

FAUT-IL JETER LE BÉBÉ DE L'INDIVIDUALISME METHODOLOGIQUE AVEC L'EAU DU BAIN?

[Soumis à *Critique* en 1998 refusé et jamais publié]

Philosophie de l'esprit et philosophie des sciences sociales

L'une des questions qui ont le plus occupé la philosophie de l'esprit récente est la suivante: l'image courante de l'individu comme un agent autonome pourvu de croyances, de désirs et d'autres états mentaux "intentionnels" n'est-elle pas menacée par les découvertes des neurosciences et de la psychologie cognitive qui semblent montrer que les états mentaux que la psychologie du sens commun considère comme unifiés et spécifiques sont en réalité sous-tendus par tout un ensemble de représentations, de processus et de causes qui ne sont pas directement accessibles à la conscience des agents? Ce que les agents considèrent comme des *raisons* de leurs comportements et de leurs actions ne sont-elles pas en réalité des descriptions qu'une psychologie plus fine révèle comme le produit de *causes*? Et s'il existe un tel niveau "inférieur" de description des causes du comportement, la conception du sens commun selon laquelle des états psychologiques causent nos actions n'est-elle pas obsolète, puisque ces états apparaissent comme de simples épiphénomènes d'une causalité psychique ou physique plus fondamentale? L'image des agents comme personnes, responsables de leurs actes, n'en est-elle pas compromise?¹ Il y a bien longtemps que la psychanalyse et toutes les théories déterministes de l'esprit nous ont rendus familiers avec ce genre de contraste. Les sciences sociales n'y ont pas échappé, puisque de nombreux théoriciens ont suggéré que si nous évaluions mieux les

¹ Sur cette problématique en philosophie de l'esprit, cf. notamment P.Engel, *Introduction à la philosophie de l'esprit*, Paris, La découverte, 1994, Pierre Jacob, *Pourquoi les choses ont-elles un sens?* Paris, O. Jacob 1996, et J. Proust, *Comment l'esprit vient aux bêtes*, Paris Gallimard 1996.

facteurs à l'oeuvre dans la vie sociale, nous devrions comprendre que notre conception usuelle de nous-mêmes en tant qu'agents autonomes est radicalement erronée: ce sont ces forces sociales, extérieures aux individus, qui sont réellement à l'oeuvre dans la production de nos actions, et non pas nos états psychologiques individuels. Il y a ainsi un parallèle frappant entre la position de ceux que l'on appelle des "éliminativistes" pour qui en toute rigueur des états mentaux tels que des croyances et des désirs *n'existent* pas et n'ont aucune efficacité causale², et la position, en philosophie des sciences sociales, des auteurs qui soutiennent que les croyances et les désirs individuels n'existent pas, ou en tout cas n'ont aucune efficacité causale, comparés aux facteurs causaux qui opèrent au niveau des groupes, des communautés ou des entités collectives.

Ce parallèle n'est pas l'effet du hasard. L'opposition entre les explications du comportement humain "par les causes" et les explications "par les raisons" a des affinités étroites avec la distinction classique weberienne entre les modèles d'explication causale en sciences sociales et les modèles fondés sur la "compréhension", et avec la distinction entre les modèles d'explication fondés sur les causes "irrationnelles" du comportement et les modèles fondés sur les "bonnes raisons" des agents.³ Le problème que cette menace causaliste pose pour la philosophie de la psychologie est symétrique de celui qui se pose en sciences sociales. En particulier, l'un des ses modèles est celui du choix rationnel, selon lequel les agents individuels ont des désirs (des utilités) et des croyances (des estimations de probabilité) et cherchent à accomplir les actions qu'ils jugent capables de satisfaire leurs désirs ou de "maximiser leur utilité espérée". Ce modèle est parfaitement conforme au sens commun, et c'est l'une des raisons de son attrait. Mais s'il existe, comme le soutiennent les conceptions "holistiques", des régularités de niveau supra-individuel qui ne se laissent pas réduire à l'agrégation des actions des agents, c'est précisément ce modèle de sens commun qui se trouve menacé, "par en haut", pour ainsi dire, plutôt que "par en bas" comme c'est le cas pour les explications causales infra-rationnelles.

Les tentatives récentes pour essayer de repenser ces oppositions et si possible les dépasser, ne manquent pas.⁴ Celle d'un philosophe des sciences

² Cet éliminativisme est représenté, au sein de la philosophie des sciences cognitives contemporaine par des auteurs comme P.M. et P. Churchland aux Etats Unis, et par des auteurs comme Changeux en France.

³ Pour reprendre la terminologie de R. Boudon, *Le juste et le vrai*, Paris Fayard 1995, introduction.

⁴ cf par exemple, dans des styles très distincts, John Searle, *The Social Construction of reality*, The Free Press, New York, 1995, et Philip Pettit, *The Common Mind*, Oxford University Press 1993

comme Mario Bunge ⁵ se propose de “trouver la philosophie dans les sciences sociales”. Mais il ne le fait pas en ayant l’air de vouloir apprendre aux chercheurs en sciences sociales comment poser des questions qu’ils devraient se poser mais qu’ils ne se posent pas. Il reprend leurs propres questions, et en systématisant les réponses qu’ils donnent eux-mêmes à des problèmes classiques comme celui de l’objectivité des sciences sociales, de la différence entre explication et compréhension, entre des approches “holistiques” et des approches “individualistes méthodologiques”, et entre des conceptions réalistes, subjectivistes et constructivistes quant à la nature des entités sociales. En dépit de cette revendication de modestie à l’égard de la sociologie, le livre, en fait très ambitieux, de Bunge n’est pas dénué de soucis normatifs. Il conçoit la philosophie des sciences sociales, au même titre que la philosophie des sciences en général, comme une forme de “métathéorie”, destinée à articuler des réponses à sept types de questions: 1) des questions relevant de la *logique* des sciences sociales et analysant la “forme logique des concepts, des propositions et des inférences effectuées dans les sciences sociales” et la structure formelle de leurs théories; (2) des questions relevant de la *sémantique* des sciences sociales, quant à ce que leurs théories désignent (des individus, des systèmes sociaux?) et quant à leur prétention à énoncer des vérités; (3) des questions relevant de l’épistémologie des sciences sociales, portant sur la nature de la connaissance des faits sociaux (est-ce une connaissance empirique?); (4) des questions relevant de la *méthodologie* de ces sciences, portant sur les modes de validations de leurs hypothèses, ou sur la nature de leurs méthodes (quantitatives ou pas); (5) des questions concernant leur *ontologie* (les entités en cause sont-elles collectives ou individuelles?); (6) des questions portant sur l’*axiologie* de ces disciplines, analysant question de savoir si leurs descriptions et leurs explications peuvent être libres de tout concept normatifs; et enfin (7) des questions portant sur l’*éthique* de ces sciences, et sur les codes de conduites des savants qui les pratiques, notamment relativement à leurs applications sociales.

Les réponses que Bunge apporte à ces questions, sont fortement inspirées de celles qu’il a proposées dans ses ouvrages sur la philosophie des sciences “exactes” ⁶: il défend une conception hypothético-déductive des théories scientifiques, une conception sémantique des théories comme modèles, une

⁵ Je ferai désormais référence à ce livre au moyen de l’abréviation “B”, suivi du numéro des pages.

⁶ Voir par exemple son livre *Philosophie de la physique*, tr.fr.Paris, Seuil 1975.

conception de la confirmation des hypothèses qui se veut à mi-chemin entre inductivisme et le déductivisme strict (B: 90) et une conception foncièrement réaliste de la science, selon laquelle les théories scientifiques sont susceptibles d'être vraies et de faire référence à des entités réelles. Ce réalisme le conduit notamment à fustiger les conceptions relativistes inspirées par de nombreux travaux de sciences sociales, et particulièrement de sociologie de la connaissance scientifique, ainsi que les diverses conceptions "herméneutiques" et idéalistes inspirées par les courants phénoménologiques et "subjectivistes". A l'exception de ce dernier aspect, il n'est pas absolument clair que la présentation des problèmes généraux de philosophie des sciences proposée par Bunge soit ancrée dans les méthodes et les travaux effectifs de sociologues contemporains, même si son réalisme contraste fortement avec les tendances instrumentalistes et positivistes qui ont dominé durant de longues années la méthodologie de l'économie.⁷ Bunge présente ce réalisme sous une forme passablement dogmatique. Celui-ci repose sur l'admission d'entités problématiques: son ontologie postule l'existence de *propriétés* et de *relations* indépendantes des individus qui les possèdent, en même temps que des *faits*, tenus comme autonomes par rapport aux individus, aux propriétés et aux relations. Je ne discuterai pas ici directement ces engagements ontologiques, ni les principes de la philosophie réaliste des sciences de Bunge. Je me concentrerai plutôt sur sa conception "systémique" de la nature des faits sociaux qui se veut une voie médiane entre un individualisme et un holisme, ses critiques de la théorie du choix rationnel, et sa critique du relativisme en sociologie de la connaissance. Sur tous ces points, ses analyses se prêtent explicitement à une comparaison avec des thèses défendues par des sociologues qui ont cherché à élargir le cadre de l'individualisme méthodologique et à repenser le modèle classique du choix rationnel.

