knowledge", *Journal of Economic Perspectives*, vol. 6, n° 4, p. 33-82.
- GILBERT M. (1990), "Rationality, Coordination and Convention", *Synthese*, vol. 84, p. 1-21.
- GILBERT M. (1992), *On social facts*, Princeton University Press, 1ère édition 1989.
- HARSANYI J.C., SELTEN R. (1988), *A General Theory of Equilibrium Selection in Games*, Cambridge, MIT Press.
- KLEIN D. (1997) "Convention, social order and the two coordinations", *Constitutional Political Economy*, n° 8, p 91-106.
- KREPS D., MILGROM P., ROBERTS J., WILSON R. (1982), "Rational cooperation in the finitely repeated prisoners dilemma", *Journal of Economic Theory*, vol. 27, p. 253-279.
- LAURENCE S. (1996), "A chomskian alternative to conventions-based semantics", *Mind*, vol. 105, p. 269-301.
- LEWIS D. (1968), "Counterpart theory and quantified modal logic", *Journal of philosophy*, 65, p 113-126, rééd. Dans *Philosophical Papers*, vol. 1, 1983, Oxford U. P.
- LEWIS D. (1969), *Convention. (A Philosophical Study)*, Harvard University Press, Cambridge.
- LEWIS D. (1972), "General Semantics", rééd. dans *Philosophical Papers*, vol. 1, 1983, Oxford U. P.
- LEWIS D. (1973), *Counterfactuals*, Oxford, Blackwell
- LEWIS D. (1983), "Languages and Language", *Philosophical Papers*, Oxford University Press, p. 163-188.
- LEWIS D. (1993), "Langages et langage", *Réseaux*, n° 62, p. 11-18.
- Mc ADAMS R. (2001), "Conventions and norms (Philosophical aspects)", dans Smelser N., Bates P., *International Encyclopedia of the Social and Behavioral Sciences*.
- Mc DERMOTT M. (2001), "Quine's holism and functionalist holism", *Mind*, vol. 110, p. 977-1025.
- Mc KELVEY R. D., PALFREY T. R. (1992), "An Experimental Study of the Centipede Game", *Econometrica*, vol. 60, n° 4, July, p. 803-36.
- MARMOR A. (1996), "On Convention", *Synthese*, vol. 107, n° 3, p. 349-371
- MILLER S. (1986), "Conventions, Interdependence of Action, and Collective Ends", *Noûs*, vol. 20, n° 2, p. 117-142
- MILLER S. (1987), "Conventions, Expectations and Rationality", *The Southern Journal of Philosophy*, vol. 25, n° 3, p.357-380
- MILLER S. (1990), "Rationalising Conventions", *Synthese*, vol. 84, p. 23-41
- MILLIKAN R. G. (1998), "Language conventions made simple", *The Journal of Philosophy*, vol. 94, p. 161-180.
- QUINE W. V.O. (1936), "Truth by Convention", réédité in *The Ways of Paradox and Other Essays*, Harvard University Press, 1977.
- QUINE W. V.O. (1951), "Two Dogmas of Empiricism", *The Philosophical Review*, January, vol. 60, n° 1, p. 20-43.
- QUINE W. V.O. (1960), *Word and object*, Cambridge, MIT Press.
- ROUSSEAU J-J. (1754), *Discours sur l'origine de l'inégalité parmi les hommes*, édition de 1985, Bordas.
- SHELLING T. (1960), *The strategic of conflict*, Harvard University Press.
- STROLL A. (2000), *Twentieth-century analytic philosophy*, Columbia UP.
- URRUTIAGUER D., BATIFOULIER P et MERCHIERS J. (2001), "Peut-on se coordonner sur une base arbitraire ? Lewis et la rationalité des conventions", in *Théorie des conventions*, P. Batifoulier éd., Economica, p. 63-95.
- WINCH C. (1988), "Do we need conventions ? ", *Philosophical Investigations*, vol. 11, p. 133-146