

LA THÉORIE ÉCONOMIQUE DE LA CONFIANCE

ET SES LIMITES

André Orléan

(Paru dans les *Cahiers de Socio-Économie*, intitulés « La confiance en question », sous la direction de R. Laufer et M. Orillard, coll. « Logiques Sociales », Paris, L'Harmattan, 2000, 59-77)

Ce n'est pas un hasard si O. Williamson commence l'article "Calculativeness, Trust, and Economic Organization"¹ dans lequel il se propose de clarifier "l'insaisissable notion de confiance", par une réflexion sur ce qu'est l'Economie, sur les relations que cette discipline doit entretenir avec les autres Sciences Sociales. C'est bien là une des questions qu'on trouve au cœur du débat suscité aujourd'hui par la théorie économique de la confiance. La réponse que lui apporte O. Williamson a le mérite de la clarté. Elle procède en deux temps. Dans un premier temps, il caractérise l'approche économique par l'usage que cette discipline fait de la notion de *calculativeness* : "Calculativeness is the general condition that I associate with the economic approach and with the progressive extension of economics into the related social sciences" (*op. cité*, page 456). Puis, dans un second temps, il montre que tout ce que les économistes disent sur la confiance s'inscrit pleinement et sans difficulté dans ce cadre. Pour O. Williamson, les situations pour lesquelles les économistes ont eu recours à la notion de confiance, ne sont que des cas particuliers de transactions risquées : les outils classiques que la théorie économique a contruits pour les analyser, s'y appliquent aisément. Aussi, pour cet auteur, l'emploi du terme "confiance" doit-il être prohibé dans la mesure où il est source de confusion : "Calculative trust is a contradiction in terms" (page 463). C'est ainsi qu'il écrit : "I submit that calculativeness is determinative throughout and that invoking trust merely muddies the (clear) waters of calculativeness" (page 471). Ou encore "If calculative relations are best described in calculative terms, then diffuse terms, of

¹ *Journal of Law and Economics*, vol. XXXVI, avril 1993, pp. 453-486.

which trust is one, that have mixed meanings should be avoided when possible" (page 469) ; "Reference to trust adds nothing" (page 474).

Mais, par delà le rejet du terme de confiance, c'est l'idée même d'une possible collaboration entre économie et autres sciences sociales qui se trouve de facto rejetée. Tel est, me semble-t-il, un des enjeux centraux du débat contemporain autour de la notion de confiance : celui de la nature des relations théoriques qui doivent être tissées entre l'économie et les autres Sciences Sociales. Il est naturel que ce problème soit posé à propos d'une notion comme celle de confiance dont l'ambiguïté vient précisément du fait qu'elle se situe à la frontière de plusieurs disciplines, en l'occurrence l'économie et la sociologie, comme le rappellent les réflexions socio-économiques de M. Granovetter². C'est dans cette perspective que nous aborderons les réflexions de D. Kreps. Les travaux de D. Kreps illustrent une stratégie intellectuelle qui vise à penser la production de la confiance en des termes strictement économiques. De ce point de vue, la critique de O. Williamson est parfaitement justifiée : la notion de réputation que propose D. Kreps, devrait permettre de résoudre la question de la confiance sur la seule base du calcul des intérêts privés, sans qu'il soit fait appel à des formes sociales extérieures. Cependant, notre thèse sera que, malgré l'originalité et la force de sa construction, D. Kreps n'a pas parfaitement atteint son but. Nous montrerons que le mécanisme réputationnel repose implicitement sur la mobilisation de ressources non économiques ; à savoir l'appartenance à certains réseaux sociaux.

Avant de développer cette analyse, il convient de rappeler que l'indépendance revendiquée par le discours économique participe de ce vaste mouvement historique qui voit l'ordre économique, jadis province particulière de l'activité collective, étroitement imbriquée dans le tissu social, "acquérir son indépendance, se séparer de la société et constituer à lui seul un système distinct auquel le reste du social doit être soumis"³. Or, comme le montre K. Polanyi, la prétention des valeurs économiques à dominer la totalité sociale conduit à un dangereux émiettement de l'expérience sociale⁴ dans la mesure où la logique de la marchandise s'avère incapable de gérer à elle seule la complexité des liens sociaux fondamentaux. Les débats contemporains expriment ce double mouvement : d'une part, l'autonomie progressive que la valeur économique tend à acquérir par rapport à l'ensemble des valeurs globales ; d'autre part, l'incapacité dans laquelle elle se trouve de parvenir à ses fins, son incapacité à

² M. Granovetter, "Economic Action and Social Structure: The Problem of Embeddedness", *American Journal of Sociology*, vol. 91, n°3, novembre 1985, pp. 481-510.

³ L. Dumont, préface à *La grande transformation* de K. Polanyi, Editions Gallimard, 1983.

