# Articles

# Une forme minimale de coopération

CÉDRIC PATERNOTTE *Bristol University*

*RÉSUMÉ : La plupart des nombreuses définitions existantes d'une action coopérative en fournissent des conditions suffisantes plutôt que nécessaires. Nous définissons ici une forme minimale de coopération, correspondant aux actions de masse, telles des manifestations. Nous en détaillons les aspects intentionnel, épistémique, stratégique et téléologique, généralement obtenus par affaiblissement spécifique de concepts classiques. Parallèlement, nous soulignons le rôle crucial de concepts issus de la théorie des jeux pour la définition d'une action coopérative. Enfin, nous soutenons que la rationalité d'une action coopérative minimale est cruciale à sa réalisation et pas seulement possible ou souhaitable comme le soutiennent les analyses habituelles.*

*ABSTRACT : Definitions of cooperative action usually provide sufficient rather than necessary conditions. I here define a minimal form of cooperation, which encompasses mass actions, such as demonstrations. Intentional, epistemic, strategic and teleological aspects, mainly obtained from a weakening of pre-existing concepts, are discussed. I also emphasize the key role played by game theoretic concepts in the definition of joint action. Finally, I claim that rationality is a necessary condition of minimal cooperation, whereas classical analysis merely sees rationality as a possible or desirable characteristic.*

***Dialogue* 48 (2009), 235-267. Printed in the U.S.A.**
**© 2009 Canadian Philosophical Association/Association canadienne de philosophie**
**doi:10.1017/S0012217309090246**

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

## 1. Les définitions classiques

Quand un ensemble d'actions individuelles représente-t-il un cas de coopération? Cette question de la juste définition d'une action coopérative, ou action conjointe, occupe les philosophes depuis quelques décennies. Une première évidence est que le seul comportement observable des individus ne suffit pas à déterminer s'ils coopèrent ou non. Fondamentalement, on distingue une action coopérative d'une simple action collective[1]. Cette dernière existera dès qu'un ensemble d'actions individuelles produit un effet global remarquable. Un mouvement de foule quelconque peut par exemple être qualifié d'action collective. Autrement dit, l'existence d'une action collective dépend simplement de caractéristiques observables de son résultat[2]. Au sens large, toute juxtaposition d'actions individuelles pourrait être qualifiée d'action collective.

Toute action coopérative, en ce qu'elle suppose au moins plusieurs actions individuelles, est donc une action collective. Elle nécessite cependant à l'évidence comme ingrédient supplémentaire un certain lien entre les agents. Coopérer signifie agir ensemble, de concert, d'intelligence. Ces expressions, qu'il s'agit d'élucider, indiquent une certaine proximité ou cohérence entre les états mentaux des individus. Autrement dit, pour coopérer des agents doivent avoir certains états mentaux qui prennent notamment en compte la présence d'autrui, c'est-à-dire certains de ses propres états mentaux supposés et de ses actions. La coopération consiste donc en un tissu d'actions et de certains états mentaux. Plus précisément, de même que l'analyse orthodoxe de la connaissance la voit comme une croyance à laquelle s'ajoutent des propriétés supplémentaires (être vraie et justifiée), une action coopérative peut être comprise comme une action collective à laquelle s'ajoutent des conditions appropriées sur les états mentaux des agents.

Quels types d'états mentaux sont impliqués dans la coopération? Regan (1980), dans l'une des premières analyses de la coopération (bien qu'il n'ait pas eu pour intention de la définir), propose les pistes suivantes. Deux actions coopérative et non coopérative étant déterminées d'avance, la coopération peut échouer si un agent se trompe d'action (ce qui serait observable); s'il choisit la bonne action par erreur, la confondant avec l'autre; ou s'il choisit la bonne action en se méprenant sur l'action coopérative d'autrui. Les individus doivent donc choisir la bonne action avec l'intention, la motivation et la compréhension appropriées à la situation. Ils doivent également croire qu'autrui a ces motivation, intention et compréhension correctes; croire qu'autrui le croit également, et ainsi de suite : «la coopération implique une hiérarchie potentiellement infinie de croyances réciproques» (*ibid.*, p. 129). Autrement dit, les conditions précédentes doivent être de connaissance (ou croyance) commune. La définition d'une action coopérative a donc une *structure tripartite* : les *présuppositions* — la donnée préalable de combinaisons d'actions acceptables caractérisant la situation de départ; les *prémisses* — un ensemble d'états mentaux stipulant les modalités de l'action; et la *réalisation* — l'accomplissement correct de ces actions[3]. Les prémisses se décomposent elles-mêmes en un premier ensemble fondamental

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

d'états mentaux, auquel s'ajoute la condition générale de croyance commune. C'est la détermination de ce premier ensemble des prémisses de la coopération qui constitue ordinairement l'essentiel du problème de sa définition. L'ensemble des prémisses correspond habituellement à ce qu'on appelle une *intention de groupe* dans la littérature. En résumé, on dira donc que des individus coopèrent lorsqu'ils accomplissent (réalisation) certaines actions (présuppositions) avec une intention de groupe.

À titre d'exemple, voici une définition simple tirée de Tuomela et Miller (1988). Soit une action conjointe A composée de deux action A1 et A2 respectivement effectuables par les agents P1 et P2. P1 aura une *intention de groupe* par rapport à A si :

i) P1 a l'intention de faire A1;
ii) P1 croit que P2 fera A2;
iii) P1 croit que P2 croit que P1 fera A1; et ainsi de suite.

Ces prémisses sont extrêmement simples, puisqu'elles se résument à un ensemble d'intentions individuelles qui sont de croyance commune. La plupart des définitions ultérieures ont, tout en conservant cette structure, détaillé la façon dont sont liées intentions et croyances. Pour Gilbert, des agents coopèrent s'ils sont d'abord «conjointement disposés à [s'engager ensemble] dans l'action A dans des circonstances C si et seulement si il est de connaissance commune entre eux qu'ils ont mutuellement exprimé leur quasi-disposition à [s'engager]» (1989, p. 197) et qu'ils forment ensuite des intentions individuelles d'agir, qui sont de connaissance commune. Elle enrichit donc la définition précédente en détaillant la façon dont les intentions se forment : les agents ont d'abord des intentions conditionnelles mutuelles publiques, que leur simultanéité transforme en intentions inconditionnelles. Les définitions ultérieures complexifieront cette analyse en introduisant de nouveaux concepts et en rendant explicites les raisons d'agir des individus — les raisons pour lesquelles ils forment leurs intentions. Bratman (1992), dans un cadre dynamique, montre l'importance des plans d'action suivis par les agents, la coopération disparaissant lorsque les plans de chacun ne sont pas cohérents entre eux. Miller (2001) marque le rôle crucial des buts collectifs (ne faisant pas eux-mêmes l'objet d'une définition), autrement dit des raisons pour lesquelles les agents agissent : à action et croyances égales, des agents peuvent ou non coopérer selon qu'ils visent la réalisation d'un but collectif ou non.

Mais c'est Tuomela qui a établi ce que nous considérerons comme la définition classique de l'action coopérative. Ce dernier, là où les autres auteurs livraient une seule analyse, a construit tout un ensemble de définitions imbriquées nuancées, afin d'englober tous les cas, allant de l'action collective la plus simple à la coopération la plus indiscutable. La construction de Tuomela procède par emboîtements successifs de définitions de concepts distincts, conformément à l'analyse précédente : en résumé, des agents coopèrent s'ils accomplissent un

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

but collectif avec une intention conjointe, ces deux éléments étant de connaissance commune et cette connaissance commune étant elle-même une des raisons de formation de l'intention (cf. Tuomela, 2007). Tuomela met l'accent sur la nature collective du but, c'est-à-dire le fait qu'il a été choisi ou construit collectivement par les agents avant l'action. Cela garantit que les agents agissent selon «un mode de groupe», sans lequel la coopération dégénérerait en une sorte d'action conjointe d'aide mutuelle.

Notons que les définitions ici présentées partagent un présupposé commun : l'individualisme méthodologique. Toutes supposent en effet que leurs objets, des concepts collectifs, peuvent être ramenés à un ensemble de concepts individuels imbriqués; autrement dit, qu'ils ne sont pas irréductibles[4]. Il existe au moins une grande analyse, celle de Searle (1990) qui suppose au contraire que les intentions individuelles découlent d'intentions collectives irréductibles et premières. La qualité d'une telle théorie est de mettre en évidence un rapport causal entre intention de groupe et intentions individuelles : c'est parce que la première existe que les agents décident effectivement d'accomplir leur part de l'action. Mais son inconvénient est qu'en considérant une intention collective irréductible, elle se contente de postuler son existence sans l'analyser. Or, comme l'ont remarqué Chant et Ernst (2007), il est possible de sauvegarder la première intuition tout en s'épargnant un tel coût ontologique : on peut très bien supposer que les intentions individuelles ne sont adoptées que parce qu'elles entretiennent certaines relations. Autrement dit, une intention collective est analysable en termes d'intentions individuelles imbriquées, qui ne sont des intentions qu'en raison de leur imbrication. Par la suite, nous nous limiterons donc à une vision de la coopération compatible avec l'individualisme méthodologique.

Ce rapide passage en revue révèle la nature des principaux ingrédients potentiels de la coopération : intentions, croyances, buts, raisons d'agir. Notons que d'autres notions envisageables, comme les accords ou les promesses, ne sont pas considérés comme des ingrédients de la coopération, puisqu'ils ne font rien d'autre que garantir certaines croyances chez les individus et n'ont donc pas leur place en tant qu'éléments fondamentaux : ils représentent des manières dont les ingrédients nécessaires à la coopération peuvent apparaître. Un autre point important est l'importance de l'intuition dans ces définitions : engagements, plans d'action, buts collectifs, mode de groupe, sont respectivement considérés comme des éléments cruciaux par chacun des auteurs précédents. Autrement dit, les théoriciens de l'action collective ont privilégié la recherche de conditions *suffisantes*. La démarche d'un auteur consiste souvent à citer des exemples de coopération manifeste, avant de l'analyser pour en dégager la structure et les éléments qui la composent. Selon les intuitions des auteurs, on aboutit ainsi à des analyses sensiblement différentes. Au contraire, l'analyse de Regan, presque jamais citée dans la littérature des actions conjointes et nettement moins sophistiquée que les analyses précédentes, a toutefois le rare avantage de rechercher des conditions *nécessaires* de la coopération. Regan se

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

demande explicitement dans quels cas de figure la coopération pourrait échouer, et énumère ainsi des éléments indispensables à son existence.

En d'autres termes, il manque dans la littérature existante une caractérisation minimale de la coopération, qui se concentrerait sur des aspects suffisants mais surtout nécessaires. De fait, chacune des analyses précédentes semble pouvoir être affaiblie sans que la coopération ne disparaisse. Il semble en effet qu'il puisse y avoir coopération sans engagement, sans plan d'action détaillé, sans but construit collectivement, et surtout sans connaissance ou croyance commune. Notre travail va maintenant consister à exposer ces cas de coopération minimale qui échappent aux analyses classiques et à en donner une définition satisfaisante. Nous comptons ainsi nous rapprocher du cœur conceptuel de la coopération humaine.

Le plan de la suite de l'article sera le suivant. Dans la section 2, nous introduirons les cas de coopération minimale pour lesquels aucune définition existante n'est satisfaisante et suggérerons qu'ils peuvent être compris à partir de l'affaiblissement méthodique d'une définition classique. La section 3 soulignera la pertinence, habituellement ignorée, de la théorie des jeux pour la définition d'une intention de groupe, et suggérera qu'un certain type d'équilibre peut être interprété de façon cognitivement réaliste et convient aux cas de coopération minimale. Nous construirons ensuite, dans la section 4, la définition analytique proprement dite, en soulignant l'importance de ses aspects rationnels. Enfin, la section 5 sera consacrée à la discussion des caractéristiques de notre définition, qui évite certains problèmes classiques des actions conjointes tout en apportant des précisions conceptuelles bienvenues.

