

Quelles préférences expliquent la coopération ?

Les théories du choix rationnel – en situation de décision comme d’interaction – fondent leurs explications des actions humaines sur deux types d’états mentaux : d’une part les désirs, ou préférences, des agents, et d’autre part leurs croyances. Le schéma explicatif est le suivant : un agent rationnel choisira l’action dont il croit qu’elle mène aux conséquences qu’il préfère. L’un des problèmes de cette analyse est que l’observation des actions d’un agent sous-détermine le couple préférence-croyance qui aurait pu y mener. Le choix d’ouvrir mon parapluie lorsqu’il pleut pourrait également être expliqué par mon désir de ne pas être mouillé et ma croyance qu’un parapluie est efficace, aussi bien que par mon désir de laver mon parapluie et ma croyance que la pluie fournira un lavage efficace... Certaines croyances, et certains désirs, peuvent eux-mêmes paraître plus raisonnables que d’autres, mais la théorie du choix rationnel ne s’occupe pas de leur origine. Ce problème est amplifié dans le cas de situations d’interaction entre plusieurs agents : l’ensemble des désirs et préférences pouvant expliquer une action collective donnée devient considérablement plus vaste, d’abord parce que les agents sont plus nombreux mais surtout parce que leurs préférences peuvent dépendre les unes des autres et que certaines de leurs croyances portent sur les préférences et croyances d’autrui.

Nous nous intéresserons ici aux situations de dilemme coopératif, tel le dilemme du prisonnier, dans lesquelles les actions des agents ne peuvent être expliquées par des fonctions d’utilité (reflétant leurs préférences) qui s’aligneraient sur leurs gains matériels. Dans de tels cas, les théoriciens des jeux introduisent habituellement des préférences « sociales », qui reflètent différents *types* possibles d’agents – correspondant par exemple à leurs tendances altruistes ou égalitaristes. Or de nombreux profils de préférences peuvent fournir une telle explication. En particulier, certains modèles expliquent rationnellement le comportement coopératif en postulant un grand nombre (voire une infinité) de fonctions d’utilité – autrement dit, la coopération est fondamentalement expliquée par le fait que les agents pourraient avoir un large éventail de goûts plus ou moins pour la coopération. Ce type de solution nous paraît insatisfaisant pour au moins trois raisons : 1/ formellement, le nombre de paramètres nécessaires est trop élevé par rapport au nombre de comportements à expliquer ; 2/ empiriquement, rien n’indique que des agents peuvent effectivement avoir des préférences aussi variées ; 3/ conceptuellement, expliquer la coopération par la donnée de préférences *pour* la coopération elle-même est proche de la circularité.

En conséquence, nous soutenons que, d’une part, l’inflation ontologique de tels modèles peut être évitée, et que d’autre part, une préférence sociale doit satisfaire certaines propriétés pour être acceptable. Pour cela, nous fournissons d’abord un modèle explicatif de la coopération qui prend en compte des comportements coopératifs divers mais ne postule qu’un petit nombre de préférences sociales et situe plutôt l’origine de la variété des types d’agents dans la palette de *croyances* qu’ils ont les uns sur les autres. Autrement dit, nous utilisons la flexibilité des explications en termes de choix rationnel pour mettre l’accent sur les croyances au lieu des préférences, ce qui ne pose aucun problème ontologique. Nous soutenons que le pouvoir explicatif de ce modèle est comparable à celui de ces concurrents, pour une complexité moindre. Dans un second temps, nous soutenons qu’une fonction d’utilité sociale, de par son lien avec le concept d’objectif collectif, doit remplir certaines conditions formelles. L’ensemble des préférences sociales est donc contraint, et fondé non sur un goût fondamental pour la coopération mais par l’adoption de certains objectifs. Nos analyses répondent ainsi aux insuffisances mentionnées et suggèrent que l’explication formelle de la coopération peut rester simple et réaliste.