⁴ Dont le contre-modèle se trouve dans le don comme "phénomène social total" où s'enchevêtrent dimensions économiques, religieux et juridiques (Marcel Mauss, *Sociologie et anthropologie*, PUF).

supporter la totalité de l'édifice social.

A l'occasion du travail de D. Kreps, nous chercherons à montrer comment l'économie cherche à rendre l'ordre économique indépendant des autres dimensions sociales, à en faire un domaine séparé et partiellement autonome d'activités ; et comment elle échoue. Plus précisément nous montrerons que la capacité d'engagement d'un individu, ce qui n'est rien moins que sa capacité à promouvoir une action collective en convoquant la confiance d'autrui, ne trouve pas dans l'ordre économique des ressources suffisantes. Pour s'engager de manière crédible, l'individu doit pouvoir mobiliser une valeur plus solide, moins précaire, que celle qui caractérise les relations marchandes. On se rend crédible auprès d'autrui par une représentation de soi plus complète que sa seule appartenance économique, ce qu'on appellera l'être social. Cette fragilité de la valeur économique est mis au jour par A. Caillé avec force et talent quand il écrit que derrière le désir d'argent "se dessine l'espoir, toujours dénié mais toujours renouvelé, d'opérer la conversion de la valeur économique précaire en une valeur sociale indubitable et substantielle, le fantasme de transmuier la monnaie moderne en une véritable monnaie sauvage"⁵. L'archaïsme, c'est alors la possibilité d'une évaluation solide, définitive, *hic et nunc*, de l'être ; c'est à dire une évaluation qui se soit rendue indépendante des fantaisies et des aléas qui caractérisent l'univers de la marchandise.

1. Présentation du problème⁶

Considérons la situation formalisée par le schéma qui suit et l'analyse qu'en a proposée D. Kreps. Rappelons que le couple de nombres (x,y) signifie que, dans la situation considérée, l'utilité de l'individu A vaut x et celle de B vaut y. Ces nombres n'ont qu'une valeur ordinale.

		Joueur B	
		Tricher	Honorer
Joueur A	Faire confiance	-5 , 20	5 , 5
	Ne pas faire confiance	0 , 0	0 , 0

⁵ A. Caillé, "Monnaie des sauvages et monnaies des modernes", *Bulletin du MAUSS*, mars 1987, pp. 143-150.

⁶ Cette présentation reprend la présentation proposée dans les pages 23 et 24 de « Sur le rôle respectif de la confiance et de l'intérêt dans la constitution de l'ordre marchand », *Revue du MAUSS*, n°4, semestre 1994, pp. 17-36.

L'interaction ainsi formalisée se comprend de la manière suivante. Le joueur noté A joue en premier. Il a le choix entre « Ne pas faire confiance », ce qui signifie qu'il se refuse à entrer en relation avec le joueur noté B ; ou bien « Faire confiance » et entrer en relation avec lui. Si A et B n'interagissent pas, leur situation est caractérisée par les deux nombres (0 , 0). C'est la situation de référence. Elle correspond à "l'état de nature", état qui prévaut avant que les hommes inventent la coopération.

A peut décider d'entrer en relation avec B. Cette décision, dans l'interaction particulière que décrit ce schéma, signifie que A donne unilatéralement à B « quelque chose » et fait confiance à B pour que celui-ci en fasse bon usage. C'est alors au tour de B de jouer. Celui-ci peut « Honorer » la confiance que A a placée en lui. Cela conduit à la situation décrite par (5 , 5). Dans cette situation, la satisfaction des deux agents est nettement supérieure à celle qu'ils connaissaient dans l'état de nature. Autrement dit, la coopération entre A et B crée un état social plus satisfaisant pour tous. Malheureusement B peut choisir de « Tricher » et accaparer à son seul profit ce que A lui a apporté unilatéralement pour que la relation se noue. On obtient alors l'état caractérisé par les utilités (-5 , 20). A a une utilité moindre que celle qu'il obtient dans l'état de nature puisqu'il a perdu ce qu'il avait investi dans la relation sans en retirer quoi que ce soit. B qui a spolié A sans fournir aucun effort en contrepartie, obtient une satisfaction de 20 qui est supérieure à celle que lui aurait procurée une coopération équitable.