## 2. Vers une définition minimale

### *2.1. Actions de masse*

Imaginons un groupe de cent individus souhaitant tous le succès d'une manifestation et ayant tous dit qu'ils viendraient certainement. Le lieu ou les conditions de la manifestation sont tels que s'y rendre effectivement est coûteux. Les agents ne peuvent communiquer et ne découvriront qu'une fois à la manifestation l'identité des autres participants. Imaginons encore que la manifestation soit un succès à partir de cinquante participants. *A priori*, la participation d'aucun agent n'est assurée. Quinze minutes avant le début, tous ont pris leur décision et soixante agents ont décidé de venir : la manifestation est un succès. Notre intuition est que l'on peut légitimement parler de coopération dans ce cas. Tous les agents avaient le même but collectif dont ils souhaitaient la réalisation; mais tous n'avaient pas besoin de participer. Le but collectif a bien été atteint, et les non participants en sont également satisfaits; ils ne sont pas venus mais ne s'y sont pas opposés et n'ont pas rendu sa réalisation impossible. Or cette situation n'est pas coopérative pour l'ensemble du groupe au sens des définitions que nous avons présentées auparavant, pour une première raison : certains membres n'ont pas participé, et ce de manière volontaire — ils n'avaient pas l'intention de participer.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

Imaginons maintenant également que dans les mêmes conditions, les cent agents soient effectivement venus. Tous ont cette fois eu l'intention correcte, mais ils ne l'ont pas formée pour les bonnes raisons. On peut imaginer que les agents présents avaient décidé de venir en estimant qu'il y aurait assez de participants pour que la manifestation soit un succès. L'intention de venir a donc été formée non sur la connaissance des intentions des autres participants, mais, par exemple, à partir d'une estimation de la propension de chacun à venir. En particulier, il n'y avait pas de connaissance commune de l'ensemble des participants puisque ceux-ci ne connaissaient pas même leurs identités respectives avant d'arriver à la manifestation. Il nous semble pourtant, dans cette variation encore plus que dans le premier exemple, que l'on peut légitimement parler de coopération.

On pourrait cependant accuser ces cas d'irréalisme, en soutenant par exemple qu'une manifestation est en réalité un phénomène dynamique où la communication préalable joue de surcroît un grand rôle. Un deuxième exemple, similaire mais plus convaincant, montre qu'il n'en est rien. Il y a quelques années est apparu un phénomène social appelé «foule éclair»[5] : dans un lieu public, un groupe d'individus se forme; ils accomplissent au même moment une action identique, avant de se disperser aussitôt. L'une des particularités de ces événements est que les individus ne se connaissaient pas auparavant : ils auront lu sur un site Internet, ou reçu électroniquement, les instructions, puis choisi de participer. L'action en question est toujours gratuite (le succès de l'action est le seul bénéfice des participants), et les individus agissent sans avoir besoin communiquer entre eux. Par exemple, fin 2003, des personnes se sont assemblées dans le hall du Louvre, se sont simultanément mises à téléphoner en marchant, avant de s'arrêter, claquer des mains vers le ciel, et se disperser. Les foules éclairs sont un cas d'école de coopération minimale : lorsque les individus décident ou non de participer, ils ignorent l'identité et le nombre des autres participants. Ils ne peuvent fonder leur décision que sur la source du message (Est-elle suffisamment publique? Combien de personnes ont lu le message? Quelles sont les chances qu'un lecteur quelconque décide de participer?), ainsi que sur la réussite passée de tels événements (plus le phénomène des foules éclairs devenait connu, plus les actions ont eu de participants et d'impact).

Les foules éclairs sont des phénomènes d'action collective, mais surtout de coopération. Le but de l'opération est de produire un effet perceptible massif. Un participant arrivant sur les lieux et ne voyant aucun autre participant potentiel renoncerait sans doute à cancaner, seul au milieu de la foule, avant de jeter un canard en plastique dans une fontaine…[6] le but recherché est précisément l'étonnement des passants devant une action subite, dont le début, le déroulement et l'arrêt paraissent sans cause, et sans signification. On pourrait cependant contester la valeur de cet exemple : le message d'origine contenant les instructions semble témoigner d'une organisation qui contredit l'aspect «émergent» de l'action collective. En réalité, le message initial joue un rôle simple et qui

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

s'inscrit dans notre cadre : il décrit la nature de l'action coopérative. Les lecteurs auront ensuite à faire un choix élémentaire : suivre les instructions (coopérer) ou non. Quant aux éventuelles interactions sur place entre individus se reconnaissant, elles ne rendent pas la situation plus complexe : l'action coopérative consiste d'abord à venir ou non; elle ne pourrait pas avoir lieu si les participants présents étaient trop peu nombreux.

La théorie des jeux expérimentale nous fournit un troisième type d'exemples. Depuis des décennies[7] et encore ces dernières années, des expériences ont permis d'étudier le comportement d'individus face à des situations de coopération potentielle, le plus souvent représentées par un dilemme du prisonnier ou un jeu de bien public — le terme de coopération étant ici pris au sens observable d'obtention d'un bénéfice mutuel. Dans une partie considérable de ces expériences, les agents font face à une situation ponctuelle (c'est-à-dire qui ne se produit qu'une fois), sans communication possible et dans des conditions d'anonymat. Ils doivent donc décider de coopérer en l'absence totale d'information directe sur autrui, ne connaissant que le nombre de participants potentiels. Nous soutenons que lorsque des individus choisiront l'action pouvant conduire au bénéfice mutuel dans une telle situation, il peut y avoir coopération minimale au même sens que dans les exemples de manifestation et de foule éclair. Il faut bien sûr pour cela que les agents aient choisi l'action pour de bonnes raisons, c'est-à-dire au moins en vue d'atteindre le bénéfice mutuel considéré comme un but collectif. De tels cas comportent même moins d'incertitude que ceux des manifestations, puisque dans ces derniers les agents ignorent le nombre d'individus impliqués dans la situation (la percevant) et qu'ils doivent donc prendre en compte son degré de publicité lors de leur décision. Face à un jeu expérimental, les agents savent au moins combien d'autres personnes font face à la même situation qu'eux. De ce point de vue, la coopération dans un jeu expérimental est plus proche des définitions classiques que dans les exemples de manifestation ou de foule éclair.

La différence majeure entre ces exemples de coopération et les définitions existantes est que l'identité des participants n'est pas connue de tous — et *a fortiori* n'est pas de connaissance commune, ni au sein du groupe entier, ni du sous-groupe des participants. Le type d'exemples de coopération minimale contient donc en particulier (mais pas seulement) ce que l'on peut appeler des actions de masse. En effet, plus les participants possibles d'une action coopérative sont nombreux, plus il est plausible qu'ils ne se connaissent pas d'avance. Les conditions exigées notamment par les définitions existantes seraient donc trop fortes. À la lumière de l'exemple précédent, les quelques modifications souhaitables semblent être les suivantes. D'abord, il faut un concept de croyance commune qui n'implique pas de contact direct entre les agents. Ensuite, il faut remplacer l'exigence de croyance commune des intentions d'agir par celle d'éléments moins exigeants — par exemple des propensions à agir d'une certaine façon — et celle de l'identité des participants par celle de l'ensemble des participants possibles. Enfin, il faut assouplir ou modifier les raisons fondant

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

les intentions d'agir, en les débarrassant au minimum de la référence à d'autres intentions particulières clairement identifiées. En résumé, une définition de la coopération devrait être assez faible et générale pour se passer de la condition selon laquelle les membres du groupe connaissent les participants et leurs états mentaux avant l'action. D'une façon générale, il serait utile de distinguer le groupe par rapport auquel un but collectif est défini, et le groupe de participants effectifs. La distinction devra être plus ou moins marquée selon que l'on accepte comme exemple(s) de coopération les deux situations ou seulement la seconde.

Revenons une dernière fois sur nos exemples et sur les raisons pour lesquelles ils peuvent sembler trop frustes pour être de la coopération. Les manifestations, et encore davantage les foules éclair, mettent en évidence les raisons pour lesquelles il est difficile de se décider à parler de coopération pour de tels cas. Il existe deux catégories de problèmes auxquels un agent peut faire face : les problèmes de décision et les problèmes d'interaction. Dans les premiers, l'agent doit faire un choix face à la nature : la présence de l'agent n'a pas d'influence sur les événements, qui sont extérieurs. La résolution de tels problèmes est du ressort de la théorie de la décision. Dans les problèmes d'interaction, plusieurs agents étant en présence, la situation à laquelle fait face un agent dépend de la façon dont il est perçu par les autres. Il peut leur donner des informations, ou en recevoir; et surtout, sa stratégie rationnelle doit posséder certaines vertus de stabilité, elle dépend des autres, qui dépendent d'elle : l'ensemble des stratégies envisagées doit former un tout cohérent.

Dans les situations qui nous intéressent, la communication et l'échange d'information sont absents, ce qui élimine la première forme de rapport avec autrui. Quant à la façon de choisir une action, elle se rapproche du raisonnement tenu dans un problème de décision, particulièrement quand le nombre d'agents est élevé : car un agent ne fondera son raisonnement que sur son estimation des chances qu'autrui a d'agir en vue d'un but collectif. Autrement dit, la réflexion sur la participation d'autrui se trouve assez sèchement rapprochée de la simple analyse d'un aspect concret de la situation et le problème d'interaction d'un problème de décision. C'est ici que nous situons la cause de la difficulté à penser ce type de situation en termes de coopération. Selon nos intuitions fondamentales, coopérer signifie agir de concert avec une autre personne, qui nous fait face et à laquelle on adapte son action; c'est une pleine situation d'interaction. Le cas d'un ensemble d'agents décidant de participer après avoir chacun résolu un problème individuel de quasi-décision se rapproche d'une conception mécaniste du fonctionnement d'un groupe : tels des automates, les agents appliquent presque une routine de décision qui inclut autrui dans un environnement régi par des forces étrangères.

Il faut parfois se méfier de l'intuition. S'il est compréhensible et même normal de douter que le phénomène qui nous intéresse est bien de la coopération, c'est parce qu'il s'agit d'une forme minimale, qui a vocation à nous rapprocher de conditions nécessaires à toute forme de coopération. Les définitions

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

classiques ne rencontrent pas ce problème puisqu'elles se développent à partir d'exemples intuitivement convaincants. En recherchant au contraire des cas minimaux, il est naturel de se rapprocher de la réelle limite séparant la coopération de la simple action collective; et il est prévisible que l'intuition, en atteignant sa limite d'acuité, ait davantage de difficultés à discriminer au niveau de détail nécessaire à l'analyse. Autrement dit, l'appréhension de la limite inférieure de la coopération se situerait aux limites de l'intuition. Cependant, nos situations minimales comportent encore les caractéristiques fondamentales de la coopération, puisque nous verrons que leur définition en conserve la structure caractéristique.

### *2.2. Affaiblissements*

Faut-il construire une définition nouvelle de la coopération minimale, ou pouvons-nous l'obtenir en affaiblissant une définition existante? Peu de définitions peuvent se prêter à une telle manipulation. La théorie holiste de Searle ne convient pas puisqu'elle est étrangère au projet même d'analyser une intention collective en tant qu'assemblage d'éléments plus fondamentaux. La définition de Gilbert n'est pas adaptée aux cas de coopération minimale, car en l'absence de contact direct entre individus, aucun engagement ou quasi-disposition ne peut être exprimé. Le cadre de telles actions n'étant pas nécessairement dynamique, les conditions de cohérence des plans d'action à la Bratman ne jouent plus de rôle. Par contre, la théorie de Tuomela est adaptable, ce qui n'est pas surprenant puisque ses définitions sont volontairement structurées de façon à pouvoir traiter un large éventail de cas. Dans cette sous-section, nous allons successivement suggérer que les notions de buts collectif et de croyance commune sont, sous une forme affaiblie, nécessaires pour définir la coopération minimale.

