Contribution au bien public et préférences sociales : Apports récents de l'économie comportementale

Marie Claire Villeval

▶ **To cite this version:**

Marie Claire Villeval. Contribution au bien public et préférences sociales : Apports récents de l'économie comportementale. Revue Economique, 2012, 63 (3), pp. 389-420. 10.3917/reco.633.0389 . halshs-00681348

HAL Id: halshs-00681348
https://shs.hal.science/halshs-00681348

Submitted on 19 Jul 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Presidential address, 60th Conference of the French Economic Association
Revue Economique, 63 (3), 389-420, 2012.

Contribution au bien public et préférences sociales

Apports récents de l'économie comportementale

Marie Claire Villeval*

Comprendre comment la coopération émerge entre individus non apparentés offre une clé de compréhension de l'évolution des sociétés. L'analyse du rôle de l'hétérogénéité des préférences sociales dans la décision de contribution aux biens publics révèle que la coopération conditionnelle explique le déclin des contributions au cours du temps. Les règles morales peuvent cependant contrecarrer ce processus. Les comportements de sanction altruistes conduisent en effet à ce qu'un deuxième problème de passager clandestin s'avère finalement une source majeure de coopération. Si en l'absence d'institutions, une minorité d'égoïstes pousse la majorité des individus dotés de préférences sociales à se comporter en passagers clandestins, l'introduction d'institutions appropriées conduit les égoïstes à imiter les individus dotés de préférences sociales.

CONTRIBUTIONS TO PUBLIC GOODS AND SOCIAL PREFERENCES: RECENT INSIGHTS FROM BEHAVIORAL ECONOMICS

How cooperation emerges between non-kins is a key issue for understanding the evolution of societies. Exploring the role of the heterogeneity of social preferences in the decision to contribute to the provision of public goods reveals that conditional cooperation explains the decay of contributions over time. Moral rules can, however, counter this evolution. Indeed, altruistic sanctions can turn a double moral hazard problem into a major source of cooperation. While in the absence of institutions a minority of selfish individuals leads the majority of other-regarding individuals to free-ride, the implementation of appropriate institutions can lead them to behave as the selfless individuals.

Classification *JEL* : C92, C93, H41, D6, D8

INTRODUCTION

L'évolution des sociétés humaines et d'un certain nombre de sociétés animales s'explique sans doute autant par la capacité à coopérer des individus qui les composent que par la compétition. Dans la plupart de ces sociétés, la coopération est vitale pour accéder aux ressources alimentaires et pour se protéger contre les prédateurs. Et l'on peut avancer l'hypothèse que seules les sociétés qui ont su créer un degré suffisant de coopération entre

* Université de Lyon, F-69007 Lyon ; CNRS – GATE Lyon St Etienne. *Correspondance* : 93 chemin des Mouilles, F-69130 Ecully, France. *Courriel* : villeval@gate.cnrs.fr

leurs membres peuvent survivre. Pourtant toutes les espèces ne vivent pas en société. Les spécialistes de biologie de la reproduction animale avancent qu'il n'y a pas davantage de vie en société chez les espèces animales parce que les coûts sont généralement supérieurs aux bénéfices ; la vitesse de reproduction des individus est supérieure chez les animaux solitaires et l'augmentation de la taille des groupes ou des réseaux requiert une coopération coûteuse à établir et à maintenir (Aron et Passera [2008]). La coopération s'instaure généralement au sein de la parentèle en s'appuyant sur des corrélats génétiques. Elle est en revanche beaucoup plus improbable et complexe entre individus n'appartenant pas à la même parentèle. Il est fascinant de noter à cet égard que la taille du cerveau semble être corrélée à la dimension du groupe dans lequel vit l'individu (Seabright [2012]). Comprendre comment et à quelles conditions la coopération entre individus non apparentés émerge, se stabilise, décline puis disparaît constitue certainement une clé de la compréhension de l'évolution des sociétés.

L'étude de la décision de contribution volontaire à la fourniture de biens publics est étroitement associée à la compréhension de l'émergence de la coopération dans ces sociétés.[1] Outre ses propriétés de non rivalité et de non exclusion, un bien public est caractérisé par un rendement marginal individuel inférieur à celui d'un bien privé. Il existe donc un coût individuel à sa fourniture alors que chacun peut en user, indépendamment de son effort de contribution. Le comportement individuel de passager clandestin est donc une stratégie dominante et l'équilibre de Nash du jeu correspondant est une absence de contribution individuelle. La poursuite de la stratégie dominante conduit au défaut de fourniture du bien public. L'optimum social requerrait, au contraire, que tous les individus contribuent l'intégralité de leur dotation, ce qui, sans communication et sans possibilité d'engagement crédible, est très peu probable. Pourtant, nous observons une multitude de biens publics fournis volontairement, sans intervention de l'autorité publique (par exemple, l'effort individuel dans une équipe de travail, les dons à des associations, les journées de solidarité nationale, le tri sélectif des ordures ménagères, l'évaluation des vendeurs sur les plateformes de commerce électronique, etc.).

En présence d'une telle stratégie dominante, comment peut-on expliquer que des individus appartenant à des sociétés si évoluées et anonymes que les sociétés humaines contemporaines parviennent pourtant à coopérer pour fournir des biens publics ? L'accroissement des capacités cognitives des individus grâce aux investissements massifs en capital humain, ainsi que le développement des échanges au sein de réseaux de plus en plus anonymes, distants et complexes devraient pourtant favoriser la convergence rapide vers l'équilibre. Pour expliquer ces déviations durables des comportements, deux dimensions peuvent être sollicitées : (*i*) la présence de préférences sociales telles que l'altruisme, la réciprocité ou la volonté de guider le groupe vers l'optimum de contribution pleine (l'individu contribue parce que sa contribution améliore le sort de tous les autres) ;

[1] Il existe des parallèles intéressants dans certaines sociétés animales. Par exemple, le toilettage social chez les primates peut s'apparenter à une contribution au bien public. Aron et Passera [2008] montrent qu'il existe un coût individuel à cette activité (une accidentalité supérieure chez les petits par défaut de vigilance de la mère). Si cette activité entre non apparentés existe néanmoins, c'est parce qu'elle contribue à la formation de coalitions soutenant l'accès aux ressources et à la protection contre les agressions. Seyfarth et Cheney [1984] ont ainsi mesuré que la durée de réponse d'une femelle au cri d'une autre femelle non apparentée sollicitant un secours – et donc la probabilité d'entraide - dépend de l'existence préalable d'un toilettage entre les deux primates. Le mécanisme sous-jacent peut s'interpréter comme de l'altruisme et de la réciprocité.

et (*ii*) les croyances des individus sur le comportement des autres membres (l'individu contribue parce qu'il anticipe que les autres membres du groupe vont contribuer). Ces deux dimensions invoquent l'hétérogénéité des motivations dans un même groupe. Eclairer le rôle de ces deux dimensions permet d'aborder des questions fondamentales sur les conditions de la cohabitation au sein des mêmes espaces de l'Homo Reciprocans et de l'Homo Oeconomicus. Dans quelles conditions cette coexistence produit-elle une défaillance de la fourniture de biens publics ou au contraire une coopération durable ? Enfin, les groupes sont-il capables de produire par eux-mêmes des règles informelles permettant de soutenir la coopération ou celle-ci ne peut-elle émerger que sous la contrainte d'institutions formelles ?

L'économie comportementale et expérimentale s'est particulièrement intéressée à ces questions à partir de jeux de contribution volontaire à la fourniture des biens publics, afin d'identifier les mécanismes les plus efficaces pour soutenir la coopération au sein des groupes. L'objectif de cet article est d'offrir une synthèse, nécessairement partielle et orientée, des travaux conduits en ce domaine lors des quinze dernières années (voir Ledyard [1995] pour une synthèse exhaustive des expériences plus anciennes sur les jeux de bien public ; Zelmer [2003] pour une méta-analyse des jeux de bien public linéaires ; Chaudhuri [2011] pour une revue des travaux plus récents).

Alors que l'économie comportementale produit des modèles théoriques enrichissant le raisonnement des agents économiques par l'introduction de dimensions psychologiques, en particulier dans le domaine des préférences sociales, l'économie expérimentale fournit quant à elle une méthode permettant de tester les hypothèses issues de ces modèles. Croyances et préférences sont particulièrement difficiles à mesurer à partir de données naturelles. L'expérimentation offre la possibilité de générer des données sur ces dimensions dans un environnement contrôlé. Elle permet de varier les valeurs des paramètres du modèle *ceteris paribus* (en l'occurrence, le rendement marginal du bien public, la taille des groupes, ou le degré d'inégalité des dotations initiales). Elle rend possible la manipulation du degré d'information des individus sur le comportement des autres (absence d'information, information sur la moyenne des contributions ou sur les choix individuels). Elle autorise la mesure des effets de la variation des liens sociaux (création de parentèles ou maintien de liens minimaux), du temps à travers la répétition, ou encore de la manipulation des échantillons de participants pour la réalisation de comparaisons internationales. Les recherches en neuro-économie permettent enfin d'identifier les activations neuronales et leurs processus associés conduisant à une décision de contribution ou à un comportement de passager clandestin, ou encore à une décision de sanction en cas de violation de la norme du groupe.

Cette méthode soulève encore parfois des interrogations, notamment quant à la petite taille des échantillons, au niveau des incitations et à la validité externe des résultats de laboratoire (Levitt et List [2007a] [2007b]). Il convient pourtant d'admettre que si un modèle est réfuté en laboratoire, il ne peut plus prétendre à une validité générale. Dans sa réponse critique à Levitt et List, Camerer [2011] montre que la très grande majorité des études expérimentales comparant les comportements de populations étudiantes en laboratoire et d'autres populations sur le terrain conclut à une bonne comparabilité des résultats. Plus fondamentalement, il rappelle que l'économie expérimentale cherche avant tout à établir une théorie liant les règles, les incitations et les normes aux comportements, et que de ce point de vue la question de la généralisation des résultats de laboratoire aux

réalités de terrain est peu pertinente. La méthode impose toutefois des contraintes (en particulier, les participants sont conscients de participer à des expériences ; les expériences se déroulent sur un temps court, même si la répétition des choix sur un grand nombre de périodes permet d'identifier des dynamiques de groupe). Tout en conservant à l'esprit certaines limites et contraintes de cette méthode, il convient toutefois d'apprécier les nombreuses avancées permises par les travaux récents en économie comportementale et expérimentale quant à la compréhension des comportements individuels de contribution aux biens publics.

Le reste de cette contribution est structuré autour de trois parties. La première partie se consacre à l'identification du rôle respectif des préférences sociales et des croyances dans la décision de contribution volontaire au bien public. Elle met en avant le concept de coopération conditionnelle et l'importance de la prise en compte de l'hétérogénéité des préférences des joueurs pour expliquer la tendance au déclin des contributions au cours du temps. La seconde partie s'intéresse au poids des règles morales et étudie l'usage par les individus de mécanismes institutionnels permettant de soutenir la coopération dans la production des biens publics. Elle traite principalement des comportements individuels de sanctions altruistes contre la violation de la norme de coopération dans les groupes. Elle montre comment ce qui crée a priori un double problème de passager clandestin peut finalement conduire à l'optimum social. Elle souligne la complexité de la détermination des normes et les effets incertains des sanctions sur l'efficience économique. Enfin, la troisième partie s'interroge sur les modèles théoriques de préférences sociales les plus à même de rendre compte des comportements de contribution et de sanction dans ce type de dilemme social. Elle conclut sur l'importance de théoriser le concept de réciprocité forte par rapport aux théories privilégiant les considérations distributives.

IRRATIONALITE OU CONDITIONNALITE DE LA COOPERATION ? HETEROGENEITE DES PREFERENCES ET ROLE DES CROYANCES

Une coopération spontanée qui décline au cours du temps

Dans un jeu de bien public linéaire standard, les individus sont affectés à un groupe de taille N et reçoivent une dotation initiale identique (D). Ils décident de leur contribution individuelle au bien public (g_i), sachant que le rendement marginal par tête du bien public (α) est inférieur au rendement marginal du bien privé (égal à 1, par simplicité). Le montant de la dotation non contribué au bien public est affecté au bien privé. La fonction de gain de l'individu est définie ainsi :

$$\pi i = D - g_i + \alpha \sum_{1}^{n} g_j$$

Cette situation représente un dilemme social dans la mesure où l'unique stratégie dominante et l'équilibre de Nash de ce jeu correspondent à une contribution nulle au bien

public (g_i=0 puisque $\frac{\delta \pi_i}{\delta g_i} = -1 + \alpha < 1$) alors que l'optimum social requiert une contribution totale de tous au bien public (g_i=D puisque Nα>1).

Ce jeu est testé en laboratoire, en utilisant généralement les paramètres suivants : N=4, D=20, α=0.4 (les résultats sont toutefois robustes à d'autres valeurs des paramètres), et il est répété de 10 à 50 fois avec ou sans brassage des groupes (protocole d'appariement fixe ou variable) entre les périodes de jeu. Typiquement, il est observé *i*) une contribution initiale moyenne au bien public correspondant à 40% à 60% de la dotation ; *ii*) un déclin régulier de la contribution moyenne tout au long du jeu ; iii) une contribution moyenne de l'ordre de 10% de la dotation lors de la dernière période de jeu. Ainsi, en l'absence d'institution et de possibilité de mobilité des joueurs entre les groupes, la norme qui tend à s'imposer au fil du temps est celle qui correspond à la stratégie dominante.