De nombreuses interactions répondent à cette logique. C'est le cas par exemple de l'échange séquentiel de deux marchandises. A et B décident d'échanger deux biens : la chaise de A contre la montre de B. Cet échange permettrait à chacun d'améliorer sa situation, car ce qu'il reçoit est, pour A comme pour B, plus important que ce qu'il donne. La réalisation de l'échange ferait passer de l'état (0 , 0) à (5 , 5). Cependant, l'échange étant séquentiel, il est nécessaire que, dans un premier temps, A donne à B sa chaise, pour que, dans un second temps, B livre à A sa montre. Aussi, quand B a reçu la chaise, s'ouvre la possibilité pour lui de s'en emparer sans livrer sa montre. On obtient alors la situation (-5 , 20). A a perdu sa chaise et B obtient, pour rien, celle-ci. C'est bien le problème de la confiance qui est ici posé. Pour que A entre en relation avec B, A doit faire confiance à B ; il doit supposer que B ne tirera pas avantage de la situation. Cette confiance est source d'utilité collective puisqu'elle permet que se réalise l'état (5 , 5), nettement préférable à l'état de nature (0 , 0). Cependant cette confiance peut être trahie. La trahison met A dans une situation pire que celle qu'il connaissait dans l'état de nature. Cependant, si la confiance fait défaut, aucune relation n'est nouée et la société demeure dans l'état de nature (0 , 0) sans bénéficier des effets positifs de la coopération. C'est ici le problème général de la confiance nécessaire à la formation de liens sociaux qui est très schématiquement exposé. D. Kreps nomme d'ailleurs ce jeu, « le jeu de la confiance ». A la racine du

phénomène monétaire, se trouve cette même question de la confiance, mais dans un contexte plus complexe : A peut décider de livrer à B sa chaise contre une certaine quantité de monnaie. Mais il n'agira de cette manière que pour autant qu'il anticipe que, dans le futur, grâce au signe monétaire qu'il vient d'acquérir, il lui sera possible d'obtenir la montre de B. Ce n'est plus B qui fait l'objet de la confiance de A mais le signe monétaire dans sa capacité présumée à conserver sa valeur.

Que se passe-t-il lorsque des individus se trouvent dans une telle relation ? La théorie économique analyse cette situation en supposant que les agents A et B sont parfaitement rationnels, ce qui signifie en particulier qu'ils agissent toujours de manière à maximiser leur utilité. Très clairement, la coopération (5, 5) est meilleure pour tous que la situation (0, 0). Cependant A, analysant la situation, voit bien que s'il fait confiance à B et si B est rationnel, celui-ci va nécessairement choisir de tricher puisqu'ainsi il obtiendrait une utilité de 20 supérieure à celle de 5 qu'il aurait obtenue en coopérant avec A. Aussi, si A ne fonde son action que sur la stricte rationalité, il ne fera pas confiance à B et n'entrera pas en relation avec lui, car il sait que la rationalité conduit B à trahir. Ce résultat est troublant dans la mesure où il indique que la rationalité ne peut fonder la coopération entre A et B.

Pour éviter ce résultat insatisfaisant, B pourrait être amené à faire des promesses à A : il promet à A que, si A lui fait confiance, il honorera cette confiance. La théorie des jeux montre alors simplement, en répétant le raisonnement précédent, que cette promesse n'est pas crédible. B aura beau promettre, A sait qu'une fois qu'il aura passé la main à B, ce dernier aura intérêt à revenir sur sa parole. Ce résultat est intéressant car il montre que, même si la volonté de coopération de B était véritablement sincère, B n'aurait aucun moyen de communiquer à A son authentique désir de coopérer : l'interaction rationnelle bloque les individus dans une situation de suspicion généralisée dont ils ne peuvent sortir, même si tous deux la jugent néfaste. B ne peut réussir à convaincre A de lui faire confiance. Ce blocage correspond à ce que nous avons appelé « l'incomplétude de la logique marchande pure »⁷ ; à savoir une configuration où la stricte « horizontalité » de la relation, au sens où les individus en présence ne partagent rien si ce n'est leur désir de maximiser leur utilité personnelle, conduit à une impasse. Si, dans une première approximation, on identifie la valeur économique au principe collectivement reconnue selon lequel chaque agent agit de façon à maximiser son intérêt personnel, il apparaît que la seule mobilisation de cette valeur ne suffit pas à engendrer la confiance et à sortir de l'état de nature. Examinons maintenant les différentes solutions imaginées par les économistes pour que, malgré tout, la coopération soit rendue possible. On en distinguera trois : le contrat, le serment et la réputation.

⁷ A. Orléan éditeur, Analyse économique des conventions, collection Economie, P.U.F., avril 1994, p. 16.

La solution apparemment la plus évidente pour un économiste est celle du **contrat**. On est en effet dans une situation où les deux protagonistes peuvent passer *ex ante* un contrat puisque chacun préfère *ex ante* la situation (5 , 5) au statu quo (0 , 0). Notons que cette solution repose sur l'existence d'un tiers, à savoir l'appareil judiciaire, qui contrôle l'application du contrat et sanctionne les contrevenants.

Une deuxième solution est le **serment**. B pour prouver sa bonne foi, son authentique volonté de coopérer, jure sur un texte sacré de respecter sa parole. Pour autant que A n'a aucun doute sur les convictions religieuses de B, A fera confiance à B car il sait que le non-respect du serment coûterait à B bien plus que le gain qu'il pourrait en retirer. Le rôle du tiers est ici tenu par la divinité. C'est elle qui sanctionne le contrevenant.