Individualisme, holisme et systémisme

L'opposition entre l' "individualisme" et le "holisme" en sciences sociales fait partie de ces oppositions qu'on continue d'utiliser par paresse ou pour des besoins polémiques, mais dont finalement tous s'accordent pour reconnaître qu'elle ne veut pas en elle-même dire grand chose. On tend en particulier à

⁷ Locus classicus: M. Friedman, *Essays in Positive Economics*, Chicago University Press 1953

confondre deux sortes de problèmes.⁸ L'un est celui de savoir si l'existence de régularités sociales ou collectives compromet l'explication usuelle que les agents individuels donnent de leurs comportements en termes intentionnels. C'est le problème de savoir si une totalité est plus que la somme de ses parties, ou si les entités collectives sont plus que des agrégats d'individus. Il recoupe l'opposition entre l'*individualisme* (il n'y a que des individus) avec ce que l'on peut appeler un *collectivisme* (il y a des entités et des régularités collectives, distinctes des individus qui y figurent). Le second problème est distinct: c'est celui de savoir si les actions et les croyances d'un agent individuel dépendent ou non des relations sociales qu'il entretient avec les autres. Ce problème conduit à se demander si l'on peut concevoir les individus comme des atomes séparément des groupes et des structures sociales auxquels ils appartiennent. Il est plus approprié de caractériser cette opposition comme celle de l'*atomisme* et du *holisme*, en désignant par "holisme" non pas une doctrine qui affirme l'existence autonome des tous par rapport à leurs parties, mais une doctrine selon laquelle aucune unité (de pensée ou d'action) n'est strictement isolée d'autres unités.⁹ Malgré les apparences, les deux oppositions, individualisme contre collectivisme, et atomisme contre holisme, ne sont pas équivalentes. La première est, pour ainsi dire, verticale: quelle est la relation des individus à des contraintes sociales d'ordre supérieur (non individuelles)? La seconde est horizontale: quelles sont les relations des individus entre eux? On a tendance à confondre ces deux oppositions parce que l'atomisme et l'individualisme sont couramment assimilés à des formes de réductionnisme qui nient toute réalité aux totalités en dehors de leurs parties, alors que le collectivisme et le holisme sont souvent associés à des formes d'antiréductionnisme, qui mettent au contraire l'accent sur l'autonomie des totalités par rapport à leurs parties. Pour voir comment les deux oppositions ne sont pas équivalentes, il suffit de réfléchir au fait qu'on peut avoir la combinaison

⁸ J'utilise ici la manière dont ces questions sont formulées par Philip Pettit dans *The Common Mind*, op.cit., mais on pourrait aussi bien trouver des formulations équivalentes chez Boudon, op.cit., p.563, qui oppose d'une manière très similaire "atomisme" et "individualisme", et soutient que l'on peut souscrire au second sans souscrire au premier. Pour un diagnostic très semblable sur ces problèmes, cf. l'essai d'Alain Boyer, "Comment être individualiste méthodologique? Querelle des universaux et querelle des sciences sociales", à paraître in *Cahiers de philosophie de l'Université de Caen*, Actes du Premier Colloque de la Société de Philosophie Analytique

⁹ En ce sens le terme de "holisme" est bien plus proche de la doctrine que les philosophes du langage et de l'esprit désignent par ce terme: le holisme sémantique est habituellement la thèse selon laquelle aucune signification d'une expression au sein d'un langage n'est indépendante des significations des autres expressions, et le holisme "du mental" est la thèse selon laquelle aucun contenu de pensée n'est indépendant d'autres contenus de pensée.

suivante: on peut être un *individualiste* (qui nie l'existence d'*entités* collectives distinctes des individus) tout en étant un *holiste* (qui admet la dépendance des capacités et des pensées individuelles par rapport à celles des autres individus au sein d'une société)¹⁰. En effet, il est parfaitement possible de soutenir que mes croyances, par exemple, ou mes actions, ne pourraient pas exister s'il n'existait pas d'*autres* individus qui rendent possibles ces croyances ou ces actions— par exemple parce que ces croyances impliquent la possession de concepts que je dois pouvoir communiquer et partager avec d'autres— *sans* que cela implique qu'il existe des régularités ou des entités collectives qui influeraient par elles-mêmes sur mes croyances ou mes actions. En d'autres termes, l'interdépendance sociale des actes et des pensées n'entraîne pas que nous devions *réifier* les entités collectives qui sont les sujets de ces croyances ou de ces pensées.¹¹ Il y a, en ce sens, un usage tout à fait raisonnable du terme "holisme", à condition qu'on ne désigne pas par là une conception qui postulerait des entités supra-individuelles comme la "conscience collective" ou le *Volkgeist*, ou qui supposerait que tous les rapports sociaux sont liés au sein de totalités plus ou moins organiques, mais seulement une conception selon laquelle les actions des individus ne peuvent être analysées que dans des relations d'interdépendance avec celles des autres. Tout le problème est de savoir comment nous pouvons rendre compte de l'interdépendance en question sans réintroduire l'idée que les régularités sociales et collectives menacent l'intégrité des explications du comportement en termes psychologiques intentionnels qui font appel aux croyances et aux désirs individuels des agents.

Une manière courante de contribuer à la résolution de ce problème est de distinguer la question *ontologique* de la relation de partie à tout de la question *méthodologique* correspondante. C'est ce que fait habituellement la tradition "individualiste méthodologique" en sciences sociales (IM), qui entend faire abstraction du problème ontologique et se concentrer seulement sur les schèmes d'explication. Usant du modèle classique du choix rationnel et de la théorie de l'utilité, l'IM, dans un grand nombre de versions, soutient que les seuls types d'explications valides en sciences sociales sont de ce type. La méthodologie individualiste semble nous fournir une solution élégante au problème qui nous

¹⁰ Bien que la combinaison converse — un atomisme collectiviste— soit clairement incohérente.

¹¹ Ce type de réification est au contraire typique d'une certaine tradition hégélienne et marxiste, qui passe de l'idée — correcte— selon laquelle en toute rigueur un individu ne peut pas penser ou agir seul, mais doit penser ou agir au sein d'une communauté, à l'idée — incorrecte— selon laquelle il existe quelque chose comme un "esprit objectif" propre aux communautés elles-mêmes.

occupe, que nous pouvons formuler ainsi: quand un phénomène social manifeste l'existence d'une régularité causale à un niveau macrologique d'ordre supérieur ou collectif, il est toujours possible de recourir à l'existence de régularités descriptibles au niveau micrologique d'ordre inférieur ou individuel. Supposons par exemple que nous énoncions la régularité "macro" selon laquelle le chômage produit l'augmentation du crime. Elle énonce un lien causal entre deux sortes de faits descriptibles en termes macrologiques; mais ce lien causal lui-même *dépend* de régularités existant au niveau "micro": impliquant, par exemple des relations entre des familles, des individus, leurs besoins, leurs désirs et leurs décisions. C'est ce que Bunge appelle, à la suite de nombreux auteurs (B: 145), des liens "macro-micro". Un exemple assez typique de ce genre de liens est fourni par Thomas Schelling ¹². Il identifie une régularité reliant les attitudes de différentes parties au sein d'une communauté de voisinage mixte racialement ou religieusement— Blancs et Noirs, Protestants et Catholiques, etc. — à l'émergence d'une ségrégation résidentielle globale. Chaque partie a un minimum de huit voisins immédiats, comme une case sur un damier. Supposons que chacun soit soucieux de ne pas faire partie d'une minorité dans son groupe immédiat de voisins. De même que sur un damier, le résultat d'une distribution aléatoire de familles noires ou blanches conduira des familles qui se trouvaient au préalable satisfaites de leur statut de voisinage seront désireuses de changer de position. Une fois qu'elles auront changé de place elles produiront de nouvelles configurations qui conduiront de nouvelles familles à vouloir changer de place, et le résultat global sera une trame conduisant à une ségrégation quasi totale. Ce modèle simple est séduisant, parce qu'il construit les relations macro-micro exactement sur le modèle dont la mécanique statistique construit les relations entre la température d'un gaz et la distribution de ses molécules. Les régularités macro apparaissent ainsi dépendre, de manière causalement continue, des régularités micro, et inviter le modèle de la réduction des uns aux autres. Comme le remarque Bunge (B: 247) "l'individualiste considère la société comme analogue à un gaz de faible densité, les agents individuels étant les analogues des molécules, et leurs fonctions d'utilité les homologues des lois de leur mouvement. Comme dans la théorie cinétique élémentaire des gaz, les individus sont supposés agir individuellement les uns des autres, et ne subir que la contrainte de leur contenant (ou son analogue, le cadre institutionnel du marché).