Ce déclin des contributions au cours du temps a généré un très grand nombre de travaux depuis les années quatre-vingts. Les explications ont longtemps privilégié la confusion des joueurs (Andreoni [1988]), la dynamique d'apprentissage du jeu (Binmore [2005], Andreoni et Croson [2008]), l'existence d'erreurs de décision (Palfrey et Prisbrey [1997], Anderson, Goere et Holt [1998]). Pourtant, ces explications d'ordre cognitif ont du mal à résister à l'observation d'un effet de redémarrage du jeu, aux effets de changement du mode d'appariement ou à l'introduction dans le jeu d'un équilibre intérieur. En effet, la relance d'une série de périodes de jeu non annoncées à l'avance aux joueurs se traduit par une reprise immédiate de la coopération à un niveau proche de la contribution en première période de jeu même si le groupe avait convergé vers l'équilibre à la dernière période précédente. Ceci n'est pas compatible avec une explication en termes de confusion ou d'apprentissage de la stratégie dominante. De même, l'introduction d'un équilibre intérieur, à l'aide d'une fonction de gain quadratique par exemple, permet de tester l'importance de la confusion : en effet, si les déviations par rapport à la stratégie dominante s'expliquent par des erreurs, ces erreurs devraient se situer en proportions égales au-dessus et en-dessous de l'équilibre intérieur. Or, cette égalité est rejetée car les individus sur-contribuent dans ce jeu comme dans le jeu à stratégie dominante de non-contribution. A l'aide de la méthode stratégique, Brandts et Schram [2001] définissent des fonctions de contribution en présentant aux participants divers taux de substitution faisant de la contribution une stratégie dominante et efficiente, une stratégie dominante mais non efficiente ou une stratégie efficiente mais dominée. Quelles que soient leurs préférences, tous les individus devraient coopérer dans le premier type de situation mais pas dans le second type de situation. Les déviations permettent ainsi d'identifier les erreurs de décision. Les résultats montrent cependant que la confusion ne peut pas expliquer la plupart des comportements. Enfin, en tentant de séparer le rôle de la bienveillance de celui de la confusion, Andreoni [1995] ne peut rejeter la responsabilité des deux phénomènes dans le choix de contribution. Mais il explique le déclin de la coopération au cours du temps plutôt par l'échec répété de tentatives de comportements bienveillants que par l'apprentissage du *free-riding*.

Aujourd'hui, les explications dominantes des sur-contributions mettent l'accent sur les croyances sur les intentions des autres et sur les préférences sociales. Une interprétation en termes de croyances suppose que les individus font des prédictions sur la volonté des autres membres de leur groupe de jouer ou non leur stratégie dominante et sur les

croyances de ces autres membres sur les croyances et intentions de l'individu (pour un modèle de jeu psychologique reposant sur les intentions et les croyances, voir Rabin [1993], cf. infra). Un individu peut dès lors décider de contribuer au bien public s'il pense qu'un suffisamment grand nombre de membres de son groupe vont également contribuer.

De leur côté, les interprétations en termes de préférences sociales mettent en avant notamment l'altruisme ou le *warm glow* (Andreoni [1990], Palfrey et Prisbrey [1997]). Des individus contribuent au bien public parce qu'ils aiment faire le bien ou parce que leur utilité est accrue par un gain d'égo lié au fait de donner. Ces explications admettent une diversité des préférences sociales possibles des membres d'un groupe mais ne laissent pas de place à la conditionnalité. Or, plusieurs expériences récentes ont pu identifier la coexistence d'altruistes, de passagers clandestins et de coopérateurs conditionnels (ou individus « réciproques »). Fischbacher et Gächter [2010] définissent comme altruiste un individu qui contribue en moyenne la totalité de sa dotation et dont la pente de régression de la contribution en fonction des contributions moyennes des autres membres de son groupe est nulle. A l'opposé, un passager clandestin est défini comme un individu qui, en moyenne, ne contribue rien au bien public et pour lequel la pente de régression de la contribution en fonction des contributions moyennes des autres membres de son groupe est également nulle. Un coopérateur conditionnel est défini comme un individu qui contribue en moyenne la moitié de sa dotation et pour lequel la pente de régression est égale à 1 ; le coefficient de corrélation de Spearman entre la contribution de l'individu et celle des autres membres de son groupe doit être positif et significatif à 1%. Enfin, une catégorie de « contributeurs triangulaires » est identifiée : ces joueurs accroissent leur contribution avec la contribution moyenne des autres jusqu'à un certain seuil au delà duquel leur contribution décline avec la contribution des autres ; il est également requis que le coefficient de corrélation de Spearman entre la contribution de l'individu et celle de son groupe soit significatif à 1% à la fois en deçà et au delà du seuil. Fischbacher et Gächter [2010] identifient ainsi dans la population de joueurs une majorité de coopérateurs conditionnels (55%), une minorité de passagers clandestins (23%) et de contributeurs triangulaires (12%), les 10% restant incluant des altruistes et des joueurs aux stratégies indéfinissables.

A l'aide d'une autre technique d'identification par laquelle chaque membre du groupe peut successivement modifier sa contribution initiale après avoir été informé de la contribution moyenne des autres joueurs de son groupe, Kurzban et Houser [2005] identifient une répartition des participants très proche de celle de Fischbacher et Gächter [2010] : 63% des participants sont identifiés comme étant des coopérateurs conditionnels, 20% comme des passagers clandestins et 13% comme des coopérateurs non conditionnels. Burlando et Guala [2005] utilisent une autre méthode, incluant notamment des questionnaires sur les valeurs, et identifient 35% d'individus réciproques, 32% de passagers clandestins et 18% de coopérateurs.

Il convient de définir si les explications en termes de croyances et celles en termes de préférences sociales sont complémentaires ou substituables et si elles permettent de rendre compte à la fois de la déviation des joueurs de leur stratégie dominante et du déclin des contributions moyennes au fil des répétitions du jeu.

Préférences sociales ou croyances sur les intentions des autres ?

Un test direct du rôle respectif des croyances et des préférences sociales dans les décisions de contribution au bien public a été proposé par Fischbacher et Gächter [2010] pour rendre compte du déclin de contributions au fil des répétitions du jeu. Le protocole expérimental inclut deux traitements qui permettent, d'une part, de confronter les deux explications jusqu'à présent étudiées séparément et, d'autre part, d'identifier directement les motivations des joueurs au lieu de les inférer à partir des contributions observées. Dans l'expérience « Préférences » jouée en méthode stratégique[2], les joueurs prennent une décision de contribution non conditionnelle puis une série de décisions conditionnelles à chaque niveau possible de contribution moyenne des autres membres de leur groupe (entre 0 et 20). Il est ainsi possible d'identifier les préférences sociales de chacun à partir de sa contribution moyenne et de la pente de régression de sa contribution par rapport aux contributions des autres. Jouée en méthode directe, l'expérience « Croyances » sert à éliciter les croyances des joueurs sur la contribution moyenne des membres de leur groupe et à étudier le lien entre ces croyances et les décisions de contribution, contrôlant pour les préférences sociales des individus. Les joueurs déclarent leur croyance sur le comportement moyen des autres puis prennent leur décision de contribution. Ce jeu est répété dix fois avec un réappariement des joueurs au début de chaque nouvelle période, afin d'éviter les effets de réputation.

L'analyse économétrique des résultats montre que plus la croyance est élevée, plus la contribution individuelle augmente mais contrôlant pour ces croyances, tous les groupes contribuent plus que les passagers clandestins. Ceci signifie que, une fois les croyances prises en compte, l'hétérogénéité des profils de joueurs en termes de préférences sociales conserve le plus grand pouvoir explicatif. En d'autres termes, si les passagers clandestins ont un comportement égoïste, ce n'est pas tant parce qu'ils pensent que les autres sont égoïstes mais parce qu'ils sont réellement égoïstes ; de même, si les altruistes contribuent la totalité de leur dotation, ce n'est pas parce qu'ils pensent que tous les autres sont altruistes mais parce qu'ils veulent assurer la fourniture du bien public.

Pour déterminer si le déclin des contributions au cours du temps est davantage dû à l'évolution des croyances ou aux préférences sociales, les auteurs se livrent à plusieurs exercices de simulation à l'aide de contrefactuels sur les contributions (supposant une coopération parfaitement conditionnelle) et les croyances (supposant des croyances naïves). Il apparaît ainsi que le processus de formation des croyances n'explique pas le déclin de la coopération. C'est parce que les contributions déclinent que les croyances sont révisées en baisse. Ce sont donc bien les préférences des individus qui expliquent le déclin des contributions. Ceci résulte de la combinaison de deux éléments. La présence de coopérateurs conditionnels justifie que toute contribution révisée à la baisse par un individu au sein du groupe entraine la révision à la baisse des contributions de ces coopérateurs. Mais ceci n'est pas suffisant. Il faut en effet la présence d'un élément additionnel qui rend la coopération conditionnelle imparfaite : il s'agit du biais d'auto-complaisance qui pousse les coopérateurs conditionnels à s'ajuster à la moyenne des autres moins un epsilon afin de gagner davantage.

[2] Sous la méthode « stratégique » (proposée initialement par R. Selten), l'individu prend une décision pour chaque décision possible d'un ou plusieurs autres individus. Elle n'exclut en aucun cas l'existence d'incitations. Elle s'oppose à la méthode dite « directe » où chaque individu prend une décision simultanément.

Au total, ces travaux montrent : *i*) l'existence d'une grande hétérogénéité des préférences sociales, avec une majorité de coopérateurs conditionnels (Burlando et Guala [2005], Fischbacher, Gächter et Fehr [2001], Kurzban et Houser [2005], Bardsley et Moffatt [2007], Kocher et al. [2008], Fischbacher et Gächter [2010]) [3]; *ii*) une très forte cohérence entre préférences et comportements, laquelle s'accroît au fil du temps. C'est l'hétérogénéité des préférences sociales, en plus de l'information sur les comportements des autres, qui explique la baisse temporelle des contributions. Une conclusion majeure est qu'en l'absence d'institutions, la stratégie dominante de non-contribution finit par s'imposer, *même si au départ il existe une majorité d'individus prêts à coopérer* et seulement une minorité de passagers clandestins. Cela suggère aussi qu'une modification des règles de base du jeu peut aider les coopérateurs conditionnels à élever leur niveau de contribution au fil du temps au lieu de l'abaisser.

Une plus grande homogénéité des préférences favorise t-elle la coopération ?

En raison de l'importance de la proportion des coopérateurs conditionnels dans la population identifiée par les précédents travaux, il convient de tester si la fourniture de biens publics est mieux assurée lorsque la composition des groupes est plus homogène en termes de préférences sociales. La méthode expérimentale permet de mesurer l'effet sur le niveau des contributions de la modification de la composition des groupes de manière exogène ou de manière endogène.

Certains protocoles expérimentaux manipulent de manière exogène la formation de groupes homogènes en termes de préférences sociales. Ainsi, après chaque période de jeu ou après un premier ensemble de périodes avec composition aléatoire des groupes, le programme informatique regroupe les joueurs en fonction de leurs contributions relatives passées. Ce principe de recomposition des groupes fait l'objet d'une connaissance commune (Burlando et Guala [2005], Gächter and Thöni [2005], de Oliveira, Croson et Eckel [2009]) ou pas (Gunnthorsdottir, Houser et McCabe [2007]). Les coopérateurs inconditionnels sont regroupés entre eux, de même que les coopérateurs conditionnels, ainsi que les passagers clandestins. Dans tous les cas, en raison de l'exclusion des passagers clandestins du groupe des coopérateurs, le niveau de coopération moyen est très accru et ce sont les contributions moyennes au sein du groupe des coopérateurs conditionnels qui se rapprochent le plus de l'optimum social. Plus encore, le niveau de coopération est stabilisé durant la totalité des périodes de jeu, mettant fin au déclin habituel des contributions au cours du temps. Ces résultats montrent que ce déclin observé dans le jeu standard est principalement le fait des coopérateurs qui perdent confiance au contact des passagers clandestins. En revanche, le niveau moyen de contribution des passagers clandestins reste très faible tout au long du jeu. L'effet de la recomposition des groupes par affinité ainsi identifié est encore renforcé chez les coopérateurs lorsque l'information sur le principe de recomposition est commune car la connaissance du partage des mêmes valeurs au sein du groupe facilite la coordination sur un niveau élevé de contribution.

[3] D'autres sources d'hétérogénéité ont également été identifiées. Ainsi la génération ou l'âge influencent les normes, les enfants de 6 à 12 ans contribuant au même niveau que les adultes mais augmentant leur contribution au cours du temps (Harbaugh et Krause [2000]) et les individus plus âgés (au delà de 50 ans) étant plus prônes à la coopération que les plus jeunes (Charness et Villeval [2009]) ; En revanche, on n'a pas identifié jusqu'à présent d'effet univoque du genre (Croson et Gneezy [2009]).

D'autres protocoles expérimentaux laissent les groupes se former de manière endogène. Dans Page, Putterman et Unel [2005], les participants ont la possibilité de classer par ordre de préférence chacun des autres participants de leur session en fonction de leur contribution. Le programme informatique reforme ensuite les groupes en respectant au mieux les préférences exprimées et en maintenant constante la taille des groupes. L'efficience est significativement accrue en raison de l'élévation de la contribution moyenne des coopérateurs conditionnels. En revanche, les effets sont plus mitigés lorsque les individus sont libres de quitter leur groupe pour rejoindre un autre groupe ou fonder un nouveau groupe, moyennant un coût de mobilité (Ehrard et Keser [1999]). En effet, alors que l'optimum est ici la grande coalition puisque le gain social maximal augmente avec la taille du groupe, celle-ci n'est jamais atteinte. On assiste à un mouvement permanent de création de nouveaux groupes par des coopérateurs fuyant les groupes à niveau de coopération élevé mais systématiquement envahis par les passagers clandestins. Les comportements diffèrent lorsque les joueurs ont à la fois la possibilité de quitter leur groupe (comme dans Ehrard et Keser [1999]) et celle d'exclure des membres de leur groupe (Charness et Yang [2008]). En effet, avec l'introduction d'un droit d'exclusion, la formation de la grande coalition avec un niveau élevé de coopération est fréquente. Cela est dû au fait que les passagers clandestins se mettent à imiter le comportement contributif des coopérateurs. Ce résultat est cohérent avec celui de Cinyabuguma, Page et Putterman [2006] qui introduisent un droit de vote pour expulser des membres du groupe. Un résultat important est donc qu'avec la sélection, la coopération s'impose *même si une fraction significative d'individus préfèrerait adopter un comportement égoïste de passager clandestin*.