Cependant ces solutions ne peuvent convenir véritablement à un économiste dans la mesure où elles font reposer la production de la confiance sur des logiques extérieures à l'ordre économique, l'appareil judiciaire dans le cas du contrat et le sacré dans celui du serment. Plus précisément, D. Kreps qui ne prend en considération que la solution par le contrat, légitime son rejet de cette solution en faisant appel à un argument de coût. Le recours à la justice est, en effet, coûteux. Si les coûts de justice auxquels A doit faire face sont supérieurs aux gains qu'il peut en attendre, en l'occurrence 5, alors faire appel à la justice n'est pas pour lui une solution. Aussi D. Kreps proposera une solution qu'on pourrait qualifier de purement horizontale dans la mesure où, apparemment, elle ne repose nullement sur l'existence préalable d'une médiation externe. La notion de réputation sera au centre de cette analyse. Examinons la

2. La réputation

L'idée centrale de D. Kreps consiste à considérer que l'interaction précédente opposant A et B ne se déroule pas une seule fois mais se répète au cours du temps. On passe d'un jeu à un coup, à un jeu répété. Dans ces conditions, le jeu se modifie radicalement. La raison en est simple à saisir : dans le jeu à un coup, A est impuissant devant la trahison de B ; dans le jeu répété, A peut réagir et se venger si B n'honore pas sa parole. En effet, si A fait confiance à B et que B trahit la confiance de A, B obtient instantanément un gain important, 20, comparé au 5 que lui aurait procuré l'autre choix. Mais si B fait ce choix, A peut désormais le punir puisque l'interaction se répète. Il peut alors décider de retirer définitivement sa confiance à B et ne plus jamais interagir avec celui-ci, de telle sorte que B aura 0 jusqu'à la fin des interactions. Supposons que l'interaction dure 10 coups, B obtient alors 20 le premier coup plus 9 fois 0 qu'on doit comparer au 10 fois 5, soit 50, que lui aurait procuré une coopération équitable. Notons que si la longueur du jeu était de 3, alors la

sanction encourue serait plus faible que le gain que procure la tricherie. Ce mécanisme qui repose sur la menace de punitions semble résoudre notre problème en ce qu'elle donne une bonne raison à B de ne pas trahir, pour autant que la longueur du jeu soit suffisamment élevée, et ceci sans qu'on ait eu recours à une quelconque extériorité. En effet, si le jeu est répété un nombre de fois suffisant, il est de l'intérêt bien compris de B de ne pas trahir. Ce qui a pour conséquence de donner une bonne raison à A de lui faire confiance.

Pour des raisons techniques simples, à savoir le classique paradoxe de la « *Backward Induction* », on ne peut supposer que la longueur du jeu répété est une donnée, comme c'était le cas dans l'exemple numérique précédent, en l'occurrence 10 coups. On évite cette difficulté en supposant que les protagonistes ne savent jamais quand le jeu va s'arrêter. Pour ce faire, on suppose qu'à chaque instant, il existe une probabilité, notée p , que A et B interagissent une nouvelle fois. Il en découle qu'avec la probabilité $(1-p)$, le jeu s'arrête au coup considéré. L'argument ainsi légèrement modifié repose alors sur la même logique de sanctions que celle présentée précédemment. Si B trahit maintenant, il obtient 20 immédiatement comparé au 5 qu'il aurait obtenu s'il avait honoré sa parole. Ce gain immédiat de 15 doit être comparé à ce que le fait d'avoir triché lui fait perdre dans le futur. Comme il a triché et qu'en conséquence A refuse de lui faire à nouveau confiance, il obtient désormais 0 pour tous les coups à venir. Si, par contre, il n'avait pas triché, il aurait pu obtenir 5 aussi longtemps que le jeu aurait continué. Cette valeur globale qui mesure son manque à gagner, n'est plus simplement égale à 5 multiplié par le nombre de coups à venir, car ce nombre de coups est désormais une variable aléatoire. Pour calculer le manque à gagner, noté M , il suffit de multiplier 5 par l'espérance mathématique de la longueur, qu'on notera EL . Cette grandeur se calcule aisément et on trouve alors (équation 1) :

$$M = 5 EL = 5 p + 5 p^2 + \dots = 5 \frac{p}{1-p}$$

1

Il en découle qu'il n'est pas dans l'intérêt de B de tricher dès lors que son gain immédiat, soit 15, est inférieur au gain à long terme que procure la coopération, soit M . Il n'est pas difficile de montrer que cela est vérifié dès que p est supérieure à 0.75, i.e. EL supérieur à 3. En effet si la longueur probable du jeu, EL , est trop faible, les sanctions qu'encourt B deviennent inférieures au gain immédiat qu'il peut obtenir en trichant. Si, par exemple, B ne doit plus rencontrer A, ce qui correspond à p égal à 0, il est clair qu'il est de l'intérêt de B de tricher et que la menace de sanctions est insuffisante.