¹² T.Schelling , "Models of segregation", *American Economic Review*, vol. 59, p. 488-493.

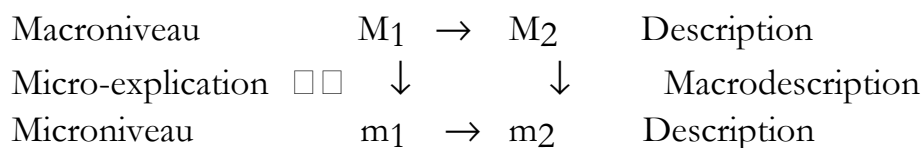
Le problème, note Bunge, est que les systèmes sociaux, de la famille aux firmes et aux Etats, ressemblent plus à des morceaux de matière condensée qu'à des gaz, et que les théories de la matière condensée ressemblent peu à celles des gaz, parce que la condensation est accompagnée de changements qualitatifs dans les structures macroscopiques qui ne sont pas continus avec les régularités microscopiques. Plus généralement, les critiques de la théorie du choix rationnel remarquent que certaines régularités n'apparaissent qu'au niveau macrosociologique et non pas au niveau individuel, que l'agrégation des comportements individuels ne suffit pas à rendre compte des régularités macrosociologiques observées parce qu'il faut au moins faire certaines hypothèses sur les motivations individuelles des agents (par exemple dans l'exemple de Schelling, on leur prête des désirs très simples de ne pas avoir de membre de l'autre communauté dans leur voisinage, mais il y a toutes sortes de circonstances dans lesquelles ils pourraient ne pas avoir ces désirs, ou bien que ceux-ci entre en compétition avec d'autres, comme celui d'être dans le voisinage de familles de revenus égal ou semblable, etc.), et *last but not least*, parce que l'agrégation des préférences individuelles peut produire des "effets pervers" qui ne sont pas continus avec les intentions individuelles des agents.¹³

Ces critiques reviennent à dire qu'il est totalement gratuit de supposer, comme le fait l'IM sous sa forme "pure" que la rationalité des agents économiques ou sociaux peut se réduire à la rationalité instrumentale d'un individu doté, de manière indifférenciée d'"intérêts", parfaitement informé sur les options dont il dispose, et libre de choisir en conséquence. Elles reviennent également à faire remarquer que l'on ne peut pas purement et simplement identifier les relations macrologiques à des relations qui seraient à la fois dépendantes des relations micrologiques et réductibles, en dernière instance, à ces dernières. S'ensuit-il que l'on doive revenir à la thèse durkheimienne selon laquelle il existe des faits sociaux *sui generis* irréductibles à des agrégations de décisions individuelles?

Les individualistes méthodologiques ont ici le choix entre renoncer à leurs postulats et les élargir. Bunge semble à certains moments soutenir que c'est cette seconde voie qui est la bonne. A d'autres moments, il est plus radical, et propose d'abandonner l'individualisme lui-même. Commençons d'abord par ses arguments en faveur de la première voie. S'appuyant sur les travaux de Boudon et

¹³ *Locus classicus*: R. Boudon, *Effets pervers et ordre social*, Paris, PUF, 1977.

de Coleman¹⁴, il soutient que ceux-ci défendent en fait une position médiane entre l'individualisme et le holisme qu'il appelle "l'indivdholisme", qu'il assimile à la position de Boudon, selon laquelle bien que tout fait macrosocial M doive être expliqué comme le produit d'un ensemble d'actions individuelles m , admet, en retour que "les actions des individus sont rendues compréhensibles, au sens weberien, en les reliant à leur environnement social, la situation S, des acteurs: $m = m(S)$ ", et telle que "finalement la situation elle-même doit être expliquée comme le résultat de certaines variables macrosociologiques, ou au moins de certaines variables situées à un niveau plus élevé que S." En d'autres termes les faits macrosociologiques sont fonction des actions individuelles, lesquelles sont fonction des situations, lesquelles sont en retour encore fonction de variables macrosociologiques de niveau supérieur.¹⁵ Selon Bunge, le "néo-individualisme" de Boudon (et de Coleman) consiste à soutenir que nous devons combiner des réductions macro-micro ou "de haut en bas" (M-m) avec des réductions micro-macro "de bas en haut" (m-M). Cela donne ce que Bunge (B: 149) appelle un "diagramme de Boudon-Coleman", qui a la forme suivante:



La thèse à laquelle on parvient, selon Bunge, et celle qu'il défend pour sa part, serait mieux appelée "systémisme". Elle consiste à admettre l'existence de certaines entités supra-individuelles, les systèmes sociaux, et à expliquer leur existence en termes d'actions individuelles, qui sont elles-mêmes expliquées en retour en termes de contextes sociaux. Selon Bunge, un système est une entité complexe, un triplet constitué de sa composition, de son environnement et de sa structure. L'ontologie du systémisme se réduit aux principes suivants: a) une société est un système de sous-système changeants, b) chaque société a des

¹⁴ Il fait référence, outre aux travaux classiques de Boudon (comme *Effets pervers et ordre social*, op.cit. et *La logique du social*, Paris, Hachette 1979), au livre-somme de Coleman, *Foundations of Social Theory*, Harvard University Press, 1990.

¹⁵ R.Boudon, "The individualistic tradition in sociology", in Alexander, J.C., Giesen, B., Munch, R., et Smelser, N.J (eds) *The micro-microlink*, Berkeley, Los Angeles, London, University of California Press, 1987, p. 45-70 (p.46), cité par B: 253. Une excellente étude récente de ce "lien mico-macro" en sociologie est celle de M. Cherkaoui, "Le réel et ses niveaux: peut-on toujours fonder la macrologie sur la micrologie?", *Revue française de sociologie*, 38, 3, 1997, p.497-524.

propriétés globales, dont certaines sont résultantes de propriétés microsociologiques, d'autres émergentes, bien que liées aux composants individuels et à leurs relations, c) l'interaction entre deux systèmes sociaux met en jeu seulement des individus. La contrepartie méthodologique de ces principes est que la sociologie est l'étude des systèmes sociaux et de leurs composants individuels, et que les hypothèses et les théories des sciences sociales sont testées par rapport à des données environnementales et sociales (B: 267-268). Le mérite de cette approche, selon Bunge, est qu'elle combine les intuitions correctes qui sont à la base à la fois de l'individualisme et du holisme: il n'y a pas d'autres interactions qu'individuelles, mais celles-ci ne peuvent se comprendre qu'au sein d'un environnement social.

Il semble que le systémisme de Bunge illustre la combinaison déjà suggérée entre un *individualisme* — compris au sens de la thèse selon laquelle les régularités sociales de niveau macroscopique ne menacent pas l'intégrité des explications micrologiques en termes de la psychologie individuelle des agents — et un *holisme* — compris comme la thèse selon laquelle les actions des agents ne peuvent pas se comprendre en dehors de leurs interactions sociales. Parce qu'il défend cette seconde thèse, le systémiste n'est pas un *atomiste*, puisqu'il refuse l'idée que les actions des agents pourraient être comprises indépendamment de toute interaction avec d'autres agents. Parce qu'il défend la première thèse, le systémiste n'est pas un *collectiviste*, parce qu'il n'admet pas l'existence d'entités supra-individuelles collectives *sui generis*.