Quelques exemples vont permettre d'introduire l'argument concernant les buts collectifs. Chwe (2001) a montré que les produits présentés dans les coupures publicitaires du Super Bowl (l'événement télévisé le plus largement regardé aux États-Unis) sont en majorité ce qu'il appelle des *produits sociaux*, que l'on préfère consommer si l'on sait que d'autres les consomment également. Il est par exemple préférable d'avoir un ordinateur de la même marque que les autres pour des raisons de compatibilité.

À première vue, la publicité qui opère dans le cadre d'un problème de coordination (où plusieurs options équivalentes sont disponibles et où les joueurs souhaitent choisir la même sans se concerter d'avance) consiste à rendre l'une des options plus saillantes que les autres afin d'augmenter les chances que tous la choisissent. Mais on peut également regarder la situation comme un problème de coopération, si l'on ajoute la possibilité pour chaque individu de ne choisir aucun des produits disponibles. La publicité rend un produit saillant et le présente comme social (bon quand au moins une catégorie de population l'achète).

On peut douter qu'il s'agisse ici de coopération, et non pas d'un phénomène plus faible dans lequel les individus tiennent compte les uns des autres sans

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

toutefois pouvoir être considérés comme agissant ensemble. Il semble toutefois que l'on puisse parler de coopération si l'on admet que certains produits seront de préférence consommés à condition que suffisamment d'autres les consomment également, *et ce parce qu'il est globalement souhaitable qu'un même produit de ce genre soit consommé*. Il est certes difficile de considérer le résultat souhaitable comme un résultat collectif dont l'accomplissement est de connaissance commune au sein des agents, et ce pour au moins deux raisons. D'abord, cet éventuel but collectif est particulièrement artificiel et volatil : il ne constitue un but que pour ceux que la publicité convainc de la supériorité (en valeur absolue ainsi qu'au niveau social) du produit présenté. Il n'est pas préexistant à la publicité, qui le suscite et en entraîne la perception. En particulier, il n'est pas, comme l'exige Tuomela dans la pleine coopération, le résultat d'une construction collective concertée entre les individus. Il n'en est pas moins collectif; la façon dont un but collectif est perçu ne saurait influer sur sa nature. Ensuite, les cas sont tout simplement rares dans lesquels les consommateurs se focalisent sur un même produit social. À ceci, nous objecterons que la rareté de l'accomplissement d'un acte coopératif ne préjuge pas de sa nature mais de sa difficulté. Or, il s'agit d'une coopération difficile : la publicité est un facteur favorisant d'une force tout relative; les produits ainsi signalés sont multiples et variés; les événements aussi publics que Super Bowl ne sont pas pléthore; et enfin, même en admettant le concept de produit social, le bénéfice tiré d'une coopération réussie n'est pas immensément supérieur au gain en cas d'échec.

Cet exemple révèle toutefois le rôle central d'un but collectif : il est indispensable que les agents pensent que la consommation généralisée du produit est souhaitable, comme le suggère un second exemple. Tuomela encore, discutant certains cas d'actions collectives faibles, qui contiennent une dimension conjointe sans toutefois pouvoir être qualifiées de coopératives, décrit la situation suivante : dans une ville, les jeunes filles préfèrent s'habiller de la même façon, en portant des mini-jupes (Tuomela, 2007, p. 153). Il s'agit manifestement d'une action collective fondée sur l'imitation et la croyance des jeunes filles qu'un nombre minimum d'autres s'habille probablement de la même façon. Ce qui différencie ce cas de l'exemple de Chwe est l'absence d'un but collectif. Les jeunes filles cherchent simplement à se coordonner, sans estimer pour autant qu'être toutes habillées de façon semblable est bénéfique au groupe. Il suffirait que les jeunes filles agissent de la sorte *afin de favoriser le groupe*, par exemple dans le but de manifester davantage la présence du groupe dans la ville, pour que l'action devienne à notre sens coopérative et semblable au cas de Chwe. Sinon, on n'a que ce que Tuomela appelle une action sociale collective parallèle. Autrement dit, toutes choses étant égales par ailleurs, et notamment pour des relations également distantes entre les individus, dans certaines situations, qu'il y ait ou non coopération dépend de la présence ou de l'absence d'un but collectif. Même affaibli, puisque ne résultant pas d'une construction collective, il s'agit donc d'un concept nécessaire.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

Dans la plupart des cas, un but collectif n'a pas besoin d'être défini : l'information qu'une manifestation a lieu fournit automatiquement ce but. Par contre, dans le cas de jeux expérimentaux, la situation est plus complexe. Les agents y font face à un jeu impliquant des gains matériels individuels interdépendants. Comment un but collectif peut-il alors être défini? La plupart des définitions existantes, extrêmement simples, considèrent comme un but collectif une combinaison d'actions offrant un bénéfice mutuel par rapport à une combinaison d'action individuellement préférables. La vérification de cette intuition et encore plus l'élaboration d'une caractérisation plus précise sortent du cadre de cet article. Nous nous contenterons ici de mentionner trois propriétés élémentaires qu'un but collectif émergeant d'un jeu doit posséder. Il doit d'abord être véritablement collectif, c'est-à-dire inatteignable par un agent seul. Il doit être atteint simultanément pour tous les membres du groupe concerné : s'il s'agit d'un bénéfice mutuel, aucun agent ne doit en être exclu. Enfin, tout agent agissant dans l'intérêt du groupe doit préférer toute situation où le but est atteint à toute situation où il ne l'est pas : le but d'un groupe définit les préférences des agents souhaitant aider ce groupe. Ces trois conditions sont toutefois minimales. Elles serviront ultérieurement à définir un but collectif en un sens faible, mais n'ont pas de conséquence directe sur la nature de la coopération faible : toute autre notion de but collectif émergent conviendrait également. C'est pourquoi leur discussion est indépendante de notre démarche.

Passons à la connaissance commune. Il est évident cette fois que la notion est nécessaire, mais sa modification est plus compliquée. Peut-on le faire sans la rendre trop faible pour permettre la coopération, et comment procéder? Dans les cas qui nous intéressent, certaines informations d'arrière-plan sont de croyance commune (la structure de la situation : participants potentiels, liste des actions disponibles…) entre les participants, alors que ceux-ci ne connaissent pas leur identité.

La connaissance commune classique est ordinairement définie comme suit : un énoncé p est dit de connaissance commune au sein d'un groupe S si :

i) p est vrai;
ii) tout individu i dans S sait que p;
iii) pour tout individu i et j de S, alors i sait que si j sait que p; et ainsi de suite[8].

On obtient une définition de la croyance commune en remplaçant le savoir par la croyance, et en retirant la condition (i) si l'on souhaite autoriser des croyances erronées. Or, dans un article récent visant entre autres à expliquer le type de raisonnement coopératif tenu par des agents en situation d'incertitude, Gold et Sugden (2007) ont introduit la notion affaiblie suivante : un énoncé p est dit de connaissance commune T-conditionnelle au sein d'un groupe S (avec T sous-ensemble de S) si :

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

i) p est vrai;
ii) pour tout individu i de S, si i est un membre de T, alors i sait que p;
iii) pour tout individu i et j de S, si i est un membre de T, alors i sait que si j est un membre de T, j sait que p; et ainsi de suite.

Autrement dit, il s'agit d'une notion de connaissance commune relative aux membres d'un groupe, qui en raison de la nature de ce groupe ont des informations sur ce que les autres membres savent sans avoir besoin de connaître leur identité. Chacun, sans aucune information directe sur la connaissance d'autrui, sait toutefois ce que saura quiconque vérifie certaines propriétés. C'est précisément de cette notion que nous avons besoin.

On pourrait faire une autre objection à l'utilisation de la connaissance commune, même conditionnelle, pour définir la coopération minimale. La connaissance commune apparaît le plus souvent à la suite d'un événement public, auquel tous les agents assistent. Cependant, dans les cas de manifestation ou de foule éclair, il sera courant que des individus perçoivent la situation à partir de sources d'information différentes. Même une publicité télévisée n'est pas aussi publique qu'un discours devant un auditoire, puisque les téléspectateurs ne se perçoivent pas mutuellement. Dans le cas de tracts ou de messages d'information variés, la situation est encore pire puisque c'est l'unicité de l'événement d'origine qui disparaît. On doit alors définir le concept de *connaissance commune indirecte*. Alors que la connaissance commune classique provient d'un événement *public*, la connaissance commune indirecte est issue d'un ensemble d'événements *typiques* et permet de décrire le statut de la masse considérable d'informations qui peuvent être de connaissance commune entre des agents ne s'étant même jamais rencontrés. Or, l'analyse de la connaissance commune indirecte est étonnamment proche de celle de la connaissance commune classique et s'obtient au prix de modifications mineures (cf. Cubitt et Sugden, 2003). Ces deux notions sont de proches parentes, et leur différence n'est pas de nature épistémique mais concerne simplement le rapport entre l'origine des sources des connaissances des agents : un même événement suffisamment public, ou des événements disjoints mais suffisamment similaires. La publicité, prise en un sens contemporain, mêle d'ailleurs ces deux aspects : la publicité d'un produit ou d'une marque, par exemple, n'est généralement pas assurée par une affiche, ou un spot unique dans des conditions exceptionnelles de rassemblement, mais par une série d'affiches et de spots, présents sur des supports multiples et répétés dans le temps; autrement dit, sur des événements similaires distribués[9]. La connaissance commune indirecte est donc encore de la connaissance commune, bien qu'affaiblie.

Similairement, la coopération minimale, comprenant les phénomènes comme les manifestations ou les foules éclair, reste de la coopération. Certes, les agents n'ont alors aucune information directe sur les autres participants, pas de dialogue ni même d'échange de regard préalable; et leur but n'a pas été choisi à l'issue d'une procédure commune. Leurs sources ont été indirectes,

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

et le degré apparent de publicité (au sens large) de ces sources y joue un rôle crucial. Les analyses précédentes montrent que qualifier notre concept minimal de coopération ne dépend pas que d'une intuition subjective de ce qu'est la coopération mais est justifié par sa proximité conceptuelle avec ses versions plus traditionnelles. Retirer d'une définition de la coopération la connaissance directe des autres participants est certes un frein à l'efficacité que le groupe pourrait atteindre, mais ne modifie pas radicalement la nature de la coopération lorsqu'elle réussit. Les croyances et intentions des agents sont simplement *fondées* de façon différente — sur la base d'indices indirects. On peut avoir coopération même lorsque l'interaction physique préalable est nulle et que la communauté résulte d'événements distribués, éparpillés au cours de l'existence des uns et des autres, mais suffisamment évidents ou publics pour qu'ils produisent la croyance que d'autres ont eu la même expérience.

## 3. Intention de groupe et théorie des jeux

### *3.1. Description*

Les cas des buts collectifs et de la connaissance commune étant résolus, reste à obtenir une définition de l'intention de groupe dans un cas de coopération minimale. Pour ce faire, nous commençons par montrer la similitude entre une intention de groupe et un équilibre en théorie des jeux.