On appliquera ce résultat au cas plus intéressant où B n'interagit plus toujours avec le même protagoniste, mais avec une série de protagonistes différents qu'on

nommera les A_i . Dans ce cas, pour que le mécanisme précédent de sanctions fonctionne, il faut qu'une tricherie faite au détriment de A_j déclenche des sanctions de la part des autres A_i . Autrement dit, il faut que chaque A_i connaisse la manière dont B s'est comporté les fois précédentes et que B sache que les A_i ont accès à cette information. C'est ici l'idée de **réputation** qui apparaît.

La réputation est définie de la manière suivante par D. Kreps. Elle prend la valeur « bonne » si B n'a *jamais* triché ; elle prend la valeur « mauvaise », et cela à jamais, dès l'instant où B a triché au moins une fois. Si on note p , la probabilité que B interagisse au coup suivant avec un des A_i , on démontre alors aisément que si B sait qu'à l'avenir les A_i refuseront de lui accorder leur confiance si sa réputation est mauvaise, il est alors de son intérêt de ne pas trahir dès lors que p est supérieur à 0.75. Cette démonstration reprend exactement la démonstration précédente.

Tels sont les résultats obtenus par Kreps. La production de la confiance repose à nouveau sur une médiation, à savoir la réputation de B. Cependant, cette médiation est interne à l'ordre marchand dans la mesure où elle repose uniquement sur l'observation que font les A_i de la conduite passée de B face aux membres de leur groupe, observation qu'il est clairement dans l'intérêt des A_i de faire. Par le jeu du mécanisme réputationnel, la valeur économique, c'est à dire le calcul rationnel des intérêts, permettrait donc de fonder la coopération entre les hommes. Examinons ce point de plus près.

3. La logique du gage

Pour bien comprendre la logique du jeu étudié, il convient de mettre l'accent, non sur les difficultés qu'y rencontre un A perpétuellement menacé par l'opportunisme de B, mais, bien au contraire, sur celles que connaît B. Pour le voir considérons un B à l'honnêteté scrupuleuse et l'incapacité dans laquelle il se trouve de faire valoir cette honnêteté aux yeux de A. En effet, comme on l'a déjà souligné, sa parole n'est pas crédible. Aussi, parce que B n'a aucun moyen de communiquer à A son désir de coopération, il obtient 0 alors même que la coopération était possible et eût été profitable pour tous. C'est là un résultat étonnant et quelque peu scandaleux. Il apparaît alors que la possibilité qui est offerte à B de pouvoir tricher est un « avantage » dont il se serait volontiers passé. D'une part B n'aura jamais l'occasion de l'exercer puisqu'elle conduit A à ne pas lui faire confiance. D'autre part, précisément parce qu'elle pousse A à ne pas lui faire confiance, elle lui interdit également d'atteindre la solution coopérative. En fin de compte B obtient 0, ce qui est pour lui la pire de toutes les solutions. Si B pouvait ex ante modifier la structure de l'interaction en éliminant radicalement la possibilité de recourir à la tricherie, il le ferait sans attendre. Cela lui permettrait d'obtenir 5 au lieu de 0. Telle est le paradoxe

de la situation étudiée : l'action qui confère à B une position dominante dans le jeu, à savoir la possibilité qui lui est offerte d'accaparer à son seul bénéfice la totalité du surplus produit par la coopération, est ce qui fait que B finira par se trouver dans la pire des situations possibles. Faute de pouvoir éliminer la possibilité même de la tricherie (i.e. transformer la structure du jeu), B doit proposer à A, pour sortir de cette impasse, un arrangement dans lequel la tricherie a cessé d'être un choix envisageable : pour rendre crédible son engagement en faveur de la coopération, il lui faut convaincre A qu'il *n'a pas intérêt à tricher*. Cet arrangement qui permet la production de la confiance et l'émergence de la coopération consiste à proposer un gage. Après en avoir exposé la logique, nous vérifierons que les solutions précédemment étudiées, à savoir le contrat, le serment et la réputation, répondent bien à cette logique générale.

Pour rendre crédible son engagement, B propose un gage G qu'il perdra s'il trahit. Il s'ensuit qu'en cas de trahison, il n'obtiendra plus 20, mais $(20-G)$. Dès lors que la valeur de G est suffisamment importante pour rendre $20-G$ inférieur à 5, il devient de l'intérêt de B de ne pas trahir : en cas de trahison, il perd son gage et obtient, en conséquence, une utilité inférieure à celle que lui aurait procurée la coopération. La possibilité qu'a B de donner un gage rend crédible sa parole. Montrons que nos trois solutions répondent à cette même logique et analysons la nature du gage qui y est proposé. On distinguera, dans cette logique, le gage proprement dit du mécanisme qui opère la saisie de G.