Considérons par exemple l'explication des phénomènes de panique collective qui ont attiré depuis longtemps l'attention des sociologues. Comme l'a reconnu Le Bon, le comportement d'une foule en panique ne peut pas s'expliquer seulement comme la sommation de comportements individuels isolés ni comme le produit d'une sorte d'esprit collectif qui s'incarnerait dans les circonstances en question. Il faut admettre que ce qui se produit est une combinaison suivie de la création de caractéristiques nouvelles, tout comme en chimie des éléments mis en contact se combinent pour produire un corps nouveau différent des corps qui le forment.¹⁶ En d'autres termes, la foule manifeste, comme la combinaison des éléments chimiques, des propriétés *émergentes*, c'est-à-dire des propriétés qui dépendent systématiquement des propriétés de leurs éléments, mais qui ne

¹⁶ Le Bon, *Psychologie des foules*, 1895, rééd. Retz, 1975.

laissent cependant pas réduire à l'agrégation de ces dernières.¹⁷ Mais il importe ici de ne pas confondre cette thèse d'émergence avec le type de dépendance des liens macrologiques par rapports aux liens micrologiques illustré par l'exemple de Schelling ci-dessus. Car ce dernier est en fait conforme à une position atomiste selon laquelle les régularités de niveau supérieur sont causalement continues par rapport aux régularités de niveau inférieur. Dans le cas d'une panique où des gens s'enfuient, par exemple d'une salle de cinéma en feu, la position atomiste serait représentée par l'application du modèle le plus connu de la théorie des jeux, le dilemme du prisonnier (DP). Si l'on prend les individus isolément, et qu'on suppose que deux joueurs-spectateurs ont chacun deux options, fuir précipitamment ou se diriger calmement vers la sortie, on peut exhiber une matrice du jeu d'après laquelle si les autres sortent calmement, l'agent a tout intérêt à fuir précipitamment, et si les autres fuient précipitamment, il a également intérêt à faire de même, ce qui produit la fuite précipitée de tous. Il s'ensuit que pour tous la seule action rationnelle en de telles circonstances est de fuir, ce qui explique le comportement de panique non pas, comme le faisait Le Bon, par des phénomènes irrationnels de suggestion et de contagion, mais comme des phénomènes explicables en termes des schèmes de rationalité instrumentale les plus classiques. Coleman, qui expose en détail ces analyses¹⁸, fait cependant remarquer que l'on n'est pas, dans un tel cas, dans une situation classique de DP, parce qu'à la différence de ce qui se passe dans ce dernier les agents n'agissent pas indépendamment des autres: tous peuvent voir ce que les autres font. L'alternative devant laquelle chaque agent se trouve en réalité, nous dit Coleman, est entre prendre précipitamment la fuite, et *transférer son pouvoir aux autres et courir s'ils courent et aller calmement vers la sortie s'il en font autant*. Si l'agent sait que son action n'affectera pas le comportement des autres, il aura intérêt à courir précipitamment vers la sortie. En revanche s'il sait que son action affectera leur comportement, il

¹⁷ cf Boudon et Bourricaud, *Dictionnaire de sociologie*, Paris, PUF, 1986, p.xiv (cité dans B: 45) : "Bien que les faits sociaux doivent être traités comme les produits de systèmes ou de processus d'action ou d'interaction, ils ne sont pas réductibles à des relations interpersonnelles, mais doivent être traités comme des phénomènes émergents, ou en d'autres termes, de composition." Il est intéressant de constater ici que la notion d'émergence, qui a ses origines dans la biologie évolutionniste de la fin du siècle dernier, est très voisine, dans le cas présent, de la notion que les philosophes de l'esprit appellent *survenance* (*supervenience*), qui caractérise la dépendance d'un ensemble de propriétés (notamment physiques) par rapport à un autre ensemble de propriétés (notamment mentales) *sans* qu'il soit vrai que les secondes soient *réductibles* au premières. cf. P.Engel, *Introduction à la philosophie de l'esprit*, *op.cit.*

¹⁸ Dues à Brown, *Social Psychology*, New York, Free Press, 1965. Coleman, *op.cit.* p. 200 sq. On trouvera dans Cherkaoui, *op.cit.* p. 518 un commentaire détaillé de cet exemple.

aura intérêt à sortir calmement. Cette interdépendance des acteurs rend à la fois la situation très distincte de celle d'un DP et en même temps ne déroge pas aux principes de l'explication par choix rationnel. Si l'on entend par "holisme" la thèse faible selon laquelle les agents n'agissent ni ne pensent indépendamment les uns des autres, et si l'on entend par "individualisme" la thèse selon laquelle les explications par comportement rationnel ne sont pas menacées par l'existence de régularités supra-individuelles, on obtient précisément la combinaison recherchée d'"individholisme" ou de systémisme.

Coleman indique lui-même très bien la nature de son approche:

"Les lecteurs familiers avec les débats et les discussions sur le holisme méthodologique et l'individualisme méthodologique reconnaîtront que [ma] position est une variante de l'individualisme méthodologique. Mais c'est une variante spéciale. Elle ne présuppose pas que l'explication du comportement systématique ne consiste en rien d'autre que des actions et orientations individuelles, prises comme des agrégats. L'interaction entre individus est considérée comme produisant des phénomènes émergents au niveau du système, c'est à dire comme des phénomènes qui ne sont ni visés ni prédits par les individus. De plus elle n'implique pas que pour un objectif particulier une explication doit descendre jusqu'au niveau individuel pour être satisfaisante. Le critère est plutôt pragmatique: l'explication est satisfaisante si elle est utile pour les types d'intervention qui la motivent. Ce critère requiert habituellement une explication qui va au delà du système pris comme tout, mais qui n'est pas nécessairement fondée dans des actions et des orientations individuelles."¹⁹

Selon Bunge, le systémisme est en fait la position implicite de la plupart des sociologues, dans la mesure où ils reconnaissent tous la spécificité des systèmes sociaux et de leurs structures par rapport aux actions des agents individuels, ainsi que les effets de rétroaction de ces systèmes sur celles-ci. En ce sens le systémisme peut être considéré comme un individualisme bien compris. Pourtant la version du systémisme de Bunge reste ambiguë. Malgré ses déclarations selon lesquelles le niveau macrosociologique du système et le niveau microsociologique des interactions individuelles sont interdépendants et irréductibles l'un à l'autre, la manière dont il caractérise les systèmes (par exemple B: 270) en fait des entités qui paraissent autonomes et autosubsistantes par rapport aux individus qui les

¹⁹ Coleman, *op.cit.* p.5

composent. Il ne rend pas parfaitement clair que sa position se distingue de celle des sociologues pour qui le niveau macrosociologique est premier est fondamental.²⁰ Il n'est pas évident que, contrairement à ce qu'il soutient, le "néo-individualisme" d'un Boudon soit un "crypto-systémisme". Les critiques qu'il adresse à l'IM en général montrent en fait qu'il est plus loin qu'il ne le dit de la position de Boudon et des sociologues qui, comme Coleman, veulent rester fidèles à certains principes individualistes. L'arbitrage entre les diverses positions qui recherchent une voie médiane entre l'individualisme et le holisme ne doit pas se faire uniquement sur le plan ontologique, mais aussi sur le plan du type d'hypothèse qu'elles adoptent au sujet de la rationalité des agents.

Les limites de la théorie du choix rationnel et leur interprétation

Bunge résume utilement les principes de la théorie du choix rationnel comme fondée sur les axiomes suivants:

A1: *Utilité et probabilité* : tout agent a une fonction d'utilité stable, et peut assigner une probabilité subjective à tout événement qui l'intéresse;

A2: *Rationalité* : tous les agents sont rationnels, au sens où ils agissent exclusivement selon leurs intérêts personnels, de manière à maximiser leurs utilités, simples ou espérées;

A3: *Individualisme ontologique*: les groupes sociaux ne sont que des collections d'individus;

A4: *Individualisme méthodologique*: les propriétés et les changements de tout groupe social ne peuvent être compris qu'en étudiant le comportement de ses membres.

On a déjà vu en quoi A3 et A4 peuvent être amendés sans renoncer à certains principes de base de l'individualisme. Depuis un vingtaine d'années au moins, les efforts des théoriciens des sciences sociales qui n'entendent pas jeter le bébé du principe de rationalité avec l'eau du bain de la théorie de l'utilité et du choix social ont porté sur les manières dont on peut réviser le modèle du choix rationnel sans renoncer à l'hypothèse de rationalité.