La notion formelle de jeu vise à analyser les interactions entre plusieurs individus. Un jeu consiste en la donnée d'un ensemble d'agents, pour chaque agent d'un ensemble d'actions et d'une relation de préférence ordonnant l'ensemble des combinaisons d'actions possibles — le plus souvent obtenue par l'intermédiaire d'une fonction d'utilité associant à chaque combinaison une valeur réelle. Cette base peut être raffinée de nombreuses façons par l'ajout de paramètres divers. Les concepts privilégiés en théorie des jeux sont les équilibres, qui ont pour but de déterminer quels choix d'actions sont rationnels; la rationalité devant ici s'entendre au sens du respect du principe de maximisation de l'utilité espérée. Un équilibre est la donnée d'un ensemble de stratégies (une pour chaque agent) optimales en un certain sens, c'est-à-dire telles qu'elles permettent à chaque agent de maximiser son utilité à partir de son estimation des stratégies potentielles d'autrui et de ses croyances sur l'état du monde. Le besoin d'une telle notion est évident : dans une situation d'interaction, le choix d'action de A dépend de celui qu'il croit que B fera, qui dépend à son tour de son estimation du comportement de A… si bien que les choix des agents doivent au minimum être fondés mutuellement quand existe une configuration stable de stratégies. Les stratégies composant un équilibre sont donc optimales à la fois mutuellement et par rapport aux données sur le monde extérieur.

De nombreux types d'équilibres sont possibles en fonction de la façon dont est effectué le calcul d'utilité espérée (et donc des paramètres fondamentaux du jeu). Ainsi, il est en général possible de justifier rationnellement un choix d'action dans une situation donnée en la modélisant par un jeu adéquat et en

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

montrant qu'il existe un équilibre dont l'action fait partie. Une telle démarche a l'avantage d'être conceptuellement claire : le mode de calcul de l'utilité espérée correspond à l'aspect rationnel du comportement, tandis que les éléments *a priori* sur lequel il se fonde figurent explicitement dans les paramètres donnés avec la structure même du jeu. Au sein d'un équilibre donné, les choix et les diverses raisons d'agir des individus peuvent donc être identifiés.

Bien que l'analyse précédente soit tout à fait orthodoxe, peu de travaux ont cependant noté la similitude conceptuelle qui existe entre les notions d'équilibres de théorie des jeux et les définitions analytiques de la coopération. Cette ressemblance repose sur des recherches qui ont battu leur plein au cours de la décennie 1985–1995 et qui ont montré qu'il était possible de trouver des ensembles de conditions épistémiques (portant sur le savoir et la croyance) des agents qui, si elles étaient réalisées, garantissent qu'ils agiraient selon un équilibre classique de théorie des jeux[10]. Ces résultats ne doivent pas être confondus avec leur réciproque : il ne saurait être question de prétendre que tout groupe d'individus agissant selon un équilibre de Nash est nécessairement dans un et un seul état épistémique déterminé d'avance. C'est au contraire un état épistémique particulier qui peut impliquer un comportement conforme à un concept d'équilibre.

Prenons un exemple : un ensemble d'actions données constitue un équilibre de Nash lorsque, les actions d'autrui étant fixées, tout agent changeant unilatéralement d'action verrait son utilité espérée diminuer. Un équilibre de Nash est par définition stable puisque aucun agent n'a intérêt à dévier de son action prescrite[11]. Or, un résultat technique établit que lorsque la structure du jeu et la rationalité des agents sont de connaissance partagée (connues de chacun), et que leur choix d'action est de connaissance commune (ou public), l'ensemble de leurs actions constitue un équilibre de Nash (Aumann et Brandenburger, 1995). Autrement dit, ces conditions épistémiques impliquent que : chaque agent a l'intention d'accomplir sa part de l'équilibre (choix de l'action), croit que les autres vont accomplir la leur (connaissance commune des choix), accomplit sa part en raison de cette croyance (rationalité), sait que les autres font de même (connaissance de la rationalité d'autrui).

Les définitions de concepts liés à la coopération et les caractérisations épistémiques des équilibres peuvent donc être proches, et sont exprimées dans des termes compatibles : les choix d'action correspondent aux intentions, les calculs renvoient aux raisons d'agir… et une partie de ces conditions doit être de connaissance commune. Autrement dit, il existe entre elles un certain degré de traductibilité. En particulier, une intention collective est fondamentalement stable : les intentions individuelles des agents doivent, au moins partiellement, se justifier mutuellement.

En principe, pour une définition de l'intention collective et l'ensemble de conditions épistémique d'un équilibre donné, trois cas de figure sont possibles. Si la première implique la seconde, avoir une intention collective implique que l'on agit selon l'équilibre. Si la seconde implique la première, alors l'état

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

garantissant l'équilibre implique la présence d'une intention collective. Enfin, il est possible que les deux ensembles de conditions n'aient aucun lien logique. En réalité, seule la première possibilité semble plausible.

Reprenons la définition déjà donnée de Tuomela et Miller (1988). L'intention collective de faire A (composée de A1 et A2) est définie par les conditions suivantes : j'ai l'intention de faire A1, tu as l'intention de faire A2, et nous croyons tous deux qu'il est de croyance commune entre nous que nous ferons respectivement A1 et A2. Autrement dit, les choix d'action sont de croyance commune. Les conditions définissant une intention collective sont ici un sous-ensemble des conditions épistémiques d'un équilibre de Nash, ce qui a plusieurs conséquences immédiates. D'abord, si l'action A correspond à un équilibre de Nash, alors elle *peut* être l'objet d'une intention collective donnée. Ensuite, si un groupe d'agents possède nécessairement une intention collective, il est nécessairement dans un état épistémique qui garantit qu'il agira selon un équilibre de Nash.
Prenons une autre définition analytique de la coopération, celle de l'«agir ensemble» de Tuomela (2000) :

1) J'ai l'intention que nous accomplissions X ensemble, et j'accomplis ma part de X partiellement en raison de cette intention
2) Je crois que tu feras ta part de X.
3) (1) en partie en raison de (2).
4) Connaissance commune de (1), (2) et (3).

La structure est ici extrêmement proche de celle des conditions épistémiques de l'équilibre de Nash. La condition 4, plus forte que les conditions épistémiques précédentes, a notamment pour conséquence que tout groupe d'agents agissant ensemble au sens de la définition précédente agit nécessairement selon un équilibre de Nash. Il faut s'attendre à ce que cette propriété soit générale : puisque la majorité des définitions alternatives sont une sophistication de la définition de Tuomela et Miller (1988), qui elle-même implique l'action selon un équilibre de Nash, alors elles auront cette même conséquence.

Notons cependant que la ressemblance entre définitions et équilibres ne vaut que *selon une certaine interprétation*. Dans l'«agir ensemble» de Tuomela, l'intention d'un agent cause ou justifie partiellement l'intention de l'autre, mais selon un processus non spécifié. Dans un cadre de théorie des jeux, on peut décrire ce processus comme suit : ma structure de préférences est telle qu'étant donné qu'autrui a telle intention d'agir, il est préférable que j'agisse de telle façon (et autrui est dans une situation similaire vis-à-vis de mon intention). On pourrait toutefois imaginer un mécanisme cognitif détaché de toute rationalité de maximisation remplissant la même tâche, par exemple un mécanisme d'imitation qui pousserait à agir similairement à l'action probable d'autrui (indépendamment de l'intérêt). Les définitions spécifient trop peu précisément la nature des processus causaux impliqués pour que l'on puisse les faire correspondre exactement à un type d'explication donné.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

L'intérêt du rapprochement entre définitions et équilibres est autre. On différencie usuellement la définition de la coopération, intéressant les philosophes, de son explication, du ressort de la théorie des jeux ou de l'évolution, voire de la psychologie sociale. Idéalement, c'est le comportement observable de coopération qui demande explication, tandis que c'est le concept intuitif qui demande à être défini. Les observations révèlent que ces deux projets ne sont pas aussi grossièrement séparés. Plus précisément, l'explication rationnelle du comportement coopératif et sa définition sont liées en raison du simple fait que dans une intention collective, les agents doivent avoir certaines «bonnes» raisons d'agir qui se justifient mutuellement. Une partie au moins de ces bonnes raisons doit être compatible avec l'explication de la coopération (dans un contexte donné). En conséquence, l'obtention d'une explication rationnelle de la coopération dans un type donné de situation est susceptible de conduire à une définition appropriée d'une intention collective.

### *3.2. Explication*

Tournons-nous maintenant vers les explications du comportement coopératif dans les cas minimaux décrits dans la section 2 ci-dessus. Une fois la bonne explication sélectionnée, nous pourrons élaborer un concept adapté d'intention de groupe.

La théorie des jeux est depuis longtemps le cadre privilégié pour la recherche d'une explication rationnelle au comportement coopératif. Le problème de la coopération est généralement illustré à travers le fameux dilemme du prisonnier, dans lequel les agents ont une stratégie dominante (ses conséquences sont dans tous les cas les meilleures) qui, si elle est choisie par tous, les prive d'un bénéfice mutuel. En d'autres termes, la stratégie de coopération, visant à obtenir ce bénéfice, y est individuellement non optimale, et donc risquée puisque à la merci de tout agent qui préférerait maximiser son intérêt individuel : coopérer est donc individuellement irrationnel en un sens fort. Le problème est alors d'expliquer pourquoi, comme l'attestent de nombreuses expériences, des individus coopèrent pourtant fréquemment dans de telles situations. L'essentiel des solutions existantes a d'abord été obtenu dans un cadre de jeux répétés, par des équilibres impliquant des mécanismes de punition des non coopérants ou de réputation[12].

La théorie des jeux classique éprouve quelques problèmes à traiter le cas des jeux ponctuels, anonymes et sans communication — c'est-à-dire précisément le type de cas correspondant à la coopération minimale qui nous intéresse. Dans le dilemme du prisonnier, le paradoxe s'exprime par le fait que l'ensemble de stratégies non coopératives est le seul équilibre de Nash classique. En suivant les conclusions de la section précédente, cela implique donc qu'il existe des situations où des agents ayant une intention de groupe, puisqu'ils doivent agir selon un équilibre de Nash, ne coopéreront jamais. Cela ne pose toutefois pas de problème aux théoriciens des actions conjointes, qui soit passent sous silence la question de la rationalité de la coopération, soit considèrent comme Tuomela que les utilités des agents ne sont pas identiques aux gains du

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

jeu — autrement dit, que la situation est un dilemme du prisonnier par rapport aux gains matériels des agents, mais que ces derniers raisonnent par rapport à des paramètres supplémentaires qui modifient la nature du jeu, le transformant en un dilemme coopératif plus faible dans lequel la coopération mutuelle devient un équilibre de Nash.

Le pouvoir explicatif d'une telle solution est cependant limité. En effet, pour qu'une intention de groupe justifie le choix d'une stratégie coopérative qui est l'équilibre de Nash d'un jeu modifié, il est nécessaire de justifier que ce jeu est bien de connaissance commune parmi les agents. Autrement dit, il faut montrer que le fait que les agents jouent subjectivement à un jeu particulier différent du jeu apparent est dû à un mécanisme psychologique qui est *identique* pour tous les agents. Or, des expériences[13] montrent qu'il existe plusieurs types de joueurs (plusieurs utilités individuelles finales peuvent découler des gains matériels) : coopérateurs (maximisant la somme des gains de chacun), compétiteurs (maximisant leur différence), individualistes (maximisant leur gain) ou altruistes (maximisant le gain d'autrui). En l'absence de toute information sur autrui et sans possibilité de communication, un agent ne peut donc savoir de quels types sont les autres.