Pour résumer, dans un jeu où la stratégie dominante est l'absence de contribution au bien public, la présence de préférences sociales plus que les croyances sur les comportements des autres conduit à une déviation systématique hors de l'équilibre. Toutefois, l'hétérogénéité des préférences sociales des individus et l'importance de la coopération conditionnelle accompagnée d'un biais d'auto-complaisance expliquent le mouvement de déclin inéluctable des contributions vers l'équilibre au fur et à mesure des répétitions du jeu. Les travaux récents ont également montré que des recompositions de groupes plus homogènes en termes de préférences sociales modifient les comportements et annulent la tendance au déclin temporel de la coopération qui est attribuable principalement aux coopérateurs conditionnels. L'introduction d'un droit d'exclusion conduit même les passagers clandestins à imiter les comportements des coopérateurs. Ceci conduit à s'interroger sur la production des normes et la capacité des groupes à les faire respecter.

REGLES MORALES ET COOPERATION OU « LE MYSTERE DES SANCTIONS ALTRUISTES »

Il n'y pas de sociétés sans normes. En matière de biens publics, la norme initiale – celle partagée par la majorité des membres d'un groupe - est la coopération conditionnelle ; en revanche, la norme qui s'impose à défaut d'institutions est l'absence de contribution au bien public. Elle correspond à l'équilibre du jeu. Dans cet environnement très épuré, les individus n'ont comme seul moyen de réaction au comportement des autres que leur propre décision de contribution. Toutefois, dans la réalité, les individus ont davantage de moyens

d'action pour essayer de changer le comportement des autres et les normes. Ils peuvent menacer, sanctionner ou exercer des représailles à l'encontre des membres de leur groupe qui ne respectent pas la norme du groupe. Par exemple, des enfants peuvent ostraciser un des leurs qui refuse systématiquement de partager, un employé peut cesser de vouloir travailler avec un co-équipier tire-au-flanc, des Etats membres d'une Union peuvent faire pression contre un des leurs qui compromet la santé économique du groupe en ne s'engageant pas dans des mesures d'assainissement budgétaire, etc.. Ces sanctions étant généralement coûteuses pour celui qui y recourt, leur usage peut toutefois surprendre. Elles soulèvent en effet un deuxième problème de passager clandestin. Les économistes comportementalistes se sont alors largement investis dans l'exploration de l'usage des sanctions, de ses déterminants et de l'impact des sanctions sur le respect des normes de contribution. En outre, les préférences sociales étant hétérogènes, ils se sont interrogés sur la capacité des groupes à se coordonner sur la même norme.

Les sanctions : un deuxième problème de passager clandestin ou une solution ?

Yamagishi [1988] et Ostrom, Walker et Gardner [1992] ont initié l'hypothèse que les sanctions coûteuses au sein d'un groupe peuvent contribuer à la lutte contre le comportement de passager clandestin, tout en montrant aussitôt que ces sanctions soulèvent un problème de passager clandestin de second ordre. En effet, même si les individus anticipent un impact de la sanction sur les futures contributions au sein du groupe, ils n'ont pas intérêt individuellement à en supporter le coût. Plus récemment, la contribution majeure de Fehr et Gächter [2000] a pourtant établi que, bien que hors équilibre, la possibilité d'attribuer des sanctions non seulement est utilisée par les joueurs mais permet d'atteindre l'optimum social.

Le protocole de Fehr et Gächter [2000] introduit dans le jeu standard de bien public linéaire présenté supra une seconde étape de jeu. Dans la première étape, les individus décident de leur contribution au bien public. A la fin de cette étape, les individus sont informés de la contribution individuelle de chacun des autres membres du groupe. Dans la seconde étape, ils ont la possibilité d'attribuer des points de sanction (P) aux autres membres. Chaque point attribué par i à j réduit de 10% le gain de première étape du joueur j. Un joueur sanctionné ne peut toutefois pas perdre plus de 100% de son gain de première étape. Attribuer des points de sanction est également coûteux pour celui qui les attribue (K par point attribué) ; la fonction de coût est convexe. La fonction de gain des joueurs en fin de deuxième étape s'écrit donc ainsi :

$$\pi_i^2 = \left(D - g_i + \alpha \sum_{i=1}^{N} g_j \right) * \frac{\max\left\{0{,}10 - \sum_{j \neq i} P_{ji}\right\}}{10} - \sum_{j \neq i} K(P_{ij})$$

Dans une version ultérieure (Fehr et Gächter [2002]), la fonction de coût de sanction est simplifiée : chaque point de sanction attribué par i à j réduit le gain de première étape du joueur j de 3 unités. La fonction de gain des joueurs en fin de deuxième étape s'écrit donc ainsi :

$$\pi_i^2 = \left(D - g_i + \alpha \sum_{i=1}^{N} g_j \right) - e \sum_{j \neq i} P_{ji} - \sum_{j \neq i} K(P_{ij})$$

Quelle que soit la fonction retenue, l'équilibre du jeu initial n'est pas modifié. La stratégie dominante consiste à ne pas sanctionner en deuxième étape du jeu et, la sanction n'étant pas crédible, à ne pas contribuer en première étape du jeu. Or, le comportement observé en laboratoire s'écarte significativement et durablement de cette prédiction. En effet, les individus sanctionnent d'autres joueurs et en subissent le coût, même lorsque les membres du groupe ne jouent qu'une seule fois et que l'on brasse la composition des groupes d'une répétition à l'autre. Ces sanctions sont qualifiées de « sanctions altruistes ». L'impact de ces sanctions est majeur puisqu'au lieu de converger vers l'équilibre de Nash, toutes les contributions convergent vers l'optimum social. Sur l'ensemble des répétitions du jeu, la contribution moyenne est de 19% de la dotation initiale en l'absence de sanction et 58% lorsqu'il est possible de sanctionner. Lorsque les groupes sont fixes, l'optimum est atteint au bout de quatre périodes. Ces résultats s'expliquent par le fait que les individus sanctionnés élèvent leur contribution à la période suivante, *y compris lorsqu'ils sont réalloués à un groupe différent*.

Ce résultat majeur a été reproduit dans de nombreuses expériences, y compris en donnant la possibilité de punir un membre d'un autre groupe que le sien (donc sans perspective de retour personnel) ou en donnant l'opportunité à un tiers de sanctionner (Carpenter et Matthews [2008], Bochet, Page et Putterman [2006]). La décision de punition par un tiers non intéressé monétairement au comportement du groupe peut s'expliquer par un effet d'indignation suscité par la violation d'une norme en dehors de toute considération de réciprocité. Un autre résultat majeur est que l'usage de sanctions non monétaires (en l'occurrence la distribution de points de désapprobation sans effet sur les gains) a également un effet d'incitation sur les individus qui les reçoivent : le fait de se voir adresser un signal de désapprobation par les membres de son groupe élève significativement la contribution de l'individu sanctionné lors de la période suivante, *même si les groupes sont reformés* (Masclet, Noussair, Tucker et Villeval [2003] ; voir également Balafoutas et Nikiforakis [2011] pour une expérience de terrain sur ce thème). Ceci montre que les individus sont sensibles aux signaux, monétaires et non monétaires, réprimant la transgression d'une norme. Avec les sanctions, même des égoïstes rationnels se mettent à coopérer, *y compris en dernière période de jeu.*

Les sanctions altruistes : un comportement irrationnel ?

Bien que situés hors équilibre, il serait erroné de considérer ces comportements de sanction comme irrationnels, et ce pour au moins quatre raisons : il est possible d'identifier la norme qui les déclenche ; ils répondent à la loi de la demande ; ils créent une utilité émotionnelle immédiate ; ils suscitent une réaction en retour de la part des individus sanctionnés qui permet d'augmenter les gains futurs des groupes.

L'intensité des sanctions dépend linéairement de l'importance de la déviation négative entre la contribution de l'individu qui reçoit la sanction et la contribution moyenne des autres membres du groupe. La norme dans le jeu de bien public avec sanction est clairement identifiée comme étant le comportement moyen ou médian du groupe mais ni le

minimum, ni le maximum des contributions (Croson [2007]). La sanction punit la violation d'une norme endogène et spécifique au groupe et non pas la violation d'une norme générique (par exemple la contribution correspondant à l'optimum social).

La distribution de sanctions répond à la loi de la demande (Anderson et Putterman [2006], Carpenter [2007]). A l'aide d'un protocole dans lequel chaque joueur est assuré de ne pas interagir plus d'une fois avec les mêmes autres personnes (écartant ainsi les sanctions stratégiques), Anderson et Putterman [2006] ont manipulé de manière aléatoire le coût de la sanction pour celui qui punit avec une gamme large de prix incluant un prix maximum de 25 fois le prix minimum. Ils ont ainsi pu montrer que pour une déviation à la norme donnée, l'intensité des sanctions est corrélée à son prix pour celui qui punit. Le plus grand nombre de points de sanction est attribué aux sujets qui dévient le plus de la norme du groupe lorsque sanctionner est gratuit ou très peu coûteux ; le plus faible nombre de points est attribué aux mêmes déviations lorsque le coût de la sanction est le plus élevé. La courbe de demande de sanctions a naturellement une pente négative. Lorsque l'attribution de sanctions est gratuite pour celui qui les inflige, les sanctions sont toujours orientées à l'encontre des passagers clandestins. Il est à noter que ces résultats montrent que les sanctions répondent à un raisonnement économique compatible avec une théorie des choix rationnels. La volonté de punir s'avère être une préférence assez conventionnelle, qui crée une satisfaction et réagit aux prix.

Les recherches précédentes montrent que la sanction est un acte délibéré, non automatique. Plusieurs recherches en neuro-économie ont tenté d'expliciter les processus neuronaux engagés dans l'acte de punition dans le cadre de dilemmes sociaux. Si l'hypothèse d'une valeur neurale commune entre monnaie et morale n'est pas réfutée, payer pour sanctionner peut créer de l'utilité. Recourant à la tomographie par émission de positrons (TEP, méthode reposant sur la détection de particules émises par une substance radioactive administrée au sujet), de Quervain et al. [2004] ont montré dans le cadre d'un jeu de confiance que les sanctions coûteuses activent le striatum, structure du cerveau impliquée dans le traitement des récompenses chez les primates humains et non humains. Les individus qui présentent la plus forte activation du striatum en punissant sont ceux qui acceptent de payer le plus pour punir. Il est aussi montré que chez les individus qui ont la plus forte volonté de punir, le cortex ventromédian est plus activé quand la punition engendre un coût pour celui qui l'inflige que lorsqu'elle est sans coût. Ces résultats soutiennent l'hypothèse que la punition de la violation des normes crée de la satisfaction et que l'activation du striatum reflète la satisfaction anticipée de la punition. Comme le notent les auteurs, ceci illustre en quelque sorte « le goût suave de la revanche ». De manière complémentaire, Joffily, Masclet, Noussair et Villeval [2011] ont étudié les réponses électrodermales des participants à un jeu de bien public linéaire avec sanctions. Ils identifient un cercle vertueux des émotions : la coopération engendre une émotion positive alors que le comportement de passager clandestin suscite des émotions négatives ; contrôlant pour l'effet négatif sur la valence des émotions de la déviation à la norme de contribution, une punition plus sévère engendre une émotion positive chez les sujets qui sanctionnent ; les individus sanctionnés augmentent d'autant plus leur contribution à la période de jeu suivante que la réception de la sanction a engendré une émotion négative. Cette dynamique émotionnelle montre que sanctionner crée une utilité morale chez une fraction significative des individus qui compense son coût monétaire, voire augmente avec lui.

Une dernière explication des comportements de sanction altruistes provient de l'impact de ces sanctions sur les individus à qui elles ont été infligées. Il est montré dans toutes les expériences de bien public avec sanction que les sanctions conduisent à un plus grand respect de la norme dans les transactions ultérieures. Une réaction monotone a été identifiée entre le ratio de coût de la sanction et les contributions ultérieures. Ce ratio est calculé comme le rapport entre le coût de la sanction pour l'individu qui punit et le coût de la sanction pour l'individu qui est puni. Nikiforakis et Normann [2008] font varier ce ratio entre 1:1, 1:2, 1:3 et 1:4. L'impact d'un ratio plus élevé sur les contributions est très fort mais surtout, les gains d'efficience, mesurés par la somme des gains des membres du groupe, ne sont accrus que si le ratio est très élevé. En dessous d'un certain seuil, l'opportunité de sanctionner ne parvient plus à empêcher le déclin des contributions au cours du temps, comme dans les jeux sans sanction. Les sanctions peuvent ainsi être définies comme un bien normal et inférieur. A l'aide d'une expérience réalisée sur Internet avec un échantillon représentatif de la population Néerlandaise, Egas et Riedl [2008] ont aussi montré qu'un mécanisme de sanction impliquant un coût faible pour celui qui punit et un coût fort pour celui qui est puni est le seul qui permette une stabilité de la coopération à un haut niveau dans la durée.

Y a t-il unicité de la norme de contribution ?