Le problème est que celle-ci est, tout autant que les notions d'individualisme ou de holisme, extrêmement ambiguë. Dans un article fameux,

²⁰ cf. par exemple P.Blau, *Inequality and heterogeneity* , New York, Free Press, 1977, et les analyses de Cherkaoui, *op.cit* p518-520.

Jon Elster a distingué au moins 16 sens du mot “rationalité”.²¹ Contentons-nous d’abord des cinq sens distingués par Bunge (B: 366):

R1: *Principe du comportement adaptable*: “Les agents agissent toujours de manière appropriée à la situation dans laquelle ils se trouvent” (Popper);

R2 *Principe de la rationalité instrumentale*: les agents adoptent toujours les moyens qui sont le plus susceptibles de produire les résultats escomptés;

R3 (= A2) *Principe de rationalité économique*: les agents agissent toujours de manière à maximiser leur utilité espérée;

R4 *Principe du moindre effort*: les agents utilisent toujours les moyens les moins coûteux pour atteindre leurs objectifs;

R5 *Principe de rationalité subjective* : les agents agissent toujours d’après leurs croyances portant sur la situation dans laquelle ils se trouvent, et sur les meilleurs moyens d’obtenir les conséquences qu’ils préfèrent.

Bunge soutient que tous ces sens du terme “rationalité” sont soit faux (R2, puisque les agents n’ont pas toujours l’information nécessaire pour adapter leurs moyens à leurs fins, R3 parce qu’il ne permet pas de dire si l’on doit maximiser son utilité à court, moyen, ou long terme, alors qu’il est clair que la maximisation à court terme peut entrer en conflit avec celle à long terme; R4 parce que “l’activité peut être sa propre récompense”), soit triviaux ou vides (R1 de l’aveu de Popper lui-même, et R5 parce qu’il est irréfutable). Toutes ces critiques sont quasiment triviales, depuis celles d’Allais contre le modèle du choix rationnel ou les études psychologiques de Kahneman et Tversky qui montrent que les maximisateurs d’utilité, bien qu’ils existent, sont rares, et que les gens obéissent bien plus souvent à des aversions au risque qu’à des soucis maximisateurs. De nombreux auteurs n’en ont pas conclu que l’hypothèse de rationalité devait être abandonnée: ils ont par exemple, à la suite de Simon, proposé des modèles de rationalité limitée ou sous contraintes (le “satisficing”) qui impliquent des hypothèses plus faibles que celles de la rationalité optimisatrice. Mais ce n’est pas la voie qu’entend suivre Bunge. Son diagnostic sur la théorie du choix rationnel est sans appel:

“Le lecteur pourra avoir l’impression que ma critique est excessive, et que je jette le bébé avec l’eau du bain. Ma réponse est qu’il n’y a pas de bébé. En fait la théorie du choix rationnel est née en 1789 dans le domaine de la philosophie

²¹ J. Elster, “Rationality”, in G. Floistad, *Contemporary Philosophy, A new survey, Vol. III, Philosophy of Action*, La Haye, Nihoff, 1981.

politique et morale (l'utilitarisme de Bentham) et en 1871 dans le domaine de l'économie néoclassique, et est donc loin d'être jeune. De plus elle est morte il y a longtemps de malnutrition mathématique et d'anémie psychologique, faute d'avoir les enzymes requis pour digérer des faits sociaux et des principes moraux simples, et à la suite de son exposition au bruit et à la fureur des orages sociaux.”(B: 385).

Le remède que Bunge entend proposer aux déficiences de la théorie du choix rationnel est cependant décevant. Il est certes marqué du sceau du bon sens: il ne peut y avoir qu'un concept large, donc vague, de la rationalité, et toute hypothèse de rationalité doit être tempérée par un “*hard-nose realism*” qui appelle en retour “des tests empiriques et l'inclusion de traits non rationnels dans le comportement humain.”(B: 386). Il y a fort à parier cependant que le “réalisme” en question soit lui-même parfaitement trivial (prenez les gens tels qu'ils sont) et que l'usage d'un principe de rationalité se réduise à une maxime toute aussi vide: traitez les gens comme rationnels, sauf si vous avez de bonnes raisons de ne pas le faire. Et quand il n'est pas trivial le réalisme de Bunge conduit à des jugements assez surprenants. C'est au nom de ce réalisme qu'il critique “Pareto, Weber et Boudon” pour avoir défendu la concept R5 de la rationalité subjective. Si l'on adopte ce principe, nous dit-il, on sera amené, comme Pareto, à traiter le sacrifice que faisaient les marins grecs anciens à Poséidon comme tout aussi “logique” que l'acte de ramer. Mais pour Bunge, “le sacrifice est objectivement irrationnel, alors que le fait de ramer est objectivement rationnel.” (B: 374) Mais depuis quand un rameur qui, pour reprendre l'exemple célèbre de Hume, serait le seul à ramer sur un bateau alors que son compagnon se repose sur ses efforts serait-il “objectivement rationnel”? Et depuis quand un rituel serait-il *en lui-même* “objectivement irrationnel”? On ne voit pas pourquoi des éléments rationnels ne pourraient pas coexister avec d'autres qui ne le sont pas, ce qui signifie aussi qu'il ne peut pas y avoir de rationalité ou d'irrationalité objective, et que nos diagnostics sur la rationalité dépendent toujours des principes d'interprétation et de normes de rationalité que nous adoptons, ainsi que des contextes dans lesquels nous les appliquons.²² Si l'on veut échapper à ce genre d'absurdité, il faut bien admettre qu'il existe un sens tout à fait respectable de la

²² Il me semble que c'est l'essence de la conception de la rationalité défendue par des auteurs comme Davidson. cf. P.Engel “Trois formes de normativité”, in P.Engel, ed. *Lire Davidson*, L'Eclat, combas, 1994.

notion de rationalité, qui correspond approximativement à R5, et selon lequel les actions d'un agent sont rationnelles si on les considère à la lumière des *bonnes raisons*, ou des raisons subjectives, que cet agent a d'agir.

Boudon a qualifié ce type de rationalité de "cognitive". Par là il désigne les types de raisons que les agents ont d'adhérer à des croyances qui rendent celles-ci plausibles à leurs yeux, et qui expliquent leurs jugements et leurs actions, quand bien même elles pourraient apparaître, au regard d'une théorie normative particulière de la rationalité, comme irrationnelles. Parmi les nombreux exemples qu'il donne, l'un des plus intéressants est celui d'une expérience bien connue de psychologie sociale, dans laquelle on demande à des médecins d'évaluer la probabilité pour qu'un sujet soit atteint d'une certaine maladie, dans l'hypothèse où elle a un taux de pénétration de 1/ 1000, et où le test permettant de détecter sa présence a un taux de 5% de "faux positifs", et où le test a été soumis à l'individu en question. La plupart des médecins répondent, à tort, que le sujet positif a dans ce cas 95% de chances d'être atteint, alors que la bonne réponse qu'il a moins de 2 chances sur cent de l'être. Si l'on se place du point de vue de la théorie normative usuelle de l'inférence statistique, celle de l'inférence bayésienne usuelle en termes de probabilités conditionnelles, les sujets paraissent clairement dévier de ces normes de rationalité, ce que l'on peut expliquer en termes de facteurs causaux tels que des "biais" ou des "heuristiques" qui influent sur le raisonnement et le parasitent. Mais Boudon suggère une autre hypothèse: les sujets surestiment les risques de maladie parce qu'ils interprètent mal la notion de "faux positif", et donc ont de bonnes raisons de se tromper, selon un raisonnement que l'on peut reconstruire.²³ Il existe donc, selon cette analyse, un grand nombre de croyances qui sont fausses, mais qui ont néanmoins un caractère "transsubjectif" (propre à la plupart des sujets et reconnaissables par eux) et qui sont, bien que fausses, parfaitement *plausibles*. On pourrait répondre ici, à l'instar de Bunge (B:374), que l'hypothèse de la rationalité cognitive ou subjective est tout aussi intenable et irréfutable que l'hypothèse freudienne selon laquelle tous les rêves qui n'ont pas un contenu sexuel explicite en ont un latent, et que le fait qu'un interprète puisse reconstruire le comportement d'un agent apparemment irrationnel comme rationnel ne prouve en rien que ce comportement *soit* rationnel. Mais il est faux qu'il soit intenable. Car de nombreux chercheurs ont remarqué que quand on modifie les termes des questions posées dans les expériences de psychologie