Cet obstacle est surmonté par le recours aux équilibres de Nash bayésiens. Dans un cadre bayésien, les agents ont différents types possibles, régis par une distribution de probabilités *a priori*; un type est simplement défini par une certaine fonction d'utilité. Alors que dans un équilibre de Nash classique, chaque agent choisit la meilleure stratégie étant données les stratégies d'autrui, dans un équilibre de Nash bayésien chaque type d'agent choisit une stratégie optimale en fonction des stratégies des autres types et des probabilités pour que les agents soient de chaque type. On peut alors montrer alors que dans un dilemme du prisonnier à deux types d'agents (individualiste ou coopérateur), il existe un équilibre où la coopération mutuelle est la stratégie choisie par les coopérateurs dès que la probabilité qu'un agent soit coopérateur est assez grande. Autrement dit, l'introduction d'une incertitude quantifiée sur les types possibles d'agents permet de rationaliser la coopération, en conciliant comportement coopératif et équilibre de Nash bayésien.

Les équilibres bayesiens ne sont pas qu'un outil abstrait, car ils peuvent être interprétés de façon à recevoir une justification empirique. De nombreuses expériences[14] montrent que le comportement coopératif est favorisé par le phénomène psychologique d'*identification de groupe*, qui désigne la faculté pour un individu de se percevoir en tant que membre d'un groupe et non en tant qu'individu isolé. Les individus sont ainsi plus nombreux à coopérer après qu'on a rendu manifeste leur appartenance au groupe (les façons de faire sont nombreuses; l'expérience significative de Brewer et Gardner (1996) montre par exemple que la présence, dans un texte lu par les agents, de pronoms personnels «nous» et «vous» suffit à favoriser le comportement coopératif). Il est notoirement difficile de décrire exactement ce qu'est l'identification de groupe et surtout quelles sont ses conséquences : adoption de préférences sociales,

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

biais envers les membres du groupe, tendance à les imiter… Il existe cependant une interprétation minimale qui a l'avantage d'être compatible avec un cadre bayésien. On peut caractériser l'identification d'un agent par le fait qu'elle entraîne l'adoption d'un but collectif ou de préférences de groupe. S'identifier à un groupe conduirait donc au minimum à agir dans l'intérêt de ce groupe, ou en termes de théorie des jeux, à adopter le type de coopérateur (défini plus haut) relativement à ce groupe. L'identification de groupe est donc compatible avec l'explication rationnelle de la coopération dans les jeux ponctuels, puisqu'elle représente le versant psychologique d'un équilibre bayésien. Ces derniers reçoivent donc une légitimité supplémentaire à inspirer la définition d'une intention de groupe pour des agents humains.

Le détour par l'identification de groupe a une conséquence supplémentaire. Dans les expériences psychologiques comme dans les équilibres bayésiens, les raisons pour lesquelles un agent s'identifie à un groupe plutôt qu'à un autre (ou est d'un certain type plutôt que d'un autre) restent inexpliquées; seule leur force relative est, dans le second cas, exprimée par le truchement de probabilités. Or, les expériences précédemment citées révèlent que, plus que l'identification de groupe, c'est la modification de la *perception* des groupes qui joue un rôle dans le comportement coopératif. Chez Tajfel et al. (1971), Brewer et Gardner (1996) et d'autres, l'effet des expériences est de rendre un groupe saillant pour l'ensemble des agents; en conséquence de quoi il est naturel que davantage d'agents s'y identifient. Ceci a pour conséquence qu'une explication fondée sur l'identification de groupe doit se passer de l'hypothèse de connaissance commune de l'ensemble des groupes possibles, ou des types. Dans un équilibre de Nash bayésien, l'ensemble des types possibles est une donnée *a priori*, un paramètre du jeu qui est de connaissance commune. Il faut donc raffiner ce modèle en distinguant les types possibles (connus du seul modélisateur) des types effectivement perçus par chaque agent, qui peuvent être moins nombreux. Formellement, on devrait donc ajouter aux probabilités d'identification à un groupe une probabilité de perception de groupe. Nous ne chercherons cependant pas à construire un tel modèle; il nous suffit ici de comprendre les éléments sur lesquels les agents fondent leurs décisions. La différence conceptuelle entre perception et identification de groupe se révélera utile dans la prochaine section, lors de l'élaboration de notre définition.

Enfin, en ce qui concerne la définition d'une intention de groupe, l'analyse des équilibres bayésiens implique que les intentions des agents n'ont pas besoin de faire référence aux intentions explicites, mais simplement aux propensions qu'a autrui de former des intentions — ce qui dépendra donc, de son type, du groupe auquel il s'identifie, ou de l'objectif collectif qu'il adopte; ces dernières formulations étant considérées comme équivalentes. Pour être formée, l'intention d'agir doit être stable avec l'ensemble des intentions possibles d'autrui, ce qui dépend cette fois de paramètres de la situation comme les valeurs des propensions d'identification et le degré de saillance des groupes possibles. Par rapport aux définitions classiques, les intentions de groupe

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

gagnent une caractéristique essentielle de stabilité mais voient disparaître le lien d'imbrication direct entre les intentions individuelles.

## 4. La coopération minimale

### *4.1. Le statut de la rationalité*

Avant de construire nos définitions, nous nous arrêtons un instant sur le rôle qu'y joue la rationalité. Dans les analyses existantes de la coopération, la question de la rationalité coopérative — et de l'explication de la coopération en général — est souvent ignorée : définition et explication se développent indépendamment. Et quand elle ne l'est pas, comme chez Tuomela, la rationalité n'est qu'une propriété supplémentaire, possible mais non essentielle, d'une action coopérative. Dans les définitions par emboîtement, une fois définie la coopération, la coopération rationnelle est obtenue par le simple ajout de certaines clauses supplémentaires. Celles-ci peuvent soit faire référence aux préférences des agents en assurant que leurs choix d'actions sont cohérents avec elles, soit ajouter des conditions de connaissance commune, assurant que la coopération des agents et leur contribution mutuelle à leurs buts sont bien publics.

La rationalité joue au contraire un rôle essentiel dans la coopération minimale. La référence à des prises de décision rationnelles dans notre définition de la coopération se justifie de la façon suivante. Dans la «pleine» coopération de Tuomela, les buts collectifs sont collectivement construits : il est simple pour les agents de se concentrer sur un même objectif puisqu'ils l'ont évoqué ensemble. En l'absence d'interaction préalable, la présence de raisons apparentes d'agir conformément à un objectif devient cruciale pour estimer le degré auquel il va être adopté. Au départ, un agent percevant l'objectif sait seulement qu'il est susceptible d'être perçu par autrui en raison de sa saillance. Les raisons d'agir fondées sur le calcul d'utilité espérée prennent donc une ampleur considérable puisqu'elles constituent les seuls indices supplémentaires permettant d'estimer l'attractivité de l'objectif. Autrement dit, s'il n'existe aucun équilibre rationnel dans lequel l'objectif est réalisable, les raisons de chercher à l'accomplir faiblissent de façon dramatique. Dans la coopération idéale, les agents peuvent se communiquer promesses, accords ou plans d'action préalables, s'influencer et favoriser l'identification au groupe de façons multiples, et tout simplement avoir des interactions. Dans la coopération minimale, les buts collectifs sont tout juste saillants, les agents sont seuls et en proie à l'incertitude généralisée sur l'ensemble de la situation. Aussi est-il naturel que les raisons d'agir issues de la pensée rationnelle — que chacun peut élaborer et dont chacun sait qu'autrui le peut également — y jouent un rôle prépondérant.

Présenté de la sorte, l'argument reste cependant insuffisant : même si les considérations rationnelles constituent de bonnes raisons d'agir dans certains cas, elles restent une simple modalité de l'apparition de telles raisons d'agir. Or ce sont ces dernières qui doivent figurer dans une définition, de façon à

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

indiquer les liens de justification entre états mentaux; mais la nature de ces raisons d'agir n'a pas à être imposée *a priori*. Le véritable argument est en fait inverse : s'il n'était pas rationnel de coopérer dans une situation de forte incertitude, alors les agents auraient des raisons fortes de ne pas agir de la sorte — et surtout de penser qu'autrui ne va pas agir de la sorte. Autrement dit, sans une justification rationnelle de la coopération minimale, celle-ci serait impossible. Le réquisit de rationalité de la coopération doit donc être vu comme la garantie de l'*absence de raisons déterminantes de ne pas coopérer*. Il est précisément nécessaire à la coopération minimale, puisqu'en l'absence de lien direct entre agents, ceux-ci ne peuvent tirer d'autrui aucune bonne raison de coopérer. La présence d'une raison de ne pas coopérer serait donc dirimante.

La référence à des équilibres rationnels dans une définition pose cependant quelques problèmes. Il n'est évidemment pas possible de spécifier dans une définition informelle tout le détail des calculs de maximisation que les agents sont censés accomplir. Les définitions classiques se contentent par exemple de mentionner qu'une intention est formée partiellement en raison d'une autre intention, sans stipuler si cette raison est fondée sur un calcul ni quels sont les autres paramètres pertinents éventuels. Toutefois, dans notre cas, puisque l'identification de groupe joue un rôle unificateur crucial, rassemblant buts collectifs et explication rationnelle, il paraît impossible de ne pas y faire référence dans la définition. Cela sera assuré en spécifiant le paramètre formel majeur qui le représente, à savoir les probabilités (ou propensions) de perception et d'identification de groupe.

Cependant, l'impossibilité de traduire informellement les calculs a également des conséquences souhaitables. D'une part, elle permet de ne pas exiger de capacités cognitives excessives chez les agents. La coopération minimale convient à la description d'actions de masse dans lesquelles les agents impliqués sont nombreux; aussi est-il peu plausible qu'ils envisagent la faisabilité d'une action coopérative par l'intermédiaire de calculs probabilistes complexes. Dans de tels cas, il est envisageable qu'ils utilisent des raccourcis, tel celui d'assimiler la probabilité de s'identifier à une équipe à la proportion des agents qui s'y identifieront certainement. Raisonner directement à partir d'estimations du nombre d'agents participants est à l'évidence un moyen simple de prévoir le degré de succès ou d'échec d'une action coopérative. Il est donc souhaitable que le raisonnement des agents imposé dans les définitions soit décrit de façon assez lâche pour autoriser de tels raccourcis de calcul.

D'autre part, le fait que les agents ne perçoivent pas nécessairement les mêmes groupes semble diminuer les chances qu'ils coopèrent. Supposons que des agents s'identifiant à un groupe décident rationnellement de participer à l'action coopérative, mais sur la base de calculs différents — s'ils n'envisagent pas les mêmes identifications possibles pour autrui. Veut-on encore parler de coopération dans un tel cas? Nous pensons que oui. Il est suffisant que les agents perçoivent le groupe dont il s'agit et les groupes dégénérés représentant les individus isolés, ce qui est toujours le cas. Dans un contexte d'incertitude

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

extrême, imposer que les participants à une véritable action coopérative aient tous exactement la même perception de la situation reviendrait à inutilement décimer le nombre de cas légitimes de coopération. Il est suffisant que les agents s'identifient au bon groupe, perçoivent au moins un autre groupe possible pour autrui, et décident ensuite de participer. Ici encore, la référence précise au calcul que doivent mener les agents ne ferait que nuire à la coopération en en faisant un concept coupé de la réalité.

Les analyses précédentes peuvent désormais être rassemblées pour former l'ossature d'une définition de la coopération minimale — se manifestant lorsque des individus inconnus, isolés interagissent en une unique occasion. Lorsqu'ils sont soumis dans ces conditions à un problème coopératif représenté par un jeu, le processus menant certains d'entre eux à la coopération passera les étapes théoriques suivantes :

— Perception d'une partie des buts collectifs de la situation.
— Détermination des groupes auxquels les agents pourront s'identifier.
— Estimation par tous des propensions de chacun à percevoir ces groupes et à s'identifier à certains — ou éventuellement du nombre de membres potentiels de chaque groupe.
— Vérification qu'il existe un équilibre dans lequel chaque type de chaque agent préfère accomplir sa part dans l'objectif collectif de son groupe.
— Adoption par chacun d'un objectif (c'est-à-dire identification de chacun à un groupe donné).
— Choix par chacun, pour son action individuelle, de sa part dans l'action collective du groupe auquel il s'est identifié.