Les travaux mentionnés précédemment montrent que les individus punissent les déviations à la norme identifiée comme la moyenne ou la médiane des contributions du groupe, mais l'identification de la norme de contribution au bien public est-elle toujours aussi clairement définie pour les agents ? Rien ne garantit en effet qu'en matière de production de bien public, l'unicité de la norme soit le cas le plus fréquent.

Une manière de mesurer l'impact de la pluralité des normes consiste à introduire une hétérogénéité des ressources des individus qui composent un groupe à travers une variation de leur dotation initiale. Les expériences ne délivrent pas de conclusion définitive. Cherry, Kroll et Shogren [2005] observent que l'hétérogénéité des dotations initiales décroît le niveau moyen des contributions, que les dotations soient attribuées de manière exogène ou qu'elles soient gagnées par les joueurs. Au contraire, Buckley et Croson [2006] identifient un effet agrégé positif dans la mesure où les joueurs les moins riches contribuent le même montant que les plus riches mais un pourcentage supérieur de leur revenu. Chan et al. [1999] montrent quant à eux qu'en l'absence de communication, l'hétérogénéité à la fois en termes de dotations et de préférences accroît le niveau moyen des contributions alors que l'hétérogénéité dans une seule de ces dimensions n'exerce pas d'effet significatif. Dès lors, il est difficile de prédire de quelle norme la déviation sera sanctionnée lorsque les ressources initiales sont différenciées. Reuben et Riedl [2009] tentent d'identifier deux logiques possibles. Une norme peut en effet être celle d'une égale contribution de chaque joueur indépendamment de sa richesse initiale, mettant en avant le principe d'égalité. Une norme alternative peut être celle d'une contribution proportionnelle à la dotation perçue, mettant en avant le principe d'équité. Il s'agit alors de déterminer si en cas de conflit de normes, une norme unique peut parvenir à s'imposer et dans ce cas, laquelle est la plus probable. Au sein de groupes de trois joueurs, Reuben et Riedl dotent un des joueurs d'une dotation deux fois plus élevée que celle perçue par les deux autres joueurs. Ils comparent deux traitements, l'un avec plafonnement des contributions pour les plus richement dotés (au niveau de la dotation des plus pauvres) et l'autre sans plafonnement. En l'absence de

plafonnement, la norme de contribution proportionnelle s'impose dans la plupart des groupes. En revanche, en présence d'un plafonnement, les joueurs plus richement dotés ne contribuent pas davantage que les autres et la norme qui émerge est alors celle d'une contribution égale de tous. Cette expérience montre que la norme qui s'impose en matière de contribution au bien public est définie contextuellement et que les membres d'un groupe aux ressources hétérogènes peuvent parvenir à s'accorder sur une même norme en dépit du fait que certains joueurs sont avantagés.

Une autre manière d'aborder la question de la multiplicité des normes de contribution est de considérer les différences culturelles. Un avantage de la méthode expérimentale réside dans la possibilité de répliquer exactement le même protocole dans des endroits différents. Certes, le modèle de l'homo oeconomicus devrait s'appliquer de manière universelle. Pourtant, issue de la coopération entre anthropologistes et économistes, la recherche réalisée dans 15 « petites sociétés » et rapportée dans Heinrich et al. [2004] montre que les comportements, notamment dans des jeux de négociation, diffèrent selon les tribus en fonction du mode d'organisation de ces sociétés, notamment le degré de coopération dans la production, et du degré d'intégration du marché. Il n'est donc pas sans pertinence de s'interroger pour savoir si les normes et les comportements de sanction des déviations sont universels ou culturellement différenciés. Gächter, Herrmann et Thöni [2010] ont ainsi répliqué le jeu standard de bien public linéaire avec sanctions dans des groupes fixes dans des pays rattachés à six cultures identifiées au sens de systèmes de valeurs (Inglehart et Baker [2000]) : des pays de culture anglo-saxonne (Angleterre, Etats-Unis, Australie), de culture protestante (Allemagne, Suisse, Danemark), de culture orthodoxe/ex-communiste (Russie et Biélorussie), de culture confucéenne (Chine, Corée), de culture européenne méditerranéenne (Grèce et Turquie) et de culture arabophone (Oman, Arabie Saoudite). Les résultats de l'étude sont fascinants. Dans toutes les cultures est constaté le même déclin de la contribution dans le traitement sans opportunité de sanctionner les autres joueurs (confortant ainsi les résultats de l'étude sur l'universalité des comportements de Brandts, Saijo et Schram [2004]). En revanche, des différences majeures apparaissent dès lors qu'est introduite l'opportunité de sanctionner. En effet, dans les cultures du sud de l'Europe et dans les pays arabes il n'y a pas de différence significative de contribution dans les traitements avec ou sans sanctions alors que cette différence est très marquée dans tous les autres lieux d'expérimentation.

Gächter, Herrmann et Thöni [2010] montrent que les normes de sanction, de même que l'efficacité des sanctions sur les contributions au bien public, diffèrent selon les cultures. Les dépenses de punition des passagers clandestins sont similaires selon les cultures mais dans les cultures du sud de l'Europe, arabes et orthodoxes/ex-communistes, est observé un recours important aux sanctions antisociales (définies comme le fait de punir les individus qui contribuent plus que soi alors même que ces individus augmentent le gain de chacun – on parle de « sanctions perverses » quand la punition vise un joueur qui contribue plus que la moyenne des autres joueurs, voir Cinyabuguma, Page et Putterman [2006]). L'effet disciplinaire des sanctions prosociales s'avère alors très inférieur dans les cultures utilisant le plus les sanctions antisociales. Utilisant les mêmes données, Herrmann, Thöni et Gächter [2008] identifient plusieurs origines aux sanctions antisociales. Un plus vif sens de l'honneur dans certaines cultures et une honte supérieure déclenchée par la réception de sanctions peuvent entraîner une vengeance aveugle (car le sujet ignore qui l'a puni) mais orientée contre ceux qui contribuent davantage. Les cultures caractérisées par des normes de coopération civique plus lâches (telles que définies par le World Value Survey) et une

moindre force de la loi en raison d'une plus faible confiance dans les institutions sont également celles où les sanctions sont moins utilisées contre les passagers clandestins et davantage utilisées contre les coopérateurs. Par une analyse de décomposition de variance, Gächter, Herrmann et Thöni [2010] montrent que dans le traitement sans sanctions, 3,9% de la variance est expliquée par la culture, 29,3% par l'hétérogénéité des groupes et 16% par l'hétérogénéité individuelle (effets fixes individuels). Dans le traitement avec sanctions, 21,3% de la variance est expliquée par la culture, 35,4% par l'hétérogénéité des groupes et 12,7% par l'hétérogénéité individuelle. Il est essentiel de prendre en compte la dimension culturelle pour comprendre les comportements de contribution au bien public en présence de sanctions.

Sanctions et efficience

Si les sanctions permettent le développement de la coopération au sein des groupes, sont-elles pour autant efficientes ? Plusieurs recherches ont identifié des effets négatifs à court terme des sanctions. En effet, les sanctions réduisent les gains à la fois de celui qui punit et de celui à qui la sanction est infligée. La somme des gains des joueurs étant ainsi réduite, l'efficience est diminuée. Dans un simple dilemme du prisonnier, Dreber et al. [2008] ont ainsi montré que les gagnants dans ce type de jeu sont ceux qui ne punissent jamais. Par ailleurs, l'introduction d'un droit de sanction change les relations au sein d'un groupe. Dans le cas d'un jeu de confiance Fehr et Rockenbach [2003] ainsi que Houser, Xiao et Smith [2008] ont montré que la présence de menaces de sanction génère un changement cognitif créant une éviction de la bonne volonté intrinsèque de respect de la norme. De leur côté, Masclet, Noussair et Villeval [2011] ont étudié si l'usage de menaces préalable à la phase de contribution permet d'élever le niveau de coopération sans avoir à subir les conséquences négatives de l'application des sanctions en termes d'efficience. Ils ajoutent une étape préalable à un jeu de bien public avec sanctions lors de laquelle les joueurs adressent aux autres membres de leur groupe des menaces de sanction pour chaque niveau possible de contribution. Ces menaces n'étant pas crédibles, l'équilibre du jeu n'est pas modifié. Les résultats montrent toutefois que si les menaces élèvent bien le niveau de contribution au sein des groupes, elles augmentent également le niveau moyen de sanction pour un niveau de contribution donné : le fait de menacer pousse les joueurs à appliquer leurs menaces en sanctionnant davantage.

Jusqu'à présent les recherches évoquées ont traité de sanctions anonymes. Or, dans la réalité l'individu qui punit ne peut pas toujours rester anonyme. La levée de cet anonymat peut alors engendrer des phénomènes de vengeance qui risquent de réduire la propension à coopérer des joueurs. Ceci peut être testé en ajoutant au jeu standard une étape supplémentaire de sanction. Si chaque joueur est informé du nombre de points de sanction qui lui ont été attribués par chacun des autres membres de son groupe et qu'il peut ensuite renvoyer des points de sanction aux joueurs qui l'ont initialement puni, on observe des comportements de vengeance et de contre-punition (Nikiforakis [2008]). Dans cet environnement, la probabilité d'une sanction initiale est réduite car les joueurs anticipent une réaction négative en retour du joueur sanctionné. Au total le niveau des contributions et celui des gains sont réduits par rapport au jeu standard avec sanctions et la coopération décline au cours du temps en raison de la présence de sanctions antisociales. Jusqu'ici le

punisseur ne peut pas à son tour réagir à une contre-punition. En augmentant de manière endogène le nombre d'étapes de jeu au cours desquelles les sanctions et contre-sanctions peuvent s'enchaîner librement, il est alors possible d'observer l'apparition de vendettas en laboratoire (Nikiforakis et Engelmann [2012]). Toutefois, ce risque est anticipé par les joueurs qui réduisent leur recours initial à la sanction mais réduisent également leur contribution au bien public. Afin d'isoler la recherche d'une vengeance personnelle de celle d'une contre-punition destinée à faire respecter une norme, Denant-Boemont, Masclet et Noussair [2007] comparent trois traitements : un traitement dans lequel chacun a une information complète sur les comportements de contribution et de sanction de chaque autre joueur vis-à-vis de chacun des membres du groupe, un traitement dans lequel chacun n'est informé que des comportements de contribution et de sanction de chacun des autres joueurs vis-à-vis de lui-même (comme chez Nikiforakis [2008]), et un traitement dans lequel chacun n'est informé que des comportements de contribution et de sanction de chaque autre joueur vis-à-vis des membres du groupe autres que lui-même (comme chez Cinyabuguma, Page et Putterman [2005]). L'efficience la plus élevée est observée dans le dernier traitement (où la contre-punition sert à faire respecter une norme de contribution) et la plus faible dans le second traitement (où la contre-punition sert d'abord à assouvir une vengeance personnelle).

Si dans le court terme les sanctions réduisent l'efficience et le bien-être, en revanche si l'on étend l'horizon du jeu, les gains d'efficience sont accrus significativement. Dans Gächter, Renner et Sefton [2008], le jeu standard de bien public avec sanctions est répété cinquante fois au lieu de dix ou vingt fois habituellement. Une réduction d'efficience est observée lors des cinq premières périodes de jeu car les sujets punissent pour instaurer la coopération dans leur groupe. Mais une fois que le niveau de contribution atteint l'optimum social, il reste stable à ce niveau jusqu'à la fin du jeu. Dès lors, les bénéfices sur le long terme de ces sanctions compensent très largement la perte d'efficience initiale et suggèrent que les sanctions de début du jeu peuvent être analysées comme un investissement.

Enfin, il existe plusieurs manières de réduire les effets négatifs de la mise en œuvre de sanctions dans le court terme : l'introduction d'un choix collectif des institutions au sein des groupes, la mise en œuvre publique des sanctions, la conception de dispositifs de sanction de fréquence aléatoire. Plusieurs articles ont ainsi introduit un mécanisme de vote par lequel les individus peuvent choisir leur environnement. Selon les cas, les individus peuvent choisir une seule fois entre un environnement sans institution ou un environnement dans lequel les membres du groupe peuvent se sanctionner (Botelho et al. [2005], Sutter, Haigner et Kocher [2010]), ou bien ils peuvent faire ce choix de manière répétée (Gürerk, Irlenbusch, Rockenbach [2008]) : si le vote unique favorise rarement le mécanisme de sanction informelle ou décentralisée, la répétition du vote montre que les individus souhaitent recourir de plus en plus à ce mécanisme au cours du temps. D'autres auteurs s'intéressent aux sanctions formelles ou centralisées qui fonctionnent comme une loi. Ils comparent les niveaux d'efficience atteints lorsqu'un tel mécanisme de sanction formelle est imposé de manière exogène et lorsqu'il résulte d'un vote des joueurs : il apparaît qu'une majorité de groupes votent pour ce mécanisme de sanction et que l'efficience est accrue à la fois par rapport à un environnement sans sanction et par rapport à la mise en œuvre du même type de sanction de manière exogène (Tyran et Feld [2006]). Cet effet positif du vote sur l'efficience des sanctions a été également identifié dans un jeu où les individus peuvent choisir entre l'instauration d'un mécanisme de sanction centralisé

et un mécanisme de sanction décentralisé (Kamei, Putterman et Tyran [2011]). L'efficience est accrue dans le cas du choix endogène des sanctions décentralisées par rapport aux autres situations, sans doute parce que le choix collectif de cette institution permet de signaler aux autres membres du groupe l'intention de punir les passagers clandestins.