²³Boudon, *Le juste et le vrai*, op.cit, p.68-73.

sociale en question, de manière par exemple à éviter les termes qui les induisent systématiquement en erreur, ou en interprétant différemment leurs estimations de probabilités, leurs réponses s'améliorent et deviennent la plupart du temps correctes.²⁴ Les effets d'irrationalité sont donc souvent dus à des restrictions trop fortes des principes de rationalité utilisés. Le principe de la rationalité cognitive n'est autre que celui qui inspirait la méthodologie weberienne du *Verstehen*, selon lequel expliquer des croyances ou des actions d'un groupe, c'est essayer de montrer comment ces croyances ou ces actions ont un sens aux yeux des agents. Il ne revient nullement à attribuer systématiquement aux agents des croyances ou des actions rationnelles, mais sert plutôt de garde-fou contre des explications qui concluraient trop hâtivement, comme le fait Bunge, à l'irrationalité "objective" des croyances et des actions. Cela ne veut pas dire non plus, à mon sens, qu'il n'y ait pas des types d'actions ou de croyances qui aient des caractères qui semblent intrinsèquement irrationnels, comme celles qui reposent sur des formes d'*akrasia* ou de duperie de soi (*self-deception*) qui ont été étudiées systématiquement dans une vaste littérature²⁵. Ce que montrent en fait les travaux sur ces phénomènes est qu'ils peuvent être décomposés en des constituants "rationnels" (des raisonnements plausibles du point de vue de la rationalité cognitive) et des constituants "irrationnels" (par exemple des facteurs émotionnels), et non pas qu'ils tombent nécessairement dans l'une ou l'autre catégorie. Un agent qui agit contre son meilleur jugement ou akratiquement n'est pas nécessairement irrationnel, parce qu'il peut être dans un processus de révision de ses propres normes de conduite; mais certaines manières dont on révisé ses croyances ou ses normes de jugement et de conduite peuvent être irrationnelles. En d'autres termes, et pour reprendre ceux de Pareto, il ne suffit pas de distinguer des actions ou des croyances "logiques" ou rationnelles d'actions et de croyances "non

²⁴ Par exemple, des auteurs comme G. Gigerenzer ont montré, à mon avis de manière convaincante, qu'un certain nombre de résultats de Kahneman et Tversky devenaient moins saillants, voire disparaissaient, quand on interprète les estimations de probabilité intuitives des agents en termes fréquentistes plutôt que subjectivistes ou bayésiens. cf. par exemple G. Gigerenzer, "How to make cognitive illusions disappear",

²⁵ Cf les travaux bien connus de Elster, *Ulyses and the Sirens*, *Studies in Rationality and Irrationality*, Cambridge University Press et Maison des Sciences de l'homme, 1979, tr.fr. partielle in *Le laboureur et enfants*, Paris, Minuit 1987, et Davidson, *Essays on actions and events*, Oxford, Oxford university Press, 1980, tr.fr. *Actions et événements*, Paris, PUF 1993.

logiques” ou irrationnelles, mais de voir que les deux peuvent coexister et que les raisons d’agir ou de croire peuvent combiner les deux éléments.²⁶

C’est ici qu’il faut introduire un autre niveau de rationalité, qui est absent ou seulement implicite dans les discussions de la théorie du choix rationnel au sens étroit (utilitariste): la rationalité qui relève non plus des *formes d’inférence*, théoriques ou pratiques, mais des *normes* d’évaluation de ces formes d’inférence. Un agent n’est pas seulement rationnel ou irrationnel parce qu’il suit ou non certaines règles, mais aussi parce qu’il a des moyens d’évaluer ses règles et de se conformer à ces évaluations. Par exemple, on n’est pas rationnel dans son comportement déductif simplement parce que ce dernier obéit à certains schèmes d’inférence valides (comme le *modus tollens* de la logique classique) mais aussi parce que l’on a conscience du fait que ces règles sont valides, et qu’on est en mesure de les prescrire à autrui. Mais il peut y avoir des cas où les normes de raisonnement que nous reconnaissons entrent en conflit avec les inférences que nous sommes intuitivement prêts à admettre. Dans ce cas on a le choix entre rejeter ces inférences ou réviser nos normes initiales. Il est essentiel, dans ces cas, de tenir compte de deux facteurs: d’une part un facteur temporel ou dynamique, par lequel nous révisons nos croyances initiales à la lumière de celles que nous acquérons, et d’autre part un facteur que l’on peut appeler, à la suite de Rawls, d’équilibrage de nos croyances et des principes de justification que nous adoptons pour celles-ci. Cette situation ne se produit pas seulement dans le cas des normes épistémiques mais aussi dans le cas des normes pratiques ou d’action. En ce sens, il se peut que des règles d’inférence pratiques, comme la règle bayésienne qui prescrit de maximiser notre utilité espérée à la lumière des probabilités conditionnelles que nous assignons à des événements, entrent en conflit avec d’autres normes que celles de la rationalité instrumentale. Il est banal de remarquer que le principe de la théorie du choix rationnel semble, sur point, être parfaitement fidèle à la

²⁶ Comme le dit Boudon (*op.cit.* p.558) commentant le “postulat de Weber-Popper” sur la rationalité: “L’homme n’est ni rationnel ni irrationnel; seuls des comportements particuliers, des actions particulières, des croyances particulières et des combinaisons de ces éléments peuvent être dits “rationnels”; ils seront dits “rationnels” si une théorie “rationnelle” permet de les expliquer et si cette théorie apparait supérieure à toute théorie “irrationnelle”. Sur Pareto, cf. A. Bouvier, “Modèles paretiens en théorie des idéologies/ sociologie des représentations. Emotivisme et cognitivisme dans la théorie des dérivations de Pareto”, *L’année sociologique*? 1992, 42, 345-368. Il est tentant ici, d’opposer les explications en termes de “raisons” (supposées forcément rationnelles) aux explications en termes de “causes” (supposées forcément irrationnelles). Mais si les remarques sommaires ci-dessus sont correctes, elles ne s’opposent pas nécessairement. C’est le sens de la thèse davidsonienne selon laquelle les raisons sont aussi des causes, que j’ai essayée d’appliquer aux explications en psychologie dans *Philosophie et psychologie*, Paris, Gallimard 1996.

conception humienne selon laquelle la rationalité des choix ne porte que sur l'adaptation des moyens aux fins, et pas sur la pertinence des fins elles-mêmes, et que la théorie instrumentale de la rationalité traite sur le même plan des “désirs” des agents, des principes altruistes de comportement ou des valeurs morales universalistes, au titre des “intérêts” qui guident leurs actions. Le même postulat conduisait Hume à s'étonner de ce que les gens puissent agir d'après des normes sociales auxquelles ils sont intrinsèquement attachés, ou par fidélité à un parti, et surtout, de ce fait, contre leur intérêt²⁷. Et c'est précisément sur ce point que les critiques de l'individualisme méthodologique ont soutenu que le modèle instrumental de la rationalité trouvait ses limites, exactement là où les conceptions utilitaristes et conséquentialistes en éthique trouvent les leurs. Le comportement “normatif” paraît précisément opaque, non déchiffrable en termes d'intérêt et d'utilité. Si l'on n'entend pas, à la différence de Bunge, rejeter simplement le principe de rationalité en sciences sociales, l'alternative devant laquelle on se trouve est la suivante: ou bien on doit introduire une autre forme de rationalité que la rationalité instrumentale et la rationalité “cognitive”, ou bien on doit adapter le cadre initial et l'élargir.