### *4.2. Définition*

Nous pouvons maintenant construire notre définition d'une action coopérative minimale, par une démarche similaire à celle de Tuomela, c'est-à-dire par empilement des définitions élémentaires d'un objectif collectif et d'une intention de groupe, auxquelles nous substituons les concepts affaiblis précédemment introduits. Remarquons que le recours à la connaissance commune indirecte oblige à la vigilance quant aux sous-groupes concernés. Dans la suite, nous distinguerons trois catégories d'agents : l'ensemble G des bénéficiaires de l'accomplissement de l'objectif, dont on suppose qu'ils peuvent tous prendre part à sa réalisation; l'ensemble P des agents ayant perçu l'objectif collectif; et l'ensemble C des agents adoptant l'objectif de G (ou s'identifiant à G).

(BCp) B est un *but collectif potentiel* pour un groupe G dans une situation S si et seulement si G est tel que :

a) B est réalisable par les membres de G et irréalisable par tout membre agissant seul.
b) Si B est réalisé, B est simultanément satisfait pour tous les membres de G.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

c) La réalisation de B est préférable à sa non-réalisation pour tout membre de G agissant dans l'intérêt du groupe.
d) Si B n'a pas été rendu saillant antérieurement à la situation S, alors la structure des préférences individuelles des agents manifeste un objectif collectif émergent (elle offre l'opportunité d'une bénéfice mutuel).
e) (a), (b) et (c) sont de croyance commune P-conditionnelle au sein de G, où P est l'ensemble des membres de G ayant perçu l'existence de B défini par les conditions (a) et (b).

Les conditions (a), (b), et (c) décrivent les propriétés nécessaires d'un but collectif, qu'il émerge du jeu ou soit préexistant : (a) est triviale, (b) stipule qu'il est de nature collective, et (c) indique la nature de la préférence collective qu'il induit. La condition (d) fait référence à la structure des préférences individuelles qui est nécessaire pour que B puisse émerger. La définition est telle qu'à cette condition pourrait être substituée toute analyse de ce qu'est un objectif collectif émergent. En l'absence d'une telle analyse, notre définition reste donc incomplète et ne prend pas en compte avec précision les cas de coopération ponctuelle dans un cadre de théorie expérimentale des jeux. Toutefois, lorsque B est déduit du jeu, (d) devrait être telle que les conditions (a) à (c) en soient des conséquences. Nous devons néanmoins les mentionner explicitement pour les cas dans lesquels le but est préexistant. Enfin, la condition (e) peut être «vide», au sens où elle n'impose pas que le but ait effectivement été perçu par quiconque. Cependant, si un membre le perçoit, il sait que le but est perceptible par autrui : le but est de nature publique, et ses caractéristiques seront accessibles à tout membre le percevant.

La définition précédente décrit donc simplement le fait qu'une situation contient un but collectif perceptible. Elle se situe entièrement au niveau des présuppositions de la coopération : même si la condition (e) fait référence à des états épistémiques, elle n'en impose aucun et ne fait qu'exprimer les conséquences de la situation sur un agent qui serait dans un certain état épistémique.

Notons qu'on pourrait au besoin obtenir une version plus objective d'un but potentiel en ajoutant une condition (f) stipulant qu'il est *effectivement* réalisable, c'est-à-dire les agents l'ayant *perçu* sont suffisamment nombreux pour l'accomplir (même si rien ne garantit qu'ils le feraient). Un tel niveau de détail n'est cependant pas nécessaire à notre analyse.

(BCd) B est un *but collectif distribué* pour un sous-groupe C de G si :

A) B est un but collectif potentiel pour G.
B) Tout membre de C a B pour but dans S et a l'intention de contribuer à la réalisation de G.
C) L'utilité espérée de tout membre de C est maximale pour l'accomplissement de sa part de B étant donné les propensions de tout membre de G à percevoir les différents buts présents en S et à s'identifier aux différentes équipes possibles.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

D) (A) et (C) sont de croyance commune P-conditionnelle au sein de G, où P est l'ensemble des membres de G ayant perçu l'existence de B.
E) (B) en raison de (C) et de (D).

Un but collectif est un but potentiel qui a été perçu et définitivement adopté par un groupe d'agents. Comme on l'a vu précédemment, son adoption par un agent exige deux conditions : qu'il s'identifie au groupe et que le but soit estimé réalisable en fonction des propensions d'autrui à s'identifier au groupe, c'est-à-dire qu'il doit exister une raison *rationnelle* de l'adopter. L'étude des équilibres bayésiens montre qu'agir dans l'intérêt du groupe n'impliquait pas forcément de choisir une action coopérative, l'estimation des propensions d'autrui étant cruciale dans la prise de décision. Si le but n'était compatible avec aucun équilibre, il serait considéré comme irréalisable et donc abandonné. Les conditions (C), (D), et (E) représentent donc simplement la traduction informelle des caractéristiques d'un équilibre bayésien fondé sur les propensions des agents à s'identifier à un groupe.

(ICf) Les agents d'un groupe C ont une *intention collective faible* d'accomplir B si et seulement si :

i) Chaque membre de C a B pour but et a l'intention d'accomplir sa part de B.
ii) L'utilité espérée de tout membre de C est maximale pour l'accomplissement de sa part de B étant donné les propensions de tout membre de G à percevoir et à adopter B comme but.
iii) Chaque membre de C croit que les autres membres de C ont une certaine propension à agir selon B.
iv) (ii) et (iii) sont de croyance commune G-conditionnelle pour un groupe G contenant C.
v) (i) en partie en raison de (ii), (iii) et (iv).

Avoir une *intention collective faible* revient à agir selon un équilibre de groupe, pour une raison commune : le but collectif distribué se comprend comme la superposition d'une intention collective faible et d'un but collectif potentiel. Les conditions sont donc semblables à celles de la définition précédente, à la différence que B n'a plus besoin d'avoir une structure préalable particulière (toute combinaison d'actions peut convenir) et que la raison d'agir peut désormais être quelconque. Des agents pourraient agir ensemble dans l'intérêt d'un bénéficiaire quelconque, ou même selon un principe ou une idéologie particulière. La nouvelle condition (C) est nécessaire pour assurer que la raison d'agir soit de nature publique (ce qui était auparavant automatiquement garanti au moyen de la définition d'un but collectif potentiel). (D) assure la croyance commune conditionnelle pour un groupe G quelconque : il suffit que le but B soit saillant au moins pour les membres de C.

Notons que dans l'ensemble des définitions précédentes, rien n'assure que les membres des groupes C soient bien capables de réaliser le but qu'ils

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

poursuivent : il suffit qu'ils le poursuivent pour une raison commune et estiment sa réalisation suffisamment probable. On ne peut exiger davantage puisque les agents agissant pour la même raison ne savent pas nécessairement qui ils sont. Autrement dit, un groupe peut avoir un but collectif distribué, ou une intention collective faible, et encore échouer à coopérer.

La définition finale s'obtient désormais naturellement à partir des précédentes.

(Cm) Les agents d'un groupe C *coopèrent minimalement* (avec succès) pour réaliser B si et seulement si :

1) Les membres de C peuvent réaliser B.
2) Les membres de C souhaitent réaliser B ensemble.
3) B est un but collectif distribué pour C.
4) Les membres de C réalisent B ensemble conformément avec et en partie en raison de l'intention collective faible de réaliser B qu'ils ont en raison de (3).

La condition (1) assure que les membres de C sont assez nombreux pour accomplir B. (2) est également nécessaire mais demande un éclaircissement. Les agents ne se connaissant pas, comment peuvent-ils souhaiter accomplir quelque chose ensemble? Dire que les membres de C souhaitent réaliser B ensemble ne signifie rien d'autre que : chaque membre de B souhaite que B ait été réalisé par des membres qui agissaient pour le groupe, c'est-à-dire pour de bonnes raisons. Cette condition traduit le fait que si un agent apprend, après que B a été réalisé par une coopération apparente, qu'un certain nombre de participants avait en réalité simplement essayé de faire croire qu'ils étaient coopératifs (dans l'optique d'une possible interaction future), il n'y aurait pas eu coopération. Cela ne signifie pas cependant pas que la perception de la coopération par l'agent est nécessaire à la coopération, puisque seul le souhait que les choses se déroulent ainsi fait partie des conditions requises. Nous reviendrons sur ce point lors de la définition du concept naïf de coopération.

La donnée d'actions cohérentes avec le but collectif distribué suffit à assurer la coopération puisque sa définition implique celle d'une intention collective faible. Aucune condition supplémentaire stipulant une croyance commune n'est nécessaire à ce stade puisque la définition de but collectif distribué en contient déjà.

## 5. Discussion

### *5.1. Conséquences*

Un aspect important de notre définition est la prédominance du concept de but collectif, ce qui la rend tout à fait cohérente avec celle de Tuomela. La définition de la coopération minimale est ainsi obtenue par la superposition de strates conceptuellement distinctes : un but collectif potentiel, auquel s'ajoute une

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

intention collective faible, aboutit à la coopération minimale. Nos strates sont mêmes plus clairement séparées, puisqu'elles renvoient précisément à la structure tripartite de la coopération (qui avait été notre première intuition, issue de l'étude de Regan) : le but collectif potentiel se situe tout entier au niveau des présuppositions de la coopération; l'intention collective faible en concerne les prémisses; et la coopération minimale en est naturellement la réalisation.

La coopération repose cependant principalement dans le concept de but collectif distribué, qui reste bien un but. Dans les analyses classiques, un but, une fois sélectionné par le groupe, reste fixe. Ici, l'état d'incertitude généralisée dans lequel se trouvent les agents a pour conséquence qu'ils peuvent abandonner ce but en cas d'absence d'un équilibre de groupe le justifiant. Le rôle du but collectif s'étend donc dans la coopération tout entière : il déclenche l'identification, sa saillance fonde les croyances de chacun, et c'est sa confirmation ou son abandon à la suite de calculs d'utilité espérée qui détermine l'action entreprise. Le but collectif constitue pratiquement la seule information publique que peuvent recevoir les agents, et est, à ce titre, essentiel. En ce sens, notre conception de la coopération minimale peut être qualifiée de *téléologique*.

La coopération minimale n'est cependant pas une forme de coopération aussi faible qu'elle le paraît. Pour Tuomela, la coopération pleine et entière correspond à ce qu'il nomme le mode de groupe, c'est-à-dire à la construction collective d'un objectif et à des agents agissant entièrement dans l'intérêt du groupe. Notre définition est manifestement trop faible pour correspondre à ce cas de figure. À l'autre extrémité, la coopération en mode privé concerne des agents agissant pour le groupe en raison de leur intérêt individuel. Par exemple, un agent dont la fonction d'utilité est altruiste, ou correspond à l'utilité d'un groupe, sera dans ce cas. Autrement dit, le mode privé est conceptuellement proche de la simple transformation de préférences, ce qui le rapproche de notre cadre. Cependant, les définitions de concepts coopératifs en mode privé que donne Tuomela ne contiennent pas de buts collectifs mais simplement des buts individuels compatibles. Notre notion de but collectif, et donc de coopération minimale, se situe par conséquent au-delà du mode privé. Or, une troisième solution existe selon Tuomela :

> […] une personne peut être prosociale, par exemple altruiste, et coopérer avec d'autres gens sur cette base. Ce n'est pas encore une raison de groupe. D'un autre côté, elle pourrait s'efforcer de maximiser le bénéfice d'un groupe d'une sorte de façon égalitaire et juste. Il s'agit souvent d'une raison de mode privé pro-groupe. Un genre plus fort de raison de groupe est fondé sur l'identification d'une personne au groupe et à son adoption de la pleine perspective de groupe […] ainsi que sur le fait qu'elle agisse en tant que membre du groupe (Tuomela, 2007, p. 200).