Dans le contrat, le gage est constitué par le contrat lui-même que signe B et qu'il remet à A. L'existence de ce contrat signé par B fait qu'en cas de trahison, A peut déclencher l'action de la justice. On peut être plus précis en distinguant deux types de sanctions pénales. Si la peine qu'encourt B est purement financière, le gage G est une somme de monnaie : si B trahit, il n'obtiendra plus 20, mais 20 moins les dommages-intérêts qu'il devra reverser à A. On peut donc dire que la valeur sur laquelle repose le gage est interne à l'ordre économique et met en jeu un classique calcul d'intérêt. Cependant, même dans cette situation, la mise en œuvre de la saisie passe par la mobilisation d'un mécanisme extérieur aux rapports marchands, l'appareil judiciaire. Le fait que ce mécanisme n'appartienne pas à l'espace marchand est tout à fait central dans la production de la confiance. Pour le voir, il suffit d'imaginer a contrario une situation dans laquelle l'action des magistrats n'est plus dictée par la seule recherche du juste mais est également soumise à la « loi » de la maximisation des intérêts individuels. Ce fait a pour conséquence radicale de transformer le rapport aux juges en un rapport stratégique de même structure que celui qui lie A et B. Pour anticiper le comportement des juges, A est alors amené à procéder au même type de calcul que ceux qu'il a utilisés pour analyser le comportement de B. Il se demande quel est le prix des juges de façon à déterminer si le surplus engendré par la trahison de B est suffisant pour s'assurer de la clémence

des juges tout en permettant à B d'obtenir un revenu supérieur à celui que lui aurait procuré la coopération, soit 5. Transformer les juges en homo oeconomicus, c'est donc brutalement replonger l'interaction dans un univers dominé par le stratégique, l'univers de la marchandise, et, en conséquence, se retrouver face aux mêmes difficultés que celles caractérisant la situation initiale. Le fait d'avoir accès à une logique extérieure à l'ordre économique apparaît donc ici comme étant tout à fait central pour rendre l'engagement contractuel crédible. Cette remarque permet de comprendre la réticence des économistes à considérer cette procédure comme constituant une solution satisfaisante puisqu'il apparaît que c'est centralement le recours à des valeurs différentes de celles caractérisant l'ordre économique qui permet ce résultat, et cela même si G est une somme de monnaie. Encore plus problématique est la situation lorsque G n'est plus une quantité de monnaie.

C'est le cas si la peine encourue n'est plus financière mais qu'elle consiste, par exemple, en une privation des droits civiques : inéligibilité et interdiction de participer aux votes. La valeur de ce gage dépend alors de l'implication de B dans le domaine politique. On pourra en utilisant des termes économiques dire que la crédibilité de l'engagement de B dépend de l'importance de son investissement passé dans l'ordre politique. Par delà ce vocabulaire quelque peu trompeur, il apparaît clairement que ce qui est mobilisé dans le gage est un registre de valeurs totalement étranger à l'ordre économique. On dira que c'est « l'être politique » de B qui constitue le gage. Dans ce cas, non seulement le mécanisme de saisie du gage, mais le gage lui-même, est extérieur à l'ordre économique. Si B est crédible, c'est parce que la perte de ces droits civiques apparaît comme une véritable mutilation de son être social.

L'engagement au travers du serment procède d'une logique tout à fait similaire. Il n'est pas besoin de s'y étendre longuement. Les « intérêts » mobilisés par le serment sont clairement non économiques ; et le gage est alors constitué par « l'être religieux » de B. B est crédible pour autant que le parjure apparaît comme une conduite affectant profondément B en tant qu'être pour lequel les valeurs religieuses sont importantes.

Il est une autre solution que nous n'avons pas prise en considération jusqu'à présent et qui est calquée directement sur la logique du gage. B verse une somme G auprès d'un organisme extérieur avec mission pour celui-ci de les lui rendre, possiblement avec intérêt, s'il honore sa parole, ou de les saisir s'il triche. Comme, dans le cas du contrat, cet arrangement n'est crédible que pour autant que la saisie du gage par l'organisme considéré est appréhendée par A et B comme un pur mécanisme⁸. L'efficacité de cet arrangement mobilise donc, tout comme le contrat

⁸ Le fait que l'organisme soit capable de déterminer *ex post*, sans ambiguïté, s'il y a eu tricherie ou non est une dimension importante du mécanisme, que nous n'aurons pas le temps d'exposer ici.

avec l'appareil judiciaire, un ordre de valeurs et de comportements totalement étranger à l'ordre économique.