Selon Boudon, c'est la première démarche qu'on doit suivre. A la suite de Weber, il soutient qu'il existe une autre forme de rationalité, la *Wertrationalität* ou “rationalité axiologique”. Elle relève de tous les sens dans lesquels les agents jugent rationnel d'adhérer à des croyances parce qu'ils acceptent des normes ou des valeurs. De plus ces normes entrent la plupart du temps en conflit avec les jugements d'utilité économique. Les exemples sont légion: un couple âgé qui se retrouve seul dans maison trop grande pour eux que leurs enfants devenus adultes ont désertée refuse de la quitter pour un appartement plus petit et moins coûteux parce qu’ils ont toujours vécu là”, les paysans et les bourgeois du XIXème siècle refusent l'assurance-vie parce qu'ils refusent d'accorder un prix à la vie²⁸, on juge que la conscription est juste et acceptable s'agissant des soldats mais non pas des mineurs parce que leur activité est vitale pour la communauté²⁹, etc. Boudon souligne que ces jugements normatifs sont non seulement irréductibles à des jugements conséquentialistes, et par suite ne se prêtent pas à des extensions du modèle du choix rationnel, mais encore qu'ils sont objectifs, comme les valeurs

²⁷ Sur cet héritage humien de l'individualisme, cf. E. Picavet, *Choix rationnel et vie publique*, Paris, PUF 1997, ch. 2, et P. Demeulenaere, *Homo Oeconomicus*, Paris, PUF 1996, ch.1.

²⁸ Boudon, *Le juste et le vrai*, p.274-276

²⁹ Walzer, *Spheres of justice*, tr.fr. *Sphères de justice*, Paris, Seuil 1996

morales qui les fondent, au même titre que peuvent l'être les vérités démonstratives. Par là il ne veut pas dire qu'elles ont une forme d'existence comparable à celle des Idées platoniciennes, mais qu'elles sont publiques, argumentables et dotées d'une forme de stabilité et de constance qui les rend irréductibles à des normes seulement relatives à telle ou telle communauté.

Je ne discuterai pas ici cette théorie "néo-rationaliste" des valeurs. Elle s'oppose, comme on vient de le remarquer, à la stratégie usuelle des tenants du modèle du choix rationnel, qui consiste à étendre ce modèle au comportement "normatif" en considérant celui-ci comme "émergent" à partir du comportement individuel maximisateur. Selon cette stratégie, les normes de coopération et de réciprocité surgissent simplement de situations du type de celle du dilemme du prisonnier et de la stabilité de stratégies conduisant à des situations d'équilibre.³⁰ Mais il est clair, selon ces analyses, que le sens de l'"émergence" en question est bien celui de la dépendance continue des régularités intentionnelles au niveau macrologiques par rapport aux régularités au niveau micrologique, et non pas celui d'une émergence de structures irréductibles. Elle ne revient pas à accepter l'idée d'une rationalité axiologique opérant de manière autonome par rapport aux intérêts individuels des agents. Il existe aussi tout un ensemble de travaux qui partent des modèles usuels de la théorie des jeux et de l'utilité, mais qui soutiennent qu'un grand nombre de situations sont inexplicables en ces termes si l'on ne fait pas intervenir des normes spécifiques (et non pas "émergentes" au sens précédent) telles que la bienveillance, la réciprocité, ou le sacrifice de l'intérêt. Ainsi dans un article célèbre Amartya Sen a-t-il proposé l'idée qu'il existe non seulement un classement des actions morales, mais aussi un classement des classements d'actions morales, ou un classement des classements de préférences, et il a aussi suggéré que ces métaclassements ne se réduisaient pas à l'effet de préférences purement individuelles, mais de préférences et de principes d'emblée collectifs.³¹ La question importante est de savoir si les actions effectuées selon des normes sont incompatibles avec une vision conséquentialiste de la moralité. Les conséquentialistes ont bien entendu soutenu que ce n'était pas le cas, et ils ont

³⁰ cf. par exemple, parmi les travaux les plus connus, E. Ullman-Margalit, *The emergence of norms*, Oxford, Oxford University Press, 1977; R. Axelrod, *The Evolution of Cooperation*, Basic Book, tr.fr. *Donnant donnant*, Paris, Odile Jacob, 1991, et David Gauthier, *Morals by agreement*, Oxford 1986. ces analyses sont souvent étendues dans le cadre de conceptions évolutionnistes de l'émergence des normes.

³¹ A Sen, "Rational Fools", *Philosophy and Public Affairs*, 6, tr.fr. in *Ethique et économie*, Paris, PUF 1993.

essayé de défendre l'idée que les règles d'universalisation par exemple, qui sont typiques des raisons "axiologiques", n'étaient qu'un prolongement des principes conséquentialistes. On peut soutenir que bien qu'elles n'en dérivent pas, nous les mesurons néanmoins en termes de leurs conséquences pour la communauté et que les "bonnes raisons" normatives des agents ne sont pas incompatibles avec des intuitions utilitaristes, mêmes si elles ne s'y réduisent pas (par exemple dans le cas de la conscription de Walzer, c'est l'intérêt de la communauté qui semble motiver les jugements). Il existe donc tout un domaine, encore à explorer, dans lequel on pourrait assister à une collaboration, plutôt qu'à une opposition, entre les motifs conséquentialistes et les motifs "déontologiques" au sein des raisons de l'action³². Cela veut dire qu'une fois que l'on distingue différents types de rationalité — instrumentale, cognitive et axiologique — on n'est pas obligé de soutenir qu'ils sont exclusifs les uns des autres. Les travaux inspirés par la théorie du choix rationnel se sont aussi étendus dans une autre direction, en modifiant les bases *psychologiques* de cette théorie. Sous sa version simple, comme on l'a vu, elle considère que les agents sont simplement dotés de croyances et de désirs, et que l'on peut isoler, à un moment donné, les fonctions d'utilité et les estimations de probabilité des agents. Cette hypothèse est apparue comme beaucoup trop simple, au regard du fait que les agents modifient leurs estimations à travers le temps et en fonction de données contextuelles, et en regard du fait, déjà noté, qu'ils n'ont pas simplement des croyances et des désirs, mais des croyances *sur* leurs croyances et des désirs *au sujet de* leurs désirs, et en particulier des désirs de ne pas avoir certains désirs (comme dans les cas d'*akrasia*). Cela revient aussi à dire qu'ils incluent dans leurs croyances et dans leurs motivations des normes d'évaluation épistémiques et pratiques. Un grand nombre de travaux récents ont aussi permis de rétablir une notion que la psychologie usuelle de la décision avait plus ou moins cherché à évacuer, celle d'intention, en montrant qu'il n'était pas clair qu'une intention soit simplement une combinaison de croyances et de désirs.³³ Les versions récentes de la théorie de la décision tiennent compte de ces facteurs dynamiques et normatifs, et permettent de fournir une image beaucoup plus complexe et réaliste de ces changements de préférence, de ces effets de planification du comportement et des stratégies épistémiques. En ce sens, le

³² Dans divers essais, E. Picavet a fourni un balisage très intéressant de ce domaine. cf. par exemple son article "L'élargissement de l'individualisme méthodologique", à paraître.

³³ cf. en particulier M. Bratman, *Intentions, Plans and Practical Reason*, MIT Press, Cambridge Mass, 1987.

modèle du choix rationnel conserve encore une valeur heuristique certaine en sciences sociales, et je crois qu'il est prématuré de jeter le bébé avec l'eau du bain. Entre l'extension linéaire du modèle du choix rationnel à tous les domaines de la vie sociale et son rejet pur et simple, il y a place pour une position intermédiaire qui combine les éléments instrumentaux et axiologiques.³⁴

Réalisme et relativisme

J'ai suggéré que le systémisme de Bunge, malgré sa recherche d'une position intermédiaire de ce genre, ne fournissait pas nécessairement les instruments nécessaires pour dépasser l'opposition entre des théories "individualistes" pures et dures et des positions "holistes" radicales, et que des conceptions néo-weberiennes avaient plus de chances de le faire. En revanche, les critiques de Bunge, le plus souvent très dures, contre les doctrines relativistes, herméneutiques, ou constructivistes en sociologie de la connaissance, devraient rencontrer la sympathie des philosophes et des sociologues qui n'ont pas le sentiment que les travaux culturalistes en anthropologie, les études sur la "vie de laboratoire" et les thèses du "programme fort" en sociologie de la connaissance sur la manière dont les théories scientifiques "construisent" le réel plutôt qu'elles ne décrivent une réalité indépendante aient réellement montré que les savants ne croient pas à leurs théories parce qu'ils estiment qu'elles ont de bonnes chances d'être vraies, ni ne montrent que les idéaux classiques de rationalité et d'objectivité sont obsolètes.³⁵ Bunge oppose à ces formes de relativisme ce que l'on pourrait appeler un réalisme "industriel" doté d'engagements ontologiques lourds: une conception correspondantiste de la vérité, une conception selon laquelle les théories scientifiques désignent des entités réelles, et comme on l'a vu, une ontologie des "faits" comme entités autosubsistantes et autonomes. Sur ce point, si son livre est destiné à exposer les controverses récentes en philosophie des sciences qui portent sur le réalisme, l'instrumentalisme, ou la confirmation des théories scientifiques, je ne crois pas qu'on puisse le recommander aux praticiens des sciences sociales qui souhaiteraient disposer d'une introduction à ces controverses. Rien n'est dit, par exemple, des discussions qui ont opposé, durant

³⁴ Il me semble que des positions comme celle de Philip Pettit, *op cit.* vont également dans ce sens.