Ainsi, entre l'identification totale et l'altruisme se résumant à une préférence existe une troisième voie qui reste dans la sphère du mode privé mais s'approche

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

du mode de groupe. Dans notre analyse, un agent s'identifiant au groupe n'adopte pas n'importe quelle fonction d'utilité, mais celle qui découle de l'existence d'un but collectif. La maximisation de l'utilité collective espérée dans les équilibres de groupe n'est que la traduction en langage économique de la volonté de réaliser l'objectif de groupe; c'est parce que l'agent cherche à réaliser l'objectif de groupe qu'il raisonne selon l'équilibre. De plus, cet objectif est la cause première de l'identification, et la réussite de la coopération dépend bien de sa réalisation. Nous suggérons par conséquent que la coopération minimale est une notion de mode privé pro-groupe. L'affirmation certes est difficile à infirmer en raison de la définition floue de la catégorie qu'elle concerne : Tuomela admet lui-même que la distinction entre mode de groupe et mode privé pro-groupe n'est pas toujours évidente[15]. Reste que le fait que notre concept minimal de coopération soit plus fort que le mode privé classique et partage des éléments avec le mode de groupe appuie l'idée qu'il constitue un stade *minimal* de la coopération humaine.

### *5.2. Circularité*

Les analyses usuelles de concepts collectifs rencontrent fréquemment des problèmes de circularité de deux types. D'abord, certaines définitions des intentions collectives sont massivement fondées sur des intentions conditionnelles, selon le principe suivant : l'intention collective pour un groupe G d'accomplir X équivaut à l'ensemble des intentions individuelles des membres de G d'accomplir leur part de X *si les autres font de même*. Se pose alors le problème de circularité suivant : je n'aurai l'intention d'accomplir ma part que si je crois que les autres ont une intention similaire, qu'ils ne peuvent eux-mêmes avoir que s'ils pensent que les autres l'ont, etc.[16] Un ensemble d'intentions conditionnelles intriquées ne peut justifier l'apparition d'intentions «pures» : elles constituent de bonnes raisons à cette apparition, mais ne permettent pas de la déduire logiquement. On trouve par exemple ce type de structure chez Gilbert justement, qui utilise la notion de quasi-disposition à agir, c'est-à-dire de disposition conditionnelle.

Il ne s'agit pas d'un problème purement abstrait : qu'il suffise d'imaginer deux concurrents auxquels un médiateur parvient à faire accepter un accord, stipulant qu'ils doivent chacun rendre un service à l'autre. Chaque concurrent a accepté de subordonner sa signature à la signature de l'autre, puisque s'il était seul à signer, il se trouverait désavantagé. La situation restera bloquée à moins qu'on ne puisse trouver un dispositif supplémentaire permettant de transformer les intentions conditionnelles de signer en pleines intentions. Notons qu'il ne s'agit pas d'une circularité *définitionnelle*, qui ferait référence au concept même qu'elle prétend définir, mais d'une circularité *explicative* qui rend la définition caduque. De même qu'une intention individuelle est l'état mental qui cause une action donnée, une intention collective est censée correspondre à un ensemble de conditions dont s'ensuit automatiquement la réalisation d'une action collective. Or, la circularité explicative empêche de déduire de la présence

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

de l'intention collective la présence d'intentions d'agir individuelles, et n'implique donc pas la réalisation de l'action collective.

Il n'existe pas de solution à ce problème au niveau de la définition si l'on souhaite conserver des intentions conditionnelles. Les parades habituelles consistent à analyser les façons pratiques dont des agents se tirent de ce type de difficulté : par un engagement unilatéral (décider d'agir inconditionnellement et ainsi créer les intentions d'agir chez autrui), par une promesse (faire croire qu'on va agir sans conditions) ou encore par des mécanismes cognitifs plus ou moins automatiques, comme le fait de se contenter d'un *indice* de l'intention d'agir d'autrui pour déclencher sa propre intention[17].

Nos définitions précédentes ne comprennent aucune intention conditionnelle : les agents forment leur intention d'agir à partir de la croyance *a priori* en la propension des autres à agir selon le groupe; si la propension est suffisamment élevée, l'agent agira. Il nous est donc facile d'éviter le problème de circularité explicative des intentions : puisque les agents ne pourront pas entrer directement en contact avant l'action, leur intention ne saurait être conditionnelle.

Il existe un second problème, cette fois de circularité définitionnelle, qui apparaît lorsque certaines conditions de la définition d'un concept collectif font référence à ce concept. Tollefsen (2002) remarque que de tels problèmes de circularité sont répandus dans la littérature sur la coopération ou les intentions collectives. Chez Gilbert, des individus formeront un sujet pluriel faisant X si et seulement s'ils ont formé un engagement conjoint de faire X ensemble (cf. Gilbert, 1989). Chez Tuomela, l'une des conditions nécessaires à ce que des agents accomplissent intentionnellement X ensemble est qu'ils agissent conformément à et en raison de leur intention conjointe d'accomplir X ensemble (Tuomela, 2007, p. 125 et 144). Notre propre définition de la coopération a un profil similaire, puisqu'elle contient un souhait d'agir ensemble chez les participants. Si l'on ramène l'«agir ensemble» à la coopération, le risque créé par une telle condition est de faire tomber l'ensemble de la définition dans une circularité et particulièrement dans la trivialité — celle consistant à donner le fait d'être un chat comme partie de la définition de ce qu'est un chat.

Une telle circularité fait-elle réellement obstacle à l'établissement d'une juste définition? Gilbert a répondu par la négative en expliquant que la définition analytique d'«agir ensemble» ne tombait pas dans la circularité en faisant référence à une notion naïve d'agir ensemble, celle qu'aurait l'être humain moyen et non le théoricien. Tuomela argumente de façon similaire, relevant les cas d'action conjointe élémentaire chez les animaux et les enfants de moins d'un an pour défendre l'idée que les agents n'ont pas besoin de posséder de concept sophistiqué de l'action conjointe. Ce schéma d'argumentation est conceptuellement satisfaisant. En effet, la circularité présente dans toute définition d'un concept collectif n'est pas pure : la définition ne fait pas référence au concept à définir lui-même mais la version qu'en ont les agents impliqués (par exemple dans le contenu de leurs intentions). Cependant, l'analyse reste insuffisante si l'on ne décrit pas précisément à son tour la version naïve du

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

concept collectif, sous peine d'avoir des conditions insuffisamment déterminées. Nous avons précédé cette objection en expliquant après la définition ce que signifiait «agir ensemble». La prochaine et dernière sous-section expose cette conception plus en détail.

### *5.3. Types de coopération*

Comment déterminer le concept naïf de coopération correspondant à l'usage quotidien des agents? Les définitions précédentes sont données du point de vue du théoricien omniscient qui, ayant accès à l'ensemble des comportements observables, des états mentaux des agents et de tous les processus de cause à effet, est capable de décider si telle interaction est ou non de la coopération. Or, certains de ces paramètres sont inaccessibles aux agents eux-mêmes. Le concept naïf de coopération utilisé par les agents correspondant aux cas *apparents* de coopération demande donc également une définition.

Imaginons une situation d'action simultanée de plusieurs individus. Avant l'action, aucun agent ne peut savoir s'il va y avoir coopération ou non : même s'il s'est identifié au groupe et estime que dans la situation donnée les propensions de chacun à faire de même sont assez élevées, il n'a *a priori* accès à aucune intention, ignore si le but collectif a été reconnu et si d'autres se sont identifiés au même groupe. Après l'action, le résultat de l'interaction est connu, ainsi par conséquent que les actions de chacun et rétroactivement ses intentions (si l'on exclut la possibilité d'erreurs); tous savent si le but collectif a été atteint ou non. Mais il arrivera fréquemment qu'aucun agent ne sache exactement pour quelles raisons les autres ont agi; subsiste alors un doute concernant l'identification de chacun.

En théorie des jeux, il est bien connu que dans un équilibre bayésien, une même action peut parfois être choisie par différents types d'agents. Dans un cadre dynamique, les agents peuvent actualiser leurs croyances : étant donné l'action collective accomplie, l'action optimale pour chaque type et la probabilité *a priori* pour chaque agent d'être d'un type donné, on peut calculer par la règle de Bayes la probabilité *a posteriori* qu'un agent soit d'un certain type. Autrement dit, ces probabilités *a posteriori* représentent la croyance qu'ont les agents sur les types possibles d'autrui après l'action. Selon leurs valeurs, plusieurs types de coopération sont alors possibles.

Idéalement, les agents ne devraient conclure à la coopération que lorsque les probabilités d'identification au groupe *a posteriori* sont égales à 1 pour tous ceux qui ont choisi une action coopérative — nous parlerons alors de *coopération avérée*. Ce critère est rarement rempli; or, il est relativement fréquent d'estimer qu'une action collective a été coopérative. Si un téléspectateur apprend le soir que la manifestation a été un succès, il aura tendance à y voir le résultat de la coopération réussie des militants et le signe de la cohésion du groupe qui l'organisait — alors qu'il est possible que bien des participants ne soient venus que pour des raisons individuelles, étrangères à l'effet politique de la manifestation : retrouver des amis, manger gratuitement, ou simplement sortir[18]. Plus encore,

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

un spectateur pourra parler de coopération du groupe entier quand bien même une partie seulement des militants a participé.

On peut alors distinguer deux types de cas. D'abord, lorsque les actions de chacun n'ont pu être précisément observées, le critère de perception de la coopération est simplement l'accomplissement du but collectif perçu. La même équipe de football pourra par exemple être qualifiée *a posteriori* de soudée ou de désunie sur la seule base du score d'un match. Appelons ceci la *coopération comportementale* — l'accomplissement d'un objectif collectif donné.

Lorsque les actions sont observables (c'est-à-dire que le jeu modélisant la situation est perceptible), une seconde approximation semble réaliste : au lieu d'exiger la certitude que chaque membre s'est identifié au groupe, il suffit que la probabilité de son identification soit supérieure *a posteriori* à toutes celles des autres identifications possibles. On conclurait ainsi à la coopération quand, pour chaque membre, c'est l'identification à l'équipe qui explique le mieux ses actions. Un militant participant à la manifestation pourra conclure à un événement artificiel s'il voit autour de lui les autres principalement occupés à chercher à manger, à se reposer, à danser… et au contraire estimer la manifestation comme un succès si les participants semblent réellement venus manifester. Nous parlerons alors de *coopération plausible*. Notons qu'elle coïncide avec la coopération avérée lorsque la seule explication des actions coopératives est l'identification au groupe, par exemple dans des situations simples comme un dilemme du prisonnier ponctuel. Autrement dit, juger la coopération plausible comme de la vraie coopération conduit peut être une bonne approximation. En résumé, selon la définition analytique, seule la coopération avérée est légitime. Mais pour des agents normaux, le concept pragmatique de coopération plausible, voire, en cas de données peu nombreuses, de coopération comportementale, suffisent.

Revenons maintenant à la question de la circularité. Pour qu'une situation soit coopérative, il est selon la définition nécessaire que les agents souhaitent agir ensemble. Cela ne signifie finalement rien d'autre que : les agents doivent au moins souhaiter que l'objectif collectif soit réalisé par des membres du groupe; et éventuellement qu'il soit plausible qu'ils aient agi de la sorte pour atteindre cet objectif. La définition ne place donc pas l'intention même de coopérer au cœur de la coopération. Au moment d'agir, les agents ne nécessitent que l'intention de faire leur part parce qu'ils croient que les autres membres potentiels feront la leur, et qu'ils suffiront à réaliser le but. Après l'action collective, les agents la percevront comme de la coopération s'il est suffisamment plausible que les participants ont agi pour les bonnes raisons (qu'ils se sont identifiés au bon groupe). Un individu n'a donc pas besoin d'avoir l'intention de coopérer pour coopérer, mais celle de faire sa part d'un objectif collectif et d'être mû par le désir qu'il soit accompli par des membres s'identifiant au groupe.