De leur côté, Houser et Xiao [2011] ont montré qu'une mise en œuvre de la punition en public de manière anonyme (afin de ne pas créer de sentiment de honte chez les joueurs) plutôt qu'en privé (où seul l'individu sanctionné est informé de l'existence et de l'importance de la sanction) annule ces effets négatifs. En effet, la mise en œuvre publique d'une sanction renforce la saillance de la norme de coopération pour la production du bien public et focalise l'attention des joueurs sur cette norme. Enfin, dans un environnement caractérisé par un mécanisme de sanction exogène comme chez Houser et Xiao [2011], Dai, Hogarth et Villeval [2012] manipulent la fréquence et la régularité de l'audit des contributions. Ils établissent qu'après le retrait définitif des audits, les joueurs continuent à coopérer à un niveau très significativement supérieur lorsque le régime de contrôle précédent s'appuyait sur des audits aléatoires et irréguliers plutôt que sur des audits systématiques et continus. Ce résultat permet de montrer que l'on peut définir un régime de punition minimal suffisant pour garantir un respect de la norme par les membres du groupe tout en préservant l'efficience.

Le bâton ou la carotte ?

Compte tenu des effets négatifs potentiels des sanctions en termes de bien-être, plusieurs expériences ont été conçues afin d'étudier si l'usage d'incitations positives serait plus efficient pour la production de biens publics que les sanctions. Dans Sefton, Shupp et Walker [2007], les joueurs participent d'abord à dix périodes d'un jeu standard de bien public ; puis, dans une nouvelle séquence de dix périodes, selon les traitements est introduite soit une opportunité de sanctionner les autres membres de son groupe, soit une opportunité de les récompenser, soit encore la possibilité de sanctionner et de récompenser. Récompenser un autre joueur engendre le même coût que sanctionner un autre joueur pour celui qui s'y emploie mais cela augmente le gain de celui qui les reçoit d'un montant équivalent au lieu de le réduire. Les résultats de cette expérience montrent que le droit de récompense seul ne parvient pas à élever le niveau moyen d'efficience par rapport au jeu de contribution simple. L'effet des sanctions et des récompenses n'est donc pas symétrique. En revanche, le traitement mixte conduit à une hausse à la fois du niveau de coopération et de l'efficience. Les joueurs qui punissent et ceux qui récompensent sont très généralement ceux qui contribuent eux-mêmes davantage que la moyenne de leur groupe. En dynamique, le traitement mixte révèle que les individus choisissent plus fréquemment de récompenser que de sanctionner au début du jeu ; mais progressivement l'usage des récompenses diminue tandis que le recours aux sanctions augmente. Le maintien d'un régime de récompenses s'avère difficile. Globalement les deux mécanismes s'avèrent donc complémentaires et entrent en synergie.

Cette préférence initiale mais non durable pour un environnement sans sanction a été également observée par Gürerk, Irlenbusch et Rockenbach [2006], Rockenbach et Milinsky [2006] et Ertan, Page et Putterman [2008]. Ainsi chez Gürerk, Irlenbusch et Rockenbach

[2006], lors de la première étape de chaque période de jeu, les participants choisissent entre un environnement sans sanction et un environnement dans lequel les joueurs peuvent attribuer des points de sanction et de récompense. Lors de la deuxième étape, les participants contribuent au bien public. Enfin, les participants qui ont choisi un environnement avec sanctions et récompenses continuent lors d'une troisième étape durant laquelle ils peuvent allouer des points aux autres membres du groupe. En début de jeu, 63% des joueurs choisissent l'environnement sans sanction ni récompenses ; il s'agit des joueurs qui en moyenne contribuent significativement moins que les autres membres du groupe. Après 15 périodes, plus de 80% des joueurs choisissent l'autre environnement et ce pourcentage augmente très vite ensuite à plus de 90%. Cette dynamique s'explique par le fait que dans l'environnement sans sanction le niveau moyen de contribution décline très rapidement vers l'équilibre de non-contribution ; au contraire, dans l'environnement avec sanction et récompenses, les contributions convergent rapidement vers l'optimum.

La communication et le *leadership*, alternatives ou compléments aux sanctions

Dans la plupart des situations réelles, les membres d'un groupe ont la possibilité d'échanger et de communiquer avant de passer à l'action sous forme de sanctions à l'encontre d'autres membres de leur groupe. A partir d'un jeu de ressources communes dans lequel les individus ont à décider d'un montant de prélèvement dans le bien public, Ostrom, Walker et Gardner [1992] ont montré que la combinaison entre une communication libre (« non structurée ») entre les joueurs avant la prise de décision et un mécanisme de sanctions volontaires était l'institution la plus à même de soutenir la coopération. Cette conclusion sur le rôle favorable de la communication entre les membres du groupe rejoint celle d'autres études dont Dawes, McTavish et Shaklee [1977], Isaac, McCue et Plott [1985], Isaac et Walker [1988] [1991], Kerr et Kaufman-Gilliland [1994], Sally [1995], Krishnamurthy [2001], Brosig, Ockenfels et Weimann [2003].

Plus récemment, Bochet, Page et Putterman [2006] ainsi que Bochet et Putterman [2009] ont revisité le rôle de diverses formes de communication sur les décisions de contribution des joueurs au bien public. Bochet, Page et Putterman [2006] comparent dans trois traitements différents trois formes de communication autorisées entre les membres de chaque groupe avant la décision individuelle de contribution : une communication en face-à-face pendant cinq minutes, une communication en ligne non structurée et préservant l'anonymat des joueurs et une communication en ligne structurée où les joueurs ne peuvent indiquer que des niveaux de contribution possibles mais non contraignants. Dans trois nouveaux traitements reprenant ces trois modes de communication, le jeu s'enrichit de la possibilité de sanctionner les membres du groupe en dernière étape. D'un point de vue théorique, l'introduction de ces modes de communication et de sanction ne modifie pas l'équilibre du jeu car il s'agit toujours de messages et de menaces non crédibles. Pourtant, les résultats de l'expérience montrent l'importance de la communication. Le mode de communication le plus efficace en termes d'impact sur la hausse des niveaux de contribution au bien public est la communication en face-à-face : plus de 78% des joueurs contribuent la totalité de leur dotation jusqu'à la dernière période de jeu incluse. La communication en ligne s'avère également efficace pour créer une confiance entre les joueurs, bien qu'à un niveau inférieur par rapport à la communication en face-à-face dans la mesure où les sujets ne peuvent pas voir les expressions visuelles ou les nuances

vocales. En revanche, la communication structurée numérique ne modifie pas les niveaux moyens de contribution par rapport à un traitement de contrôle sans communication. Contrairement à Ostrom, Walker et Gardner [1992], l'ajout de l'opportunité de sanctionner ne modifie pas significativement la coopération dès lors que les modes de communication s'avèrent déjà capables d'élever les niveaux de contribution. Ainsi c'est bien le contenu de la communication qui explique son effet. L'analyse du contenu des messages verbaux montre que la communication la plus efficace n'est pas celle qui utilise la menace de représailles mais bien celle qui souligne les avantages de l'accord et de la coopération.

Bochet et Putterman [2009] montrent cependant que la communication structurée numérique fonctionne dans certains groupes et qu'ajouter aux messages numériques des promesses améliore significativement la coopération dès lors que les participants disposent également d'un droit de sanction pour lutter contre les promesses non tenues. De ce point de vue, ces résultats sont complémentaires de ceux de Masclet, Noussair et Villeval [2012] cités supra. En effet, dans cette expérience, la possibilité d'adresser préalablement des menaces numériques de sanction pour chaque niveau possible de contribution assortie d'une droit de sanction après la phase de contribution augmente bien le niveau de coopération dans les groupes mais pas l'efficience car les individus se sentent relativement contraints de mettre leur menace à exécution. Les résultats de Bochet et Putterman [2009] entrent également en résonance avec ceux de Kroll, Cherry et Shogren [2007] qui ont observé que l'introduction d'un vote sur un niveau de contribution pour le groupe n'a que peu d'effet en soi sur la coopération, sauf si ce vote et assorti d'un droit de sanction.

Jusqu'à présent, ont été évoqués des travaux dans lesquels la communication est horizontale entre des membres du groupe qui prennent leur décision de contribution simultanément. Mais dans de nombreuses situations réelles, il existe au moins deux autres types de communication. D'une part, la communication peut prendre la forme de recommandations et conseils de la part d'autres individus confrontés préalablement au même type de situation. D'autre part, la communication peut passer par l'observabilité des choix réalisés par d'autres membres du groupe avant de prendre soi-même sa décision de contribution. Les jeux inter-générationnels illustrent la première forme de communication en dehors de tout recours aux sanctions. Ainsi, Chaudhuri, Graziano et Maitra [2005] ont conçu une expérience dans laquelle les joueurs peuvent transmettre à la fin d'une session des conseils sous forme de messages non structurés destinés à d'autres joueurs participant à une session ultérieure. Selon les traitements, le message de chacun est destiné seulement à un autre joueur de la future session (information privée), à tous les autres joueurs (information publique) ou encore à tous les autres joueurs en plus d'une lecture à voix haute par l'expérimentaliste (connaissance commune). Le traitement avec connaissance commune est celui qui élève le plus le niveau moyen de contribution car l'envoi de messages d'encouragement homogénéise les croyances sur les comportements des autres joueurs et facilite la coordination. Chaudhuri, Maitra et Skeath [2009] montrent en outre que l'échange de conseils intergénérationnels a plus d'impact sur les décisions de contribution que la communication entre pairs.

L'autre forme de transmission d'informations passe par l'introduction d'une séquentialité dans les décisions de contribution. Ainsi, même lorsque les groupes ne sont pas organisés hiérarchiquement, des leaders émergent, naturellement ou non, et peuvent influencer les autres membres de leur groupe à travers leurs choix. Ces leaders peuvent ou non être mieux informés que les autres membres de leur groupe sur l'état de la nature (le

rendement du bien public par exemple). L'article théorique de Hermalin [1998] sur le leadership par l'exemple ou par le sacrifice en présence d'information asymétrique a ainsi influencé plusieurs expériences (Meidinger et Villeval [2003], Potters, Sefton et Vesterlund [2005], Potters, Sefton et Vesterlund [2007]). Dans ces expériences, le leader reçoit une information privée et prend sa décision avant les autres joueurs ; ces derniers observent le choix du leader et prennent ensuite leur propre décision. Le leader peut ainsi influencer les choix des joueurs suivants de manière stratégique. Meidinger et Villeval [2003] montrent toutefois que le leadership fonctionne plus par réciprocité que par effet de signal.

Les expériences de biens publics avec leadership en information symétrique ne font qu'introduire une séquentialité des décisions : un joueur – le *leader* - choisit en premier ; son choix est observé par les autres joueurs – les *followers* - qui prennent ensuite leur décision simultanément. Ces expériences confirment que les followers élèvent leur niveau de contribution en fonction de la contribution initiale du leader ; le leader anticipant cet effet d'entraînement tend à élever sa propre contribution. Ceci est un résultat important pour l'analyse des préférences sociales car un leader ne devrait théoriquement pas contribuer davantage qu'un autre joueur – sauf si à la suite de son action et au lieu d'adopter un comportement de passager clandestin, les followers contribuent au total davantage que l'investissement du leader dans le bien public. Or, si le niveau moyen de contribution dans un jeu séquentiel est significativement plus élevé par rapport au jeu simultané, le gain moyen des leaders est généralement inférieur à celui des followers. Ces résultats sont observés quand les leaders sont choisis de manière aléatoire (par exemple Moxnes et van der Heijden [2003] dans un jeu de mal public), en fonction de leur contribution dans un jeu simultané joué préalablement (Gächter et Renner [2005]), ou en fonction du statut attribué de manière exogène à partir de la performance relative des joueurs dans un quizz joué préalablement au jeu de bien public (Kumru et Versterlund [2010]). Gächter et al. [2012] montrent que les groupes qui atteignent les niveaux de fourniture du bien public les plus élevés sont ceux dirigés par des leaders caractérisés à la fois par des préférences sociales coopératives et par des croyances très optimistes sur la volonté des autres de suivre leur exemple ; les leaders les plus coopératifs sont aussi les plus optimistes sur le degré de coopération des autres.

Dans les expériences précédentes, les membres du groupe ne choisissent pas leur leader. Or dans de nombreux cas, le leadership est tournant et les pairs choisissent leur leader. L'introduction d'une procédure de vote pour sélectionner le leader confirme les résultats précédents sans toutefois élever le niveau d'efficience (Güth et al. [2007], Levy et al. [2011], Levati, Sutter et van der Heijden [2007]). De plus, dans les expériences précédentes, le leader n'a pas d'autorité formelle. Or, l'octroi au leader de moyens d'action autres que son seul exemple permet d'accroître son impact. Chez Güth et al. [2007] et Levati, Sutter et van der Heijden [2007], les leaders disposent d'un droit d'exclusion des membres de leur groupe à la période suivante, ce qui engendre un coût pour le groupe en réduisant le nombre de contributeurs potentiels mais augmente les contributions moyennes. Chez Gürerk, Irlenbusch et Rockenbach [2009], les leaders peuvent utiliser des incitations pour encourager les membres du groupe à contribuer. Enfin, chez Potters, Sefton et van der Heijden [2009], les leaders décident de la règle de partage du bien public entre les membres du groupe.