³⁵ Pour des diagnostics semblables à ceux de Bunge, cf. Boudon, *Le juste et le vrai, op.cit.* ch. 12 et 13 notamment.

les dernières décennies, les partisans de formes de bayésiannisme ou d'instrumentalisme renouvelées en philosophie des sciences ³⁶ à des partisans de formes sophistiquées de réalisme ³⁷. Ces discussions montrent à la fois qu'il est possible de défendre des types d'instrumentalisme en philosophie des sciences qui n'entraînent pas le relativisme ou le renoncement à l'objectivité de l'enquête scientifique, et des types de réalisme plus faibles que la version ontologique proposée par Bunge.

En tout état de cause, et sans pouvoir discuter ce point ici, je dirai qu'il existe des formes plus intéressantes d'anti-réalisme en philosophie des sciences que les variétés de relativisme et de constructivisme examinées par Bunge, et qu'elles méritent une analyse spécifique, parce que tout anti-réalisme n'est pas un relativisme. Il y a cependant un point sur lequel les anti-réalistes et les réalistes modérés doivent s'accorder: même s'il y a de bonnes raisons de penser que le concept, usuel comme scientifique, de vérité est un concept nécessairement interne à nos théories et à nos enquêtes, et non pas externe ou transcendant par rapport à elles, il ne s'ensuit pas que nos croyances communes soient vraies simplement parce que nous les acceptons ou parce que notre communauté culturelle les accepte. Dans le même esprit que les tenants de la rationalité "cognitive" je dirais que nous tenons nos croyances comme vraies parce que nous avons de bonnes raisons de les croire telles, et nous ne nous engageons dans des enquêtes scientifiques que parce qu'elles visent une vérité indépendante de ces enquêtes qui en est une norme.³⁸ En ce sens les jugements normatifs épistémiques que nous formulons sont tout aussi irréductibles aux intérêts particuliers — ou à ceux des groupes particuliers— de ceux qui les défendent que les jugements normatifs pratiques ou éthiques sont irréductibles aux intérêts conséquentialistes des agents dans le contexte de la théorie des valeurs morales. Aussi "puritains" et "mertoniens" ces jugements peuvent-ils paraître, il n'est pas plus facile de se débarrasser de l'idée de vérité scientifique que cela ne l'est de se débarrasser de celle de vérité ou de justesse en éthique.

³⁶ C. Howson et P. Urbach, *Scientific Reasoning, the Bayesian Approach*, Open court, La Salle, Ill. 19, Van Fraassen, *The Scientific image*, Oxford, Oxford university Press, 1981.

³⁷ cf. A. Fine, *The Shaky Game, Einstein, Relativity and the Quantum Theory*, Chicago, Chicago University Press 1984.

³⁸ C'est le sens du réalisme "minimal" que j'ai exposé au sujet des théories philosophiques de la vérité in *La vérité, réflexions sur quelques truismes*, Paris, Hatier 1998.

Pour comprendre comment une position fondée sur une notion comme celle de “rationalité cognitive” n’entraîne pas une forme de relativisme ou de contextualisme quant aux bonnes raisons que les agents peuvent avoir pour leurs actions et leurs croyances, revenons au parallèle entre la menace causale en philosophie de l’esprit et la menace causale en philosophie des sciences sociales. Nous supposons que les états mentaux de la psychologie ordinaire, comme les croyances, ont un pouvoir causal sur le comportement en vertu même de leur *contenu*. Par exemple, la psychologie ordinaire explique l’action d’un individu qui en frappe un autre en disant qu’il croyait avoir été insulté et désirait se venger. Ceci suppose que nous puissions caractériser des concepts comme ceux d’“insulte” ou de “vengeance” en des termes que l’agent est lui-même capable de comprendre et de conceptualiser. Or supposons maintenant que nous disions que ces concepts mêmes ont une nature intrinsèquement sociale, et que ce qui compte comme une insulte, ou une vengeance, diffère selon les contextes sociaux (par exemple, pour caricaturer, qu’ils ne sont pas les mêmes pour un corse que pour un suédois). Dans ce cas, l’explication psychologique même que nous donnerons de l’action et de ses “raisons” sera affectée par l’“environnement” familial, social, linguistique etc. et la manière dont le groupe, par ses pratiques ou ces croyances collectives détermine le sens de ces concepts. Un contextualisme radical soutiendra que ces explications n’ont pas de sens en dehors de cette référence “externe” à un environnement, et c’est en ce sens que l’explication psychologique “individualiste” paraîtra abstraite ou incomplète, et sera menacée par ces déterminants externes. C’est en ce sens également que les philosophes de l’esprit qu’on appelle “externalistes” soutiennent que les contenus mentaux eux-mêmes de la psychologie des agents sont déterminés par leur environnement extérieur, et par conséquent que cette psychologie ne peut pas être individualiste ou “internaliste”. Le même type de raisonnement peut être tenu au sujet des visées théoriques d’un scientifique quand il teste une hypothèse: on peut soutenir que celle-ci n’est pas déterminée uniquement par sa visée “interne”, mais par les pressions que l’environnement du laboratoire auquel il appartient, ou que certains groupes de collègues, ou même des groupes sociaux entiers, font peser sur le choix de son hypothèse. C’est ainsi que l’on aboutit aux formes de constructivisme et de relativisme qui se sont répandues en sociologie de la connaissance scientifique. Ces formes de raisonnement sont tentantes, mais fallacieuses. Il peut être parfaitement correct de dire que le sens de concepts

comme “insulte” ou “vengeance” ou “hypothèse valide” n’est pas confiné dans les limites de la psychologie d’un individu et ne peut être compris que dans une certaine forme de contexte social ou pratique (par lesquels la communauté dicte ses règles), sans qu’il s’ensuive que l’action des agents puisse s’expliquer par une forme d’action à distance de ces contextes ou de ces règles. Pour que notre individu puisse se sentir insulté et désirer se venger, il faut qu’il puisse reconnaître un type d’action comme une insulte non pas seulement à la lumière de ce que sa communauté lui fait comprendre comme une insulte, mais à la lumière de ce que lui-même comprend comme tel. De même pour qu’un savant puisse proposer une hypothèse, il ne suffit pas que celle-ci compte comme acceptable pour son laboratoire, mais aussi qu’il ait de bonnes raisons de l’admettre. En d’autres termes, la contribution que l’environnement social apporte à l’individuation des concepts qui entrent dans nos interprétations et explications du comportement n’implique pas que ces explications soient de part et d’autre environnementales et contextuelles. Il est vrai, en revanche, qu’il doit pouvoir les communiquer pour qu’elles fassent sens pour lui, sans quoi elles seraient opaques à lui-même autant qu’aux autres.³⁹

Au total, un livre comme celui de Bunge illustre assez bien le mouvement qui a conduit, depuis plus d’une trentaine d’années les philosophes et les chercheurs en sciences sociales à s’écarter de plus en plus des principes et des hypothèses de base de la théorie du choix rationnel et de l’individualisme méthodologique conçus au sens étroit, et la nécessité où ils se sont trouvés de repenser ses fondements et d’en élargir les bases. Mais on est encore loin de la synthèse que l’on pouvait attendre de ces travaux, et il ne convainc pas qu’on doive rejeter unilatéralement l’hypothèse de rationalité.

Pascal Engel

³⁹ On trouvera dans le livre de R. Bhargava, *Individualism in social science*, Oxford, Oxford university Press 1992, une argumentation contre l’individualisme méthodologique fondée sur le type d’externalisme et de contextualisme en philosophie de l’esprit comme celle que je viens de suggérer. L’erreur qu’il fait, à mon sens, est de soutenir que ce contextualisme infirme les principes de l’individualisme. Pettit (*op.cit*), qui examine les mêmes argument, ne commet pas cette erreur, et considère que l’externalisme et le holisme du mental est tout à fait compatible avec des explications individualistes.