Cette analyse est compatible avec certains cas spéciaux. Un agent peut par exemple favoriser la coopération sans le savoir, s'il ne perçoit pas l'objectif collectif (ni le groupe associé) mais son action individuelle coïncide avec l'action coopérative.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

Un agent peut également coopérer en croyant qu'il n'y a pas coopération. S'il est par erreur persuadé qu'un certain nombre d'agents ne se sont pas identifiés au groupe mais vont (par coïncidence) accomplir une action bénéfique pour le groupe; et que sans leur aide (involontaire), l'objectif collectif ne pourrait être réalisé. Notre agent va accomplir sa part, afin que l'objectif collectif soit réalisé, tout en croyant qu'il ne s'agit pas de coopération puisque tant d'autres participants agissent pour une mauvaise raison.

Pour finir et à la lumière de ce qui précède, remarquons de nouveau le rôle central que joue le concept d'objectif collectif dans notre définition. La présence d'un objectif collectif (qu'il soit émergent ou préexistant) est d'abord une condition *a priori* de la coopération. Mais *a posteriori*, un objectif collectif est également le seul indice observable du succès d'une action coopérative — et donc du fait qu'il y a eu coopération. Cet indice peut évidemment être erroné puisqu'il ne signale que la coopération plausible[19].

## 6. Conclusion

Nous avons défini un concept de coopération minimale. Il correspond à des situations comportant un but collectif manifeste (émergeant éventuellement de la situation) dans lesquelles un groupe d'agents, choisissant tous leur action conformément à un équilibre, accomplit ce but. Les agents n'ont besoin ni de se connaître, ni de communiquer, ni même de se percevoir au moment de leur prise de décision. Le cœur de cette définition est donc le concept de but collectif, sans lequel n'existerait qu'une action collective «parallèle».

Nous avons montré qu'une telle conception permettait de régler simultanément certains problèmes philosophiques habituellement liés à la coopération : la circularité des définitions et la référence à un concept naïf de coopération non analysée. Selon notre analyse, parallèlement à la conception sophistiquée de la coopération minimale que nous avons définie, il en existe une version naturelle, ou intuitive, fondée essentiellement sur la réalisation d'un objectif collectif. Elle prend acte du fait que la réussite apparente de la coopération est souvent *a posteriori* un critère déterminant dans la croyance qu'il y a effectivement eu coopération. On parle alors de coopération comportementale, voire de coopération plausible dans les cas dynamiques. Cette conception naïve peut expliquer la facilité avec laquelle on suppose habituellement l'existence de groupes.

Notre concept de coopération minimale permet de répondre clairement à plusieurs questions précises concernant la nature de la coopération. Notons que presque toutes ces questions concernant les éléments que la coopération *doit* ou *devrait* comporter touchent bien l'aspect *nécessaire* de la définition de la coopération. Il est donc légitime d'y apporter une réponse à la lumière du concept de coopération minimale.

Y a-t-il encore coopération si quelqu'un se trouve à coopérer en s'étant trompé d'action? Non. Agir pour le groupe mais en ayant mal identifié l'objectif collectif implique qu'il n'y a pas coopération puisque l'agent n'a pas accompli sa part. Si «agir coopérativement en s'étant trompé d'action» est pris au sens

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

de l'accomplissement de sa part pour les mauvaises raisons, il n'y a pas davantage coopération puisque les raisons d'agir des participants doivent toutes être correctes (fondées sur le groupe).

En cas de coopération, les agents doivent-ils être conscients qu'ils coopèrent? Non. Ils doivent être conscients qu'ils agissent dans l'intérêt du groupe, ou afin de réaliser un objectif collectif. Sans information directe sur autrui, ils ne peuvent savoir au moment de leur décision si suffisamment d'agents sont également en train d'accomplir leur propre part, et *a fortiori* pour la bonne raison.

Les agents doivent-ils avoir un plan d'action général? Non. Tout d'abord, ils ne nécessitent aucun plan d'action préalablement choisi ensemble. Ils peuvent néanmoins agir conformément à un plan en un certain sens : ils savent que leurs actions doivent à la fois faire partie d'une combinaison d'actions pour laquelle l'objectif collectif serait réalisé, et être cohérentes avec un équilibre de groupe donné. Néanmoins, rien ne garantit à un agent que les autres suivent ce même «plan».

L'intention d'agir de chacun doit-elle être absolue ou conditionnelle à la croyance en l'intention d'agir d'autrui? Un agent n'a en aucune façon besoin de connaître l'intention d'agir d'autrui. Il ne se fonde que sur les propensions d'autrui à former telle ou telle intention. Les intentions des agents ne sont donc pas conditionnelles, au sens où elles ne deviendraient des intentions qu'après constatation de l'intention conditionnelle d'autrui.

Les raisons d'agir doivent-elles être identiques pour tous? Oui. Il est nécessaire que les agents aient accompli leur part en raison de leur identification à l'équipe. Par contre, les raisons d'agir possibles n'ont pas à être de connaissance commune, même indirecte. Certains agents peuvent n'avoir perçu qu'une partie des raisons d'agir possibles pour autrui : s'ils agissent pour le groupe, ils savent qu'autrui peut agir pour cette même raison. Peu importe alors que la liste des autres groupes possibles prise en compte par chaque agent ne soit pas la même pour tous[20].

## Notes

1 Dans la littérature, on parle parfois de manière équivalente de coaction. Nous n'utiliserons pas ce terme pour désigner une action collective, afin d'éviter la confusion avec certains auteurs qui l'emploient pour désigner une forme spécifique de coopération.
2 Qu'un résultat soit remarquable ou non rend ce concept dépendant de la cognition de l'observateur, mais ne change rien à la nature externe d'une action collective.
3 Ces termes des présuppositions, prémisses et réalisation sont repris du vocabulaire de Tuomela (2000).
4 Cela n'implique pas que les concepts collectifs ne jouent aucun rôle causal. Des intentions individuelles pourraient par exemple avoir des conséquences spécifiques seulement quand elles sont liées conformément à la définition d'une intention collective, quelle qu'elle soit.
5 Traduction du terme anglais consacré de «flash mob».
6 Autre exemple d'action accomplie par une foule éclair.
7 Depuis Rapoport, 1962.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

8 Nous ne discuterons pas ici du bien-fondé de cette définition; la modification apportée pourrait être transposée à n'importe laquelle des définitions alternatives de la connaissance commune.
9 L'exemple de Chwe d'une capsule publicitaire unique pendant un événement massivement regardé est plutôt l'exception que la règle.
10 Cf. Walliser, 2002, p. 698-699, pour un résumé de ces travaux. De tels résultats ont à l'époque été établis pour les équilibres de Nash, les équilibres rationalisables et les équilibres corrélés.
11 Notons que rien ne dit comment les agents en sont arrivés à choisir les actions de cet équilibre. La théorie des jeux établit la stabilité des équilibres mais reste muette sur la façon dont les joueurs pourraient effectivement les atteindre.
12 Cf. Osborne et Rubinstein, 1994; Kreps et Wilson, 1982.
13 Cf. Camerer, 2003.
14 Cf. Dawes, 1980; Orbell et Dawes, 1981; Tajfel *et al.*, 1971; Brewer et Gardner, 1996.
15 Comme il nous en a fait part au cours d'une conversation privée.
16 Ce problème est notamment abordé dans Tuomela, 2007, p. 97-103.
17 On trouve de telles explications chez Tuomela, 2007.
18 Selon la fameuse analyse de Mancur Olson, ce sont souvent les diverses incitations individuelles accompagnant une action collective qui en conditionnent la réussite.
19 Mais cette possibilité d'une erreur d'interprétation est paradoxalement un avantage du point de vue coopératif : on croira observer de la coopération réussie plus souvent qu'elle ne l'est en réalité, ce qui promeut l'existence de groupe efficaces, pousse les rivaux éventuels à une coopération accrue, rend les identités de groupe manifeste, etc.
20 Je remercie Raimo Tuomela et Jacques Dubucs pour leurs commentaires.

## Références bibliographiques

Aumann, R. et A. Brandenburger
1995 «Epistemic Conditions for Nash Equilibrium», *Econometrica*, vol. 63, p. 1161-1180.

Bratman, M.
1992 «Shared Cooperate Activity», *The Philosophical Review*, vol. 101, nº 2, p. 327-341.

Brewer, M. et W. Gardner
1996 «Who is this "We"? Levels of Collective Identity and Self Identication», *Journal of Personality and Social Psychology*, vol. 71, p. 83-93.

Camerer, C.
2003 *Behavioral Game Theory*, Princeton, Princeton University Press.

Chant, S. et Z. Ernst
2007 «Group Intentions as Equilibria», *Philophical Studies*, vol. 133, nº 1, p. 95-109.

Chwe, M.-Y.
2001 *Rational Ritual. Culture, Coordination and Common Knowledge*, Princeton, Princeton University Press.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press

Cubitt, R. et R. Sugden
2003 «Common Knowledge, Salience and Convention : a Reconstruction of David Lewis' Game Theory», *Economics and Philosophy*, vol. 19, p. 175-210.
Dawes, R.
1980 «Social Dilemmas», *Annual Review of Psychology*, vol. 31, p. 169-193.
Gilbert, M.
1989 *On Social Facts*, Princeton, Princeton University Press.
Gold, N. et R. Sugden
2007 «Collective Intentions and Team Agency», *Journal of Philosophy*, vol. 104, nº 3, p. 109-137.
Kreps, D. et R. Wilson
1982 «Reputation and Imperfect Information», *Journal of Economic Theory*, vol. 27, nº 2, p. 253-279.
Miller, S.
2001 *Social Action — A Teleogical Account*, Cambridge, Cambridge University Press.
Orbell, J. et R. Dawes
1981 «Social Dilemmas», dans G. Stephenson et J. Davis, dir., *Progress in Applied Social Psychology*, volume 1, Hoboken, John Wiley and Sons, p. 37-65.
Osborne, M. et A. Rubinstein
1994 *A Course in Game Theory*, Cambridge (MA), MIT Press.
Rapoport, A.
1962 «Three Mode of Conflict», *Management Science*, vol. 7, nº 3, p. 210-218.
Regan, D.
1980 *Utilitarianism and Cooperation*, Oxford, Oxford University Press.
Searle, J.
1990 «Collective Intentions and Actions», dans P. Cohen, J. Morgan et M. Pollack, dir., *Intentions in Communication*, Cambridge (MA), MIT Press.
Tajfel, H., M. Billig, R. Bundy et C. Flament
1971 «Social Categorization in Group Behavior», *European Journal of Social Psychology*, vol. 1, p. 149-178.
Tollefsen, D.
2002 «Collective Intentionality and the Social Sciences», *Philosophy of the Social Sciences*, vol. 32, nº 1, p. 25-50.
Tuomela, R.
2000 *Cooperation — A Philosophical Study*, Mechelen, Kluwer.
2007 *The Philosophy of Sociality — The Shared Point of View*, Oxford, Oxford University Press.
Tuomela, R. et S. Miller
1988 «We-Intentions», *Philosophical Studies*, vol. 53, p. 367-389.
Walliser, B.
2002 «Les justifications des notions d'équilibre de jeux», *Revue d'économie politique*, vol. 112, nº 5, p. 695-716.

https://doi.org/10.1017/S0012217309090246 Published online by Cambridge University Press