Très peu d'expériences cependant ont introduit un leadership volontaire (Arbak et Villeval [2012], Nosanzo et Sefton [2011], Rivas et Sutter [2011]). Or, l'analyse de la décision de se porter volontaire permet de mieux appréhender les motivations et les

préférences sociales des joueurs dans le jeu de bien public. Ainsi, Arbak et Villeval [2012] ajoutent au jeu standard une étape préalable. Lors de cette étape, les membres du groupe peuvent se porter candidats pour prendre leur décision en premier de façon à signaler aux autres joueurs leur niveau de contribution avant qu'ils ne prennent leur propre décision. En cas de multiples candidatures, un leader unique est sélectionné aléatoirement parmi les volontaires. Les candidats non retenus peuvent réviser leur contribution après avoir observé celle du candidat sélectionné. Les résultats montrent que malgré le coût associé au leadership, des candidats se portent volontaires, y compris en fin de jeu. Les auteurs identifient trois motifs à la décision de se porter volontaire. Certains joueurs espèrent tirer un bénéfice personnel du leadership en escomptant un taux de corrélation élevé entre leur contribution et celle des autres membres du groupe ; ces leaders cessent de se porter candidats dès lors que le taux de corrélation est insuffisant. D'autres joueurs sont guidés par l'altruisme et acceptent de supporter le coût du leadership de manière répétée parce que le niveau de fourniture du bien public est largement supérieur dans un groupe avec leader que dans un groupe sans leader. Enfin, d'autres volontaires cherchent à maintenir une bonne image de soi en guidant le groupe. En cas de non-sélection face à une pluralité de candidatures, ils ne suivent pas le leader sélectionné et ils réduisent leur contribution ; ce comportement est observé significativement plus fréquemment chez les hommes que chez les femmes. Toutefois, en raison des phénomènes d'auto-sélection, les leaders volontaires ne se révèlent pas plus efficaces que ceux choisis au hasard (ce résultat diffère de celui de Rivas et Sutter [2011] qui autorisent une multiplicité de leaders dans un groupe).

Pour leur part, Nosanzo et Sefton [2011] testent le modèle de bien public avec rendement quasi-linéaire de Varian [1994] qui, sous certaines hypothèses, prédit que le premier joueur peut bénéficier d'un avantage s'il s'engage à ne pas contribuer, laissant ainsi aux joueurs suivants la charge de fournir le bien public. Varian montre ainsi que si le bien public a une grande valeur pour le premier joueur et une valeur inférieure pour les autres joueurs, l'engagement du premier joueur à ne pas contribuer conduit à un niveau de fourniture du bien public encore plus faible que dans un jeu simultané. Le test expérimental de Varian [1994] à l'aide de divers dispositifs d'engagement montre toutefois qu'il existe une forte tendance chez les joueurs à éviter de s'engager dans une contribution précoce. De fait, la séquence négative prédite par Varian apparaît dans moins de vingt pour cent des cas. Ceci est expliqué, mais en partie seulement, par la présence d'aversion à l'inégalité chez les joueurs qui conduit les joueurs suivants à punir les premiers joueurs qui ne contribueraient pas assez à leurs yeux.

Au total, lorsque le jeu de bien public est joué de manière standard, les préférences sociales expliquent les déviations par rapport à l'équilibre de Nash du jeu ainsi que le processus de déclin des contributions au fil du temps. Dès lors que l'on introduit des sanctions, des récompenses, de la communication ou de la séquentialité, les contributions se rapprochent de l'optimum du jeu alors que les prédictions théoriques sous les hypothèses standard demeurent les mêmes que dans le jeu de base. Il convient donc de s'interroger sur le modèle théorique le plus à même d'expliquer ces déviations.

QUEL MODELE THEORIQUE POUR RENDRE COMPTE DU ROLE DES PREFERENCES SOCIALES ?

La théorie standard basée sur des préférences purement égoïstes et sur une connaissance commune de l'égoïsme des autres joueurs est peu performante pour rendre compte des comportements de contribution et de sanction dans le jeu de bien public linéaire, dans la mesure où les erreurs ne peuvent pas expliquer les déviations de l'équilibre. A l'inverse, la théorie de l'altruisme et celle du *warm-glow* (Andreoni [1990]) peuvent expliquer la présence de contributions positives. En revanche, elles ne fournissent pas d'explication au déclin progressif des contributions individuelles au cours du temps ni à la corrélation positive entre les contributions individuelles et la moyenne des contributions du groupe (Croson [2007]) ; de plus, elles ne permettent pas de rendre compte de la décision de sanctionner d'autres membres du groupe. D'autres théories s'avèrent nécessaires.

Théories de l'aversion à l'inégalité

La théorie des préférences sociales la plus souvent mobilisée, de par sa simplicité, pour rendre compte de ces comportements non stratégiques et de leur évolution est la théorie de l'aversion à l'inégalité. Cette théorie met l'accent sur les préférences distributives des agents. La version la plus commune est le modèle de Fehr et Schmidt [1999], mais il convient de garder à l'esprit également le modèle de Bolton et Ockenfels [2000]. Les différences principales entre les deux modèles est que chez Fehr et Schmidt [1999], les individus comparent leur gain à celui de chacun des autres membres du groupe alors que chez Bolton et Ockenfels [2000], la référence est la moyenne des gains des autres joueurs et ce dernier modèle autorise des non-linéarités dans la fonction d'utilité.

Fehr et Schmidt [1999] réécrivent la fonction d'utilité des individus de la manière suivante :

$$U\left(x_i\right) = \pi_i - \alpha_i \frac{1}{n-1}\sum_{j\neq i}\max\left(\pi_j - \pi_i, 0\right) - \beta_i \frac{1}{n-1}\sum_{j\neq i}\max\left(\pi_i - \pi_j, 0\right)$$

avec π_i le gain du joueur i, π_j le gain du joueur j, α_i le paramètre d'aversion à l'inégalité désavantageuse et β_i le paramètre d'aversion à l'inégalité avantageuse. Dans cette équation, on impose que $\alpha_i > \beta_i$ et $0 \leq \beta_i < 1$. En d'autres termes, en comparant ses gains à celui des autres joueurs, l'individu souffre à la fois des inégalités en sa faveur (culpabilité) et des inégalités en sa défaveur (envie) mais il souffre davantage des inégalités jouant en sa défaveur. Par ailleurs, on suppose que l'individu qui gagne plus qu'un autre joueur ne souffre pas de cette inégalité au point de vouloir brûler son propre argent pour restaurer une situation d'égalité de gains.

Le modèle de Fehr et Schmidt [1999] prédit qu'une petite minorité de joueurs égoïstes est suffisante pour conduite des individus averses à l'inégalité à réduire leur contribution car au cours des répétitions les individus découvrent l'égoïsme d'autres joueurs. Il prédit également que rationnellement des joueurs peuvent punir d'autres individus pour réduire l'inégalité de gain résultant de l'inégalité des contributions. En dérivant les équilibres pour le jeu avec et sans sanctions en présence de considérations distributives, Fehr et Schmidt montrent qu'en présence de sanctions, des équilibres coopératifs sont possibles car les joueurs égoïstes vont se mettre à contribuer sous la pression des joueurs suffisamment averses à l'inégalité pour appliquer des sanctions. Sous certaines conditions, le modèle peut même prévoir l'existence de sanctions antisociales à l'encontre de coopérateurs qui ne

puniraient pas les passagers clandestins (une telle extension du modèle initial est proposée par Thöni [2011]).

Pourtant les considérations distributives ne peuvent pas *seules* expliquer les comportements de contribution et leur déclin au cours du temps, pas plus que les comportements de sanction. Ceci est particulièrement évident si l'on pense aux sanctions antisociales puisque les individus qui punissent ainsi sont ceux qui ont le plus gagné en première étape du jeu et qui, en punissant les coopérateurs, accentuent encore les écarts de gains avec ces derniers (pour un test expérimental rejetant la théorie, voir Thöni [2011]). De même, les modèles d'aversion à l'inégalité ne peuvent pas prédire les phénomènes de contre punition. Pour tester la motivation de réduction des inégalités dans les comportements de sanction, Masclet et Villeval [2008] ont conçu une expérience comparant deux traitements principaux. Le traitement dit de coût inégal réplique le protocole initial de Fehr et Gächter [2000] avec réappariement des joueurs à la fin de chaque période de jeu ; le traitement dit de coût égal est similaire, à ceci près que le coût de la sanction est similaire pour celui qui punit et pour celui qui est puni. Dans ce second traitement, la sanction ne peut en aucun cas affecter l'inégalité de gain entre les joueurs. Si les comportements de sanction répondent fondamentalement à la volonté de réduire les inégalités de gain, alors les joueurs ne devraient jamais dépenser de ressources pour punir dans le traitement de coût égal. Bien au contraire, les résultats montrent que les individus sanctionnent les joueurs qui contribuent moins que la moyenne même lorsque la sanction n'affecte pas les différences de gains. Ce résultat ne signifie pas pour autant que les joueurs sont indifférents aux comparaisons de gains. En effet, l'intensité de la sanction est toujours corrélée avec l'intensité des écarts de gains entre les individus. De plus, avec le recours aux sanctions, le niveau d'inégalité entre les gains des joueurs tend à s'atténuer au cours du temps. Mais alors que dans un environnement sans sanction les comportements de passager clandestin et la coopération conditionnelle conduisent au fil du temps à une égalisation des gains à un niveau minimum, l'existence de sanctions conduit à une égalisation des gains à un niveau de richesse significativement supérieur. Ces résultats corroborent ceux de Falk, Fehr et Fischbacher [2005] qui montrent à partir de jeux de dilemme du prisonnier que la punition n'est pas motivée par des considérations distributives mais plutôt par la volonté de se venger et de blesser ceux qui ont violé la norme. Sur un plan purement théorique, Cox et Sadiraj [2005] montrent que le modèle de Fehr et Schmidt ne peut prédire que des contributions symétriques, ce qui est manifestement rejeté par les preuves expérimentales.

Emotions et théories de la réciprocité

Ces résultats tendent à montrer que d'autres sources de motivation que les préférences égalitaires sont nécessaires pour expliquer ces comportements non stratégiques. Les comportements de sanction répondent ainsi davantage à de la réciprocité négative qu'à des considérations distributives ou à de l'altruisme. Il a été ainsi montré que les émotions, comme la colère et l'indignation, la honte et la culpabilité, jouent un rôle important dans les décisions (voir de Quervain et al. [2004], Joffily et al. [2011]). Les économistes ont toutefois encore très peu analysé le rôle des émotions dans la prise de décisions et il y a fort à parier que cette réflexion va se développer dans les années à venir pour mieux rendre compte de la manière dont émotions et rationalité se combinent dans le raisonnement (Damasio [1994]).

Un support théorique à ces motivations peut-il être trouvé dans les théories des intentions et de la réciprocité (Sugden [1984], Rabin [2003], Falk et Fischbacher [2006]) ? La théorie de la réciprocité de Sugden [1984] prédit que les individus réciproques se déterminent en fonction du minimum anticipé contribué par les autres membres du groupe. Mais cette théorie ne s'applique qu'au jeu simultané (reposant donc sur les croyances des joueurs) et de plus, Croson [2007] a montré que les individus sont influencés par la moyenne ou la médiane des autres contributions et non pas par la contribution minimum. La théorie de Rabin [2003], fondée sur les jeux psychologiques et les croyances de premier ordre et d'ordres supérieurs, est basée sur la définition d'équilibres de bienveillance. Elle suggère que les individus ne sont pas intrinsèquement altruistes mais déterminent leur contribution en fonction de leurs croyances sur le comportement de contribution des autres, tentant de régler ainsi un problème de coordination. Toutefois, cette théorie n'est pas totalement adaptée pour rendre compte de jeux séquentiels ni des comportements de punition avec $N>2$. Par ailleurs, Fischbacher et Gächter [2010] ont montré que ce sont plus les préférences sociales que les croyances qui déterminent les choix dans ce jeu. Mêlant intentions et considérations distributives dans un même cadre théorique, Falk et Fischbacher [2006] ont étendu la notion de coopération conditionnelle à des jeux séquentiels répétés finis, développant la notion de réciprocité séquentielle et généralisant ainsi Rabin [2003]. Quant à elle, la théorie de Charness et Rabin [2002] avec des préférences quasi-maximin met également l'accent à la fois sur les intentions et la répartition en insistant sur les préférences des individus pour l'efficience. Dès lors, si elle peut permettre d'expliquer la coopération conditionnelle, elle peut difficilement rendre compte des comportements de sanction qui, à court terme du moins, détruisent des ressources et affectent négativement l'efficience. Cox et Sadiraj [2005] concluent plus radicalement que les complications du modèle de quasi-maximin ainsi proposé n'augmentent pas le pouvoir explicatif des comportements dans le jeu de bien public, sachant par ailleurs que les contributions individuelles sont influencées par la moyenne et non par la contribution maximum des autres joueurs.

Il n'en reste pas moins que la réciprocité mise en avant dans ces théories des intentions (qui ne sont pas nécessairement contradictoires avec la théorie de l'aversion à l'inégalité) capte certainement une dimension importante des comportements dans les jeux de bien public. L'analyse évolutionnaire de Bowles et Gintis [2000] montre comment la « réciprocité forte » (définie comme le fait altruiste d'obéir à la norme et de punir les violateurs d'une norme sociale au bénéfice du groupe, moyennant un coût personnel) peut parvenir à envahir une population d'individus non-réciproques. Elle peut alors être soutenue comme un équilibre stable dans la population. Bowles et Gintis [2000] suggèrent ainsi que l'adhésion aux normes et la punition altruiste des violateurs peut contribuer à expliquer l'évolution des sociétés humaines. Ils n'excluent pas le principe d'un héritage génétique de la prédisposition à la réciprocité forte comme résultant d'un processus de sélection naturelle. Ce faisant, la réciprocité forte ne serait plus alors une anomalie mais un trait caractéristique de l'être humain. Dans Bowles et Gintis [2011], les auteurs expliquent que l'essentiel n'est pas tant d'expliquer pourquoi des individus égoïstes se mettent à coopérer mais plutôt d'identifier les processus génétiques et sociaux qui depuis des milliers de générations d'humains poussent des étrangers à internaliser des normes sociales et à punir les violateurs de ces normes.