Parce que ces diverses solutions reposent sur l'existence de rapports sociaux autres que les rapports marchands, D. Kreps a poursuivi sa recherche dans le but de mettre en évidence un arrangement dont la logique serait construite sur la seule base de la maximisation rationnelle de l'intérêt individuel. L'équilibre de réputation lui semble répondre à ces exigences. Il s'agit d'une situation qui émergerait spontanément des recherches individuelles d'une satisfaction maximale. Tout s'y ramènerait à un calcul bien compris des intérêts. Examinons cette hypothèse.

4. Critique du mécanisme réputationnel

Notons, dans un premier temps, que la logique à la base de l'équilibre de réputation est bien celle du gage. Si A_j est amené à faire confiance à B, cela provient du fait que dans le cadre d'un jeu répété avec la probabilité p , le choix de B est entre 5 s'il coopère et $(20-M)$ s'il trahit. Ce qui est ici mis en gage, c'est la réputation de B. Ce gage a pour valeur M^9 , donné par l'équation 1, à savoir la somme actualisée des revenus futurs qui reviendront à B si sa réputation reste « bonne ». En effet, perdre sa réputation, c'est perdre la possibilité de transacter plus tard avec le groupe des A_i . Il s'ensuit qu'investir dans la réputation est rentable et que la valeur de cet investissement répond à la formule traditionnelle de valorisation des titres financiers ; p y joue le rôle d'un facteur d'actualisation comme le montre l'équation déterminant la valeur de M. Tout ici est conforme à la logique marchande. Si A fait confiance à B, c'est parce qu'il sait que B n'a pas intérêt à trahir, car trahir serait se priver d'importants revenus futurs. C'est d'ailleurs cette omniprésence du calcul des intérêts qui fait dire à O. Williamson¹⁰ que le terme de confiance est une source de confusion et qu'il faut lui préférer, par souci de clarté, la notion de calcul : "I submit that calculativeness is determinative throughout and that invoking trust merely muddies the (clear) waters of calculativeness" (p. 471). Poussant l'argument jusqu'à son terme, O. Williamson fait remarquer que la situation décrite dans le jeu de la confiance est tout à fait classique. Il s'agit d'une banale décision en environnement risqué. L'individu A a le choix entre obtenir 0 avec certitude ou opter pour une loterie qui lui donne soit -5, soit 10. Comme la théorie économique sait calculer la valeur d'une loterie, on peut en déduire aisément quel sera le choix de A. Pour ce

⁹ On se place évidemment dans le cas où M est supérieur à 15, condition pour que l'équilibre de réputation existe.

¹⁰ O. Williamson, op. cité.

faire, on supposera que A est neutre par rapport au risque¹¹ et que son estimation de la probabilité que B trahisse, vaut q . Sous ces conditions, la valeur de la loterie "Faire confiance" est : $V = -5q + 10(1-q) = 10 - 15q$. Il s'ensuit que A fera confiance à B si V est plus grand que 0, i.e. si q est inférieur à $2/3$.

Si cette solution est si simple, c'est parce que O. Williamson suppose qu'on peut décrire le comportement de B en termes probabilistes. Au cours des réflexions précédentes, notre perspective d'analyse était tout autre : les choix de A et B étaient pensés comme devant être les meilleures réponses rationnelles au choix fait par l'autre. C'est de cette manière qu'a été introduite la dimension stratégique des interactions entre A et B. Les blocages rencontrés ont alors été analysés comme révélant les difficultés propres à la rationalité stratégique. Penser le comportement de B en termes probabilistes oblitère la dimension stratégique de l'interaction entre A et B. Les individus sont vus comme des mécanismes : compte tenu de ce qu'il connaît de B, A estime que sa probabilité de trahison est égale à q . La notion d'engagement n'a évidemment ici aucune pertinence puisque la représentation probabiliste ne laisse aucune place à l'intersubjectivité. C'est cette postulation d'une objectivité des penchants individuels gouvernant les choix personnels qui transforme le problème en une pure question de calcul. Nous voudrions montrer que, contrairement aux apparences, l'équilibre de réputation a recours également à un tel postulat lorsqu'il considère la valeur de p comme une donnée exogène.

Pour le comprendre, il faut revenir à la définition de p . Elle évalue la probabilité que B rencontre dans le futur d'autres membres du groupe des A_i . On peut donc considérer p comme une mesure du degré d'appartenance de B au groupe des A_i . Si ce degré est élevé, la menace de sanctions est efficace, car B fréquente assidûment ce groupe et sa réputation au sein de ce groupe compte beaucoup pour lui. Au contraire, si p est faible, cela signifie que la relation de B au groupe des A_i est lâche. Comme il ne transacte que très épisodiquement avec eux, sa réputation au sein de ce groupe n'affecte que très marginalement ses revenus et les sanctions des A_i ne sont guère crédibles. La question centrale est de savoir pourquoi p est traitée par D. Kreps comme une variable objective. Pourquoi ne pas considérer que ce degré d'appartenance résulte d'un choix de la part de B ? On retrouve donc bien ici la même critique que celle que nous faisons précédemment à propos de la détermination de la probabilité q . Le fait de demeurer au sein du groupe des A_i n'a aucune raison d'être considéré comme exogène. Dans l'univers stratégique que considère D. Kreps, B doit choisir son lieu d'existence au prorata des gains que ce choix procure.