CONCLUSION

Cet article synthétise des travaux récents en économie comportementale portant sur les comportements de contribution volontaire à la fourniture des biens publics. Les principaux résultats de ces travaux montrent que les individus coopèrent bien davantage que ne le prédit le modèle théorique standard fondé sur l'égoïsme des individus. Face à un tel dilemme social, les préférences sont hétérogènes. A côté de passagers clandestins qui se conforment au modèle de l'homo oeconomicus, on dénote la présence d'une majorité de coopérateurs conditionnels et d'une faible minorité d'individus altruistes. Les coopérateurs conditionnels sont des individus réciproques qui respectent et font respecter une norme sociale. La norme sociale est définie par la contribution moyenne du groupe et elle est réajustée régulièrement. Le phénomène d'érosion de la coopération au cours du temps qui a intrigué nombre d'économistes durant les dernières décennies a ainsi trouvé son explication dans la conditionnalité de la coopération, en l'absence d'institutions. C'est la présence de coopérateurs conditionnels, frappés d'un biais d'auto-complaisance, qui explique la tendance régulière à la baisse des contributions jusqu'à se rapprocher de l'équilibre de Nash du jeu.

Un des autres apports majeurs de la recherche comportementale récente a été l'identification de mécanismes institutionnels permettant de soutenir la coopération en présence de dilemmes sociaux. En particulier, l'opportunité pour les individus de se sanctionner de manière endogène permet aux groupes d'atteindre rapidement l'optimum social de contribution complète. Alors qu'en l'absence d'institutions, la présence d'une minorité d'individus égoïstes suffit à convaincre la majorité des autres joueurs pourtant dotés de préférences sociales à adopter un comportement de passager clandestin, le droit de sanctionner conduit en revanche la minorité d'individus égoïstes à se comporter comme la majorité en adoptant une attitude éthique et coopérative jusqu'à la fin du jeu. Cela suppose toutefois qu'il y ait un accord sur la norme ; cet accord est plus aisé à produire quand les individus ont la possibilité de communiquer, quand il y a égalité de dotations initiales entre les membres du groupe ou quand les groupes sont homogènes en termes de préférences sociales. La pression par les pairs et la discipline de groupe mutuelle ne sont pour autant pas une solution universelle. Elles peuvent générer à court terme une perte d'efficience et un risque d'escalade dans les sanctions ; elles ne conduisent pas à l'optimum dans tous les groupes ni dans tous les environnements culturels en raison notamment de comportements de sanction pervers ou antisociaux.

Ces résultats montrent l'importance de prendre en compte dans l'analyse l'hétérogénéité des individus, des groupes et des cultures pour rendre compte de la dynamique des normes de coopération. Au delà, ils appellent à un effort de théorisation de cette hétérogénéité. Ils révèlent aussi la difficulté actuelle des théories des préférences sociales à prédire pleinement ces comportements de contribution et de sanction. L'aversion à l'inégalité rationalise une partie de ces comportements mais les résultats expérimentaux mettent en avant principalement le rôle de la réciprocité et de la réciprocité forte plus encore que des croyances. Pourtant, si l'on dispose de plusieurs théories de la réciprocité, aucune ne rend compte simplement et complètement des phénomènes de contribution et de sanction dans les jeux de bien public. Enfin, les émotions influencent considérablement les choix des individus face à un dilemme social. Là aussi, la modélisation théorique des émotions pour comprendre les comportements économiques sera un enjeu majeur pour la recherche en économie comportementale dans les années à venir.

Ces résultats issus de l'économie comportementale éclairent certains processus rendant possible la vie en société et son évolution, en explicitant notamment la construction et la déconstruction de la coopération et de la confiance entre individus non génétiquement reliés. Ils expliquent pourquoi certains groupes ou certaines sociétés se trouvent enfermés durablement dans un équilibre de non coopération alors que d'autres s'appuient sur des valeurs de réciprocité pour construire et soutenir la coopération productrice de richesses. Les préférences sociales sont donc à analyser comme une ressource à laquelle il convient d'adjoindre des politiques d'incitation appropriées pour les soutenir. La question de la complémentarité ou de la substituabilité des préférences sociales et des incitations n'était pas l'objet de cet article mais elle est fondamentale d'un point de vue d'économie publique (voir à cet égard Bowles et Hwang [2008]). Ces résultats sont majeurs si l'on pense que la coopération est à la base de l'évolution et du progrès des sociétés tout autant que la compétition et que les sociétés les plus évoluées sont celles dont les membres sont parvenus à coopérer durablement (sur ce point, voir notamment Bowles et Gintis [2011]). A ce titre, ces recherches enrichissent et s'enrichissent des apports des autres sciences (psychologie, sociologie, anthropologie, neurologie, biologie).

Enfin, nombre de ces résultats ont été produits grâce à des expériences de laboratoire. Les conditions de validité externe de ces expériences doivent être gardées à l'esprit. Mais il est de plus en plus difficile aujourd'hui de douter de la portée scientifique de cette méthode. Un témoignage intéressant sur l'évolution du regard de la profession est délivré par Falk et Heckman [2009]. Dans la préface de 12ème édition de leurs « *Principles of Microeconomics* », Samuelson et Nordhaus [1985] écrivaient : « (Malheureusement) les économistes n'ont pas la possibilité de réaliser des expériences contrôlées comme les chimistes ou les biologistes parce qu'ils ne peuvent pas contrôler facilement les facteurs importants. Tout comme les astronomes et les météorologistes, ils doivent se contenter généralement d'observer ». Les mêmes auteurs écrivent, dans la 14ème édition des mêmes *Principles* : « L'économie expérimentale est un nouveau développement excitant » (Samuelson et Nordhaus [1985]). C'est à n'en pas douter !

RÉFÉRENCES BIBLIOGRAPHIQUES

ALBEROLA E., CHEVALLIER J., CHEZE B. [2008a], « Price Drivers and Structural Breaks in European Carbon Prices 2005-2007 », *Energy Policy,* 36 (2), p.787-797.

ANDERSON S.P., GOERE J.K., HOLT C.A. [1998], « A Theoretical Analysis of Altruism and Decision Error in Public Goods Games », *Journal of Public Economics,* 70, p.297-323.

ANDERSON C., PUTTERMAN L. [2006], « Do non-strategic sanctions obey the law of demand ? The demand for punishment in the voluntary contribution mechanism », *Games and Economic Behavior,* 51 (1), p.1-24.

ANDREONI J. [1988]). « Why Free Ride? Strategies and Learning in Public Goods

Experiments », *Journal of Public Economics,* 37, p.291-304.

ANDREONI J. [1990], « Impure altruism and donations to public goods: a theory of warm-glow giving », *The Economic Journal,* 100, p.464–477.

ANDREONI J. [1995], « Cooperation in public goods experiments: Kindness or confusion ? *American Economic Review,* 85 (4), p.891-904.

ANDREONI J., CROSON, R. [2008], « Partners versus strangers: the effect of random rematching in public goods experiments", in Smith V. et Plott C. (Eds.), *Handbook of Experimental Economics Results*, New York: Elsevier.

ARBAK E., VILLEVAL M.C. [2012]. Endogenous leadership: motivation and influence. A paraître in *Social Choice and Welfare*.

ARON S., PASSERA, L. [2008], *Les sociétés animales : évolution de la coopération et organisation sociale*, Bruxelles, de Boeck Université, 336p.

BALAFOUTAS L., NIKIFORAKIS N. [2011], « Altruistic punishment in the city: A natural field experiment », University of Innsbruck, mimeo.

BARDSLEY N., MOFFATT P. [2007], « The experimetrics of public goods: inferring motivations from contributions », *Theory and Decision,* 62, p.161-193.

BINMORE K.G. [2005], « Economic man or straw man? », *Behavioral and Brain Sciences,* 28, p.817-818.

BOCHET O., PAGE T., PUTTERMAN L. [2006], « Communication and Punishment in Voluntary Contribution Experiments », *Journal of Economic Behavior & Organization,* 60 (1), p.11-26.

BOCHET O., PUTTERMAN L. [2009], « Not just babble: Opening the black box of communication in a voluntary contribution experiment », *European Economic Review,* 53 (3), p.309-326.

BOLTON G.E., OCKENFELS A. [2000], « ERC: a theory of equity, reciprocity, and competition », *American Economic Review,* 90 (1), p.166–193.

BOTELHO A., HARRISON G., COSTA PINTO L.M., RUTSTRÖM E.E. [2005], « Social Norms and Social Choice », University of Central Florida, miméo.

BOWLES S., GINTIS H. [2000], « The evolution of strong reciprocity », Working paper, Santa Fe Institute.

BOWLES S., GINTIS H. [2011], *A Cooperative Species: Human Reciprocity and its Evolution*, Princeton, Princeton University Press.

BOWLES S., HWANG S.-H. [2008], « Social preferences and public economics: Mechanism design when social preferences depend on incentives », *Journal of Public Economics,* 92, p.1811-1820.

BRANDTS J., SAIJO T., SCHRAM A. [2004], « How universal is behavior? A four country comparison of spite, cooperation and errors in voluntary contribution mechanisms », *Public Choice,* 119, p.381-424.

BRANDTS J., SCHRAM A. [2001], « Cooperation and Noise in Public Goods Experiments: Applying the Contribution Function Approach », *Journal of Public Economics,* 79, p.399-427.

BROSIG J., OCKENFELS A., WEIMANN J. [2003], « The effect of communication media on cooperation », *German Economic Review,* 4, p.217–242.

BUCKLEY E., CROSON R. [2006], « Income and wealth heterogeneity in the voluntary

provision of linear public goods », *Journal of Public Economics,* 90 (4-5), p.935-955.

BURLANDO R., GUALA F. [2005], « Heterogenous agents in public goods experiments », *Experimental Economics,* 8 (1), p.35-54.

CAMERER C.F. [2011], « The promise and success of lab-field generalizability in experimental economics. A critical reply to Levitt and List », mimeo disponible à : http://ssrn.com/abstract=1977749.

CARPENTER J. [2007], « The demand for punishment », *Journal of Economic Behavior & Organization,* 62 (4), p.522-542.

CARPENTER J., MATTHEWS P.H. [2008], « Norm enforcement: The role of third parties », Middlebury College, Department of Economics Working Paper.

CHAN K.S., MESTELMAN S., MOIR R., MULLER R.A. [1999], « Heterogeneity and the voluntary provision of public goods », *Experimental Economics,* 2 (1), p.5-30.

CHARNESS G., RABIN M. [2002], « Understanding social preferences with simple tests », *Quarterly Journal of Economics,* 117, p.817-869.

CHARNESS G., YANG C.-L. [2008], « Endogenous Group Formation and Public Goods Provision: Exclusion, Exit, Mergers, and Redemption », University of California at Santa Barbara, Economics Working Paper Series 711912.

CHARNESS G., VILLEVAL M.C. [2009], « Cooperation, Competition, and Risk Attitudes: An. Intergenerational Field and Laboratory Experiment », *American Economic Review*, 99 (3),
p.956–978.

CHAUDHURI A. [2011], « Sustaining cooperation in laboratory public goods experiments: a selective survey of the littérature », *Experimental Economics,* 14 (1), p.47-83.

CHAUDHURI A., GRAZIANO S., MAITRA P. [2005], « Social learning and norms in a public goods experiment with inter-generational advice », *Review of Economic Studies,* 73, p.357–380.

CHAUDHURI A., MAITRA P., SKEATH S. [2009], « Communication, Advice and Beliefs in an Experimental Public Goods Game », University of Auckland, miméo.

CHERRY T.L., KROLL S., SHOGREN J.F. [2005], « The impact of endowment heterogeneity and origin on public good contributions: evidence from the lab », *Journal of Economic Behavior & Organization,* 57 (3), p.357-365.

CINYABUGUMA M., PAGE T., PUTTERMAN L. [2005], « Can second-order punishment deter perverse punishment? », *Experimental Economics,* 9, p.265-279.

CINYABUGUMA M., PAGE T., PUTTERMAN L. [2006], « Cooperation under the threat of expulsion in a public goods experiment », *Journal of Public Economics,* 89, p.1421-1435.

COX J.C., SADIRAJ V. [2005], « Social preferences and voluntary contributions to public goods », Andrew Young University Working Paper 05-22.

CROSON R. [2007], « Theories of commitment, altruism and reciprocity: evidence from linear public goods games », *Economic Inquiry,* 45, p.199–216.

CROSON R., GNEEZY U. [2009], « Gender Differences in Preferences », *Journal of Economic Literature*, 47 (2), p.448-474.

DAI Z., HOGARTH R., VILLEVAL M.C. [2012], « Frequency and randomness of audits and the efficiency of sanctions », GATE Lyon St Etienne, miméo.

DAMASIO, A. [1994], *Descartes' Error: Emotion, Reason, and the Human Brain*, New-York, Penguin Putnam.

DAWES R.M., MCTAVISH J., SHAKLEE H. [1977], « Behavior, communication and assumptions about other people's behavior in a commons dilemma situation », *Journal of Personality and Social Psychology,* 35, p.1-11.

DE QUERVAIN D., FISCHBACHER U., TREYER V., SCHALLHAMMER M., SCHNYDER U., BUCK A., FEHR E. [2004], « The neural basis of altruistic punishment », *Science,* 305, p.1254-1258.

DENANT-BOEMONT L., MASCLET D., NOUSSAIR C. [2007], « Punishment, counterpunishment and sanction enforcement in a social dilemma experiment », *Economic Theory,* 33, p.145-167.

DE OLIVEIRA A.C.M., CROSON R., ECKEL C. [2009], « One Bad Apple: Uncertainty and Heterogeneity in Public Good Provision », University of Texas at Dallas, mimeo.

DREBER A., RAND D.G., FUDENBERG D., NOWAK M.A. [2008], « Winners don't punish », *Nature*, 452 (20), p.348-351.

EGAS M., RIEDL A. [2008], « The economics of altruistic punishment and the maintenance of coopération », *Proceedings of the Royal Society B : Biological Sciences,* 275 (1637), p.871-878.

EHRARD K.M., KESER C. [1999], « Cooperation and Mobility: On the Run », SFB 504 Discussion Paper 99-69, Mannheim.

ERTAN A., PAGE T., PUTTERMAN L. [2008]. « Who to punish ? Individual décisions and majority rules in mitigating the free-rider problem », *European Economic Review,* 53 (5), p.495-511.

FALK A., FEHR E., FISCHBACHER U. [2005], « Driving forces behind informal sanctions », *Econometrica,* 73 (6), p.2017–2030.

FALK A., FISCHBACHER U. [2006], « A theory of reciprocity », *Games and Economic Behavior,* 54 (2), p.293–315.

FALK A., HECKMAN J. J. [2009], « Lab experiments are a major source of knowledge in the social sciences», *Science,* 326, p.535–538.

FEHR E., FISHBACHER U. [2002], « Why Social Preferences Matter – The Impact of Non-selfish Motives on Competition, Cooperation and Incentives », *The Economic Journal,* 112, p.C1-C33.

FEHR E., GÄCHTER S. [2000], « Cooperation and Punishment in Public Goods Experiments», *American Economic Review,* 90, p.980-994.

FEHR E., GÄCHTER S. [2002], « Altruistic punishment in humans », *Nature,* 415, p.137–140.

FEHR E., ROCKENBACH B. [2003], « Detrimental effects of sanctions on human altruism », *Nature,* 422, p.137–140.

FEHR E., SCHMIDT K.M. [1999], « A Theory of Fairness, Competition, and Cooperation », *Quarterly Journal of Economics,* 114, p.817-868.

FISCHBACHER U., GÄCHTER S. [2010], « Social preferences, beliefs, and the dynamics of free-riding in public good experiments », *American Economic Review,* 100 (1), p.541-556.

FISCHBACHER U., GÄCHTER S., FEHR E. [2001], « Are People Conditionally Cooperative? Evidence from a Public Goods Experiment », *Economics Letters*, 71, p.397-404.

GÄCHTER S., HERRMANN B., THÖNI C. [2010], « Culture and cooperation », *Philosophical Transactions of the Royal Society B*, 365, 2651-2661.

GÄCHTER S., NOSANZO D., RENNER E. [2011], « Sequential vs. simultaneous contributions to public goods: Experimental evidence», *Journal of Public Economic Theory,* 94 (7-8), p.515-522.

GÄCHTER S., NOSANZO D., RENNER E., SEFTON M. [2012], « Who makes a good leader ? Cooperativeness, optimism, and leading-by-example », à paraître in *Economic Inquiry*.

GÄCHTER S., RENNER E. [2005], « Leading by Example in the Presence of Free Rider Incentives », University of Nottingham, *CeDEx Discussion Paper*.

GÄCHTER S., RENNER E., SEFTON M. [2008], « The long run benefits of punishment », *Science,* 322, p.1510.

GÄCHTER S., THÖNI C. [2005], « Social learning and voluntary cooperation among like-minded people », *Journal of the European Economic Association,* 3, p.303-314.

GUNNTHORSDOTTIR A., HOUSER D., MCCABE K. [2007], « Disposition, history and contributions in public goods experiments », *Journal of Economic Behavior & Organization,* 62, p.304–315.

GÜRERK Ö, IRLENBUSCH B., ROCKENBACH B. [2006], « The competitive advantage of sanctioning institutions », *Science,* 312, p.108-111.

GÜRERK Ö, IRLENBUSCH B., ROCKENBACH B. [2009], « Motivating teammates: The leader's choice between positive and negative incentives », *Journal of Economic Psychology,* 30, p.591-607.

GÜTH W., LEVATI M.V., SUTTER M., VAN DER HEIJDEN E. [2007], « Leading-by-example with and without exclusion power in voluntary contribution experiments», *Journal of Public Economics,* 91 (5-6), p.1023-1042.

HARBAUGH,W.T. AND KRAUSE, K. (2000), « Children's Altruism in Public Good and Dictator Experiments », *Economic Inquiry*, 38 (1), p.95–109.

HEINRICH J., Boyd R., Bowles S., Camerer C., Fehr E., Gintis H. (Eds.) [2004], *Foundations of human sociality : economic experiments and ethnographic evidence in fifteen small-scale societies*. New-York : Oxford University Press.

HERMALIN B. [1998], « Toward an Economic Theory of Leadership: Leading-by-Example», *American Economic Review,* 88, p.1188-1206.

HERRMANN B., THÖNI C., GÄCHTER S. [2008], « Antisocial punishment across societies », *Science,* 319, p.1362–1367.

HOUSER D., XIAO E. [2011], « Punish in Public », *Journal of Public Economics,* 95, p.1006-1017.

HOUSER D., XIAO E., MCCABE K., SMITH V. [2008], « When Punishment Fails: Research on Sanctions, Intentions and Non-Cooperation », *Games and Economic Behavior*, 62 (2), p.509-532.

INGLEHART R., BAKER W.E. [2000], « Modernization, cultural change, and the persistence of traditional values », *American Sociological Review,* 65, p.19–51.

ISAAC R.M., MCCUE K., PLOTT C. [1985], Public Goods Provision in an Experimental Environment », *Journal of Public Economics,* 26 (1), p.51–74.

ISAAC R.M., WALKER J.M. [1988], « Communication and Free-Riding Behavior: The Voluntary Contributions Mechanism », *Economic Inquiry,* 26 (4), p.585-608.

ISAAC R.M., WALKER J.M. [1991], « Costly Communication: An Experiment in a Nested Public Goods Problem », in Palfrey T. (Ed.). *Contemporary Laboratory Research in Political Economy*, Ann Arbor, Univ. of Michigan Press.

JOFFILY M., MASCLET D., NOUSSAIR C., VILLEVAL M.C. [2011], « Emotions, sanctions and cooperation », IZA Discussion Paper 5592, Bonn.

KAMEI K., PUTTERMAN L., TYRAN J.-R. [2011], «State or Nature? Formal vs. Informal Sanctioning in the Voluntary Provision of Public Goods », Brown Economics Working Paper, disponible à SSRN: http://ssrn.com/abstract=1752266.

KERR N.L., KAUFMAN-GILLILAND C.M. [1994], « Communication, commitment, and cooperation in social dilemmas », *Journal of Personality and Social Psychology,* 66, p.513-529.

KOCHER M., CHERRY T., KROLL S., NETZER L., SUTTER M. [2008], « Conditional cooperation on three continents », *Economics Letters,* 101 (3), p.175-178.

KRISHNAMURTHY S. [2001], « Communication Effects In Public Good Games With And Without Provision Points », in M. Isaac (Ed.). *Research In Experimental Economics*, Volume Eight, Amsterdam : JAI.

KROLL S., CHERRY T.L., SHOGREN J.F. [2007], « Voting, Punishment, and Public Goods », *Economic Inquiry,* 45 (3), p.557-570.

KUMRU C., VESTERLUND L. [2010], « The Effect of Status on Charitable Giving », *Journal of Public Economic Theory,* 12 (4), p.709-735.

KURZBAN R., HOUSER D. [2005], « An experimental investigation of cooperative types in human groups: a complement to evolutionary theory and simulations », *Proceedings of the National Academy of Sciences,* 102 (5), p.1803-1807.

LEDYARD J. [1995], « Public goods: a survey of experimental research » in KAGEL J., ROTH A. (Eds.), *Handbook of Experimental Economics*, Princeton, Princeton University Press, p.111-194.

LEVATI M.V., SUTTER M., VAN DER HEIJDEN E. [2007], « Leading-by-Example in a Public Goods Experiment with Heterogeneity and Incomplete Information », *Journal of Conflict Resolution,* 51 (5), p.793-818.

LEVITT S., LIST J.A. [2007a], « Viewpoint: On the Generalizability of Lab Behavior to the Field », *Canadian Journal of Economics*, 40 (2), p.347-370.

LEVITT S., LIST J.A. [2007b], « What do Laboratory Experiments Measuring Social Preferences Reveal about the Real World », *Journal of Economic Perspectives,* 21 (2), p.153-174.

LEVY, D., PADGITT, K., PEART, S., HOUSER, D., XIAO, E. [2011], « Leadership, cheap talk and really cheap talk », *Journal of Economic Behavior and Organization,* 77, p.40-52.

MASCLET D., NOUSSAIR C., TUCKER S., VILLEVAL M.C. [2003], « Monetary and Non-Monetary Punishment in the Voluntary Contributions Mechanism », *American Economic Review,* 93 (1), p.366-380.

MASCLET D., NOUSSAIR C., VILLEVAL M.C. [2012], « Threats and sanctions in VCM games », à paraître in *Economic Inquiry.*

MASCLET D., VILLEVAL M.C. [2008], « Punishment, Inequality and Welfare: A Public Good Experiment », *Social Choice and Welfare,* 31(3), p.475-502.

MEIDINGER C., VILLEVAL M.C. [2003], « Leadership in Teams: Signaling or Reciprocating?», GATE Working Paper 10-03, Lyon.

MOXNES E., VAN DER HEIJDEN E. [2003], « The Effect of Leadership in a Public Bad Experiment », *Journal of Conflict Resolution,* 47 (6), p.773-795.

NIKIFORAKIS N. [2008], « Punishment and Counter-Punishment in Public Good Games: Can We Really Govern Ourselves? », *Journal of Public Economics,* 92, p.91–112.

NIKIFORAKIS N., ENGELMANN D. [2012], « Altruistic punishment and the threat of feuds », à paraître in *Journal of Economic Behavior & Organization*.

NIKIFORAKIS N., NORMANN H.T. [2008], « A Comparative Statics Analysis of Punishment in Public Good Experiments », *Experimental Economics,* 11, p.358-369.

NOSANZO D., SEFTON M. [2011], « Endogenous move structure and voluntary provision of public goods: theory and experiment », *Journal of Public Economic Theory,* 13 (5), p.721-754.

OSTROM E., WALKER J.M., GARDNER R. [1992], « Covenants with and without a sword—self-governance is possible », *American Politicical Science Review,* 86, p.404–417.

PAGE T., PUTTERMAN L., UNEL B. [2005], « Voluntary Association in Public Goods Experiments: Reciprocity, Mimicry and Efficiency», *The Economic Journal,* 115 (506), p.1032-1053.

PALFREY T., PRISBREY J.E. [1997], « Anomalous Behavior in Public Goods Experiments: How Much and Why? », *American Economic Review,* 87, p.829-846.

POTTERS J., SEFTON M., VESTERLUND L. [2005], « After You - Endogenous Sequencing in Voluntary Contribution Games », *Journal of Public Economics,* 89, 1399-419.

POTTERS J., SEFTON M., VESTERLUND L. [2007], « Leading-by-example and signaling in voluntary contribution games: An experimental study », *Economic Theory,* 33 (1), 169-182.

POTTERS J., SEFTON M., VAN DER HEIJDEN E. [2009], « Hierarchy and Opportunism in Teams », *Journal of Economic Behavior & Organization,* 69 (1), 39-50.

RABIN M. [1993], « Incorporating Fairness into Game Theory and Economics », *American Economic Review,* 83, p.1281-1302.

REUBEN E., RIEDL A. [2009], « Enforcement of Contribution Norms in Public Good Games with Heterogeneous Populations », IZA Discussion Paper 4303, Bonn.

RIVAS M.F., SUTTER M. [2011], « The benefits of voluntary leadership in experimental public good games », *Economics Letters,* 112 (2), p.176-178.

ROCKENBACH B., MILINSKY M. [2006], « The efficient interaction of indirect reciprocity and costly punishment », *Nature,* 444, p.718-723.

SALLY D. [1995], « Conversation and cooperation in social dilemmas: a meta-analysis of experiments from 1958 to 1992 », *Rationality and Society,* 7, p.58-92.

SAMUELSON P.A., NORDHAUS W.D. [1985], *Principles of Economics*, New-York, McGraw-Hill, Edition 12.

SAMUELSON P.A., NORDHAUS W.D. [1992], *Principles of Economics*, New-York, McGraw-Hill, Edition 14.

SEABRIGHT P. [2012], *The War of the Sexes: How Conflict and Cooperation Have Shaped Men and Women from Prehistory to the Present*, Princeton, Princeton University Press, à paraître.

SEFTON M., SHUPP R., WALKER J. [2007], « The Effect of Rewards and Sanctions in the Provision of Public Goods », *Economic Inquiry,* 45, p.671-690.

SEYFARTH R.M., CHENEY D.L. [1984], « Grooming, alliances and reciprocal altruism in vervet monkeys », *Nature,* 308, p.541-543.

SUGDEN R. [1984], « Reciprocity : The Supply of Public Goods through Voluntary Contributions », *The Economic Journal*, 94, p .772-787.

SUTTER M., HAIGNER S., KOCHER M. [2010], « Choosing the Carrot or the Stick? – Endogenous Institutional Choice in Social Dilemma Situations », *Review of Economic Studies,* 77 (4), p.1540-1566.

THÖNI C. [2011], « Inequality Aversion and Antisocial Punishment », University of St. Gallen, Discussion Paper 2011-11.

TYRAN J.-R., FELD L.P. [2006], « Achieving Compliance when Legal Sanctions are Non-deterrent », *Scandinavian Journal of Economics,* 108 (1), p.135-156.

VARIAN H.R. [1994], « Sequential provision of public goods », *Journal of Public Economics,* 53, p.165–186.

YAMAGISHI T. [1988], « The provision of a sanctioning system in the United-States and Japan », *Social Psychology Quarterly,* 51, p.265–271.

ZELMER J. [2003], « Linear Public Goods Experiments : A Meta-Analysis », *Experimental Economics*, 6, p.299-310.