Si p n'est plus une variable exogène, le mécanisme de réputation ne fonctionne plus. En effet, l'engagement de B est crédible pour autant que le manque à

¹¹ C'est là une hypothèse purement technique qui pourrait être levée sans difficulté autre que rendre les calculs plus longs.

gagner M est important. Mais M est important dans la mesure où le degré d'appartenance de B au groupe des A_i , mesuré par p , est élevé. Si p est une variable de choix pour B, par exemple si B peut décider de changer de région ou d'environnement social à la suite de sa transaction avec A_j , M peut baisser soudainement jusqu'à rendre la trahison profitable en rendant la menace de sanctions trop faible. C'est le cas d'un employé qui se sauve aux Bahamas avec la caisse. Autrement dit, dès lors qu'on fait de p une variable stratégique, les calculs précédents perdent leur pertinence. Mais quelle justification peut faire valoir la théorie économique pour traiter p comme une donnée objective du problème ? Aucune. En effet, p n'est en rien une donnée : elle est le produit des choix économiques de B qui a décidé, pour telles et telles raisons, de commercer au sein du groupe des A_i . Dans le cadre de l'ordre économique que considère exclusivement D. Kreps, ce choix est nécessairement la conséquence d'un calcul qui a montré à B que tel était son intérêt. On doit alors considérer une causalité inverse à celle sur laquelle D. Kreps a fait reposer son argument : la valeur du gain de coût terme est l'élément central du calcul des coûts sur la base duquel B forme ses choix et adopte une certaine valeur de p . Son appartenance au groupe des A_i est le résultat d'un calcul économique parce que, dans l'ordre économique, cette appartenance n'a d'autres fondements que l'intérêt que B trouve dans les transactions avec les A_i .

Il apparaît ainsi qu'à la base du succès du mécanisme réputationnel, se trouve l'hypothèse implicite selon laquelle l'appartenance de B au groupe des A_i peut être traitée comme une grandeur dégagée des liens stratégiques. Cependant cette hypothèse ne trouve pas de justification convaincante dans l'ordre économique. Il nous semble que c'est la notion d' « être social » qui permet de justifier rigoureusement la construction ingénieuse de D. Kreps. Ce qui est mis en gage, c'est l'ensemble des liens sociaux qu'a accumulés B et qui constituent son identité sociale. Ces liens sociaux (voisinage, parenté, métier, ...) parce qu'ils ne procèdent pas des valeurs économiques peuvent constituer un gage dont la valeur échappe à la spécularité stratégique. L'étude empirique de la manière dont est gérée la confiance, par exemple, dans les communautés chinoises du sud-est asiatique (M. Granovetter) ou au sein de la communauté des diamantaires juifs new-yorkais (J. Coleman), donne du poids à cette hypothèse. En reprenant les concepts proposés par M. Granovetter¹², on dira que la confiance dérive de l'appartenance à un certain réseau social. On peut ainsi réinterpréter la notion d' « embeddedness » de l'économie dans le social.

¹² M. Granovetter, op. cité.

5. Conclusions provisoires

Est-il possible de produire une théorie purement économique de la confiance ? Telle est la question qu'examine ce texte. C'est là une question qui nous semble fondamentale dans la mesure où la confiance constitue une dimension première du lien social. Autrement dit, c'est la capacité des valeurs économiques à rendre compte d'un lien social stable qui est ici examinée. Pour aborder ce problème, nous nous sommes attachés à comprendre l'édifice théorique développé par D. Kreps. La conclusion provisoire de ce travail, qu'il conviendra de développer plus qu'il n'est fait ici, est que la pertinence du mécanisme réputationnel ne peut être établie que dans la mesure où a été supposée préalablement l'exogénéité de la variable p . Or supposer que les liens entre l'individu B et les individus A_i répondent à une détermination exogène rigide n'a guère de sens dans un univers dominé par la logique stratégique. Seule la notion d'appartenance nous semble apte à justifier une telle hypothèse. Ainsi la stratégie réductionniste de l'économie échoue en ce que la confiance et la coopération dont elle est la condition ne peuvent y être engendrées qu'en prenant appui sur l'existence de formes sociales échappant à l'ordre économique : l'engagement repose sur la mise en gage de l'être social et de ses ressources spécifiques. Ce résultat nous semble exemplaire en ce qu'il démontre l'étroite imbrication des logiques économiques et sociologiques.