



**La complexité aux limites de la rationalité. Proposition de définition de la structure de base de la complexité du couple actions-attitudes par la critique du principe du tiers-exclu**  
**Complexity on the Fringes of rationality: Proposal for a Definition of the Basic Structure Underlying the Complexity of the Action-Attitude Coupling through a Critique of the Excluded Middle Principle**

Denis Martouzet

Volume 11, Number 1, November 2015

URI: <https://id.erudit.org/iderudit/1035942ar>  
DOI: <https://doi.org/10.7202/1035942ar>

[See table of contents](#)

Publisher(s)

Prise de parole

ISSN

1712-8307 (print)  
1918-7475 (digital)

[Explore this journal](#)

Cite this article

Martouzet, D. (2015). La complexité aux limites de la rationalité. Proposition de définition de la structure de base de la complexité du couple actions-attitudes par la critique du principe du tiers-exclu. *Nouvelles perspectives en sciences sociales*, 11(1), 377–424. <https://doi.org/10.7202/1035942ar>

Article abstract

This article aims to define a basic structure of the complexity in the relationship between cognitive attitude and action by questioning the principle of the excluded middle. Taking as a starting point the human will, we show that, given its weaknesses (Aristotelian akrasia and objectives referred to as essentially secondary effects by Jon Elster), the person implements self-restraint strategies (such as commitment), which are irrational but effective. Their effectiveness arises from the human capacity to self-deceive. The structure of this «mauvaise foi», as Jean-Paul Sartre conceptualized it, is based on the result of the contradictory co-presence of incompatible beliefs and, even more, their mutual strengthening, in spite of the fact that forcing oneself to believe is an example of an essentially secondary effect. We describe the mechanism of this contradiction, thus questioning the law of the excluded middle. The result can be thought of as the «elementary brick» of human complexity.

## **La complexité aux limites de la rationalité. Proposition de définition de la structure de base de la complexité du couple actions- attitudes par la critique du principe du tiers-exclu**

**DENIS MARTOUZET**

UMR 7324 CITERES, Équipe IPAPE,  
Université François-Rabelais, Tours

**M**aurice Allais, dans un article paru en 1955, pose les principes de l'action rationnelle et de l'acteur rationnel :

un homme est réputé rationnel lorsque : a. il poursuit des fins cohérentes avec elles-mêmes; b. il emploie des moyens appropriés aux fins poursuivies. Or [...] ces deux conditions entraînent comme seules conséquences : 1° Que le champ de choix soit ordonné; 2° Que l'axiome de préférence absolue soit suivi; 3° Que l'on considère les probabilités objectives. Les points 1° et 2° sont admis par tout le monde; quant au point 3°, il paraît difficilement contestable qu'il y ait intérêt à substituer aux probabilités objectives des probabilités subjectives qui en soient distinctes<sup>1</sup>.

Confortant la notion de rationalité, à partir d'une tradition que l'on peut faire remonter à Jeremy Bentham<sup>2</sup>, Allais précise

---

<sup>1</sup> Maurice Allais, « Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine », *Annales des Mines*, vol. 144, n° spécial, 1955, p. 31.

<sup>2</sup> Jeremy Bentham, *An Introduction to the Principals of Morals and Legislation*, Oxford, Oxford University Press, 1996.

un certain nombre de postulats, centraux dans la littérature de l'économie standard. Ces postulats sont, d'une part, relatifs à l'acteur quant à ses « manières d'être » (rationnel, conséquentialiste, maximisateur et égoïste<sup>3</sup>). Ils sont réputés, par les tenants de la théorie du choix rationnel, telle que, par exemple, mobilisée par Gary Becker<sup>4</sup> ou, à leur manière, Stephen Levitt et Steve Dubner<sup>5</sup>, suffisants pour rendre compte des comportements individuels, ce que remet en cause, par exemple, à la suite de Max Weber<sup>6</sup> qui distingue la *Werkrationalität* (rationalité en valeur) de la *Zweckrationalität* (rationalité en finalité) récusant ainsi la légitimité du postulat de conséquentialisme, Raymond Boudon<sup>7</sup> avec le Modèle Rationnel Général qu'il élabore. D'autre part, des postulats renvoient aux modes de calcul sous-tendant le choix de telle ou telle option dans le champ des possibles : la préférence, l'espérance et la cohérence.

Or, ces deux ensembles de postulats ont été de multiples manières mis à mal, tant par les économistes non orthodoxes que par les sociologues et les psychologues. La préférence suppose l'existence préalable d'un ordre complet et stable des options possibles : l'acteur a préalablement hiérarchisé ses préférences, et ce, de façon durable, si ce n'est définitive. On note le caractère irréaliste de ce postulat, tant sur la stabilité (les préférences se construisent en situation elle-même évolutive) que sur le caractère complet (il peut y avoir des options en situation d'*ex æquo*). L'espérance, qui est au cœur du mode de calcul, pose immédiatement problème ainsi que l'indique Maurice Allais lorsqu'il préconise, *a contrario* des théories américaines, d'utiliser des

<sup>3</sup> Raymond Boudon, *Raison. Bonnes raisons*, Paris, Presses universitaires de France, coll. « Philosophe en sciences humaines », 2007.

<sup>4</sup> Voir, par exemple : Gary Becker, *Human Capital : A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago, University of Chicago Press, 1964; Gary Becker, *A Treatise on the Family*, Cambridge (MA), Harvard University Press, 1991 [1981].

<sup>5</sup> Steven D. Levitt et Stephen J. Dubner, *Freakonomics. A Rogue Economist Explores the Hidden Side of Everything*, New York, William Morrow, 2005.

<sup>6</sup> Max Weber, *Économie et Société*, 2 tomes, Paris, Pocket, coll. « Agora », 1995.

<sup>7</sup> Raymond Boudon, *Raisons. Bonnes raisons*, *op. cit.*

probabilités subjectives ouvrant ainsi la porte aux travaux des cognitivistes, en passant par Herbert Simon<sup>8</sup> et Daniel Kahneman<sup>9</sup>.

Il peut y avoir préférence pour une option alors même que le produit du gain final par la probabilité d'occurrence est identique pour les deux options : l'être humain n'est pas l'âne de Buridan<sup>10</sup>.

Maurice Allais va plus loin en soulignant aussi l'imbrication des postulats :

On ne saurait, en fait, considérer comme irrationnel un homme prudent qui préfère avoir une moindre espérance mathématique psychologique s'il peut bénéficier d'une dispersion plus faible. On ne saurait non plus considérer comme irrationnel un individu qui aime le risque en tant que tel, c'est-à-dire un individu qui préfère avoir une moindre espérance mathématique psychologique, à condition de pouvoir disposer de quelques possibilités d'un gain psychologique très élevé. On pourra dire, si on veut, qu'un tel individu est imprudent, ce qu'il peut être en effet, mais en aucun cas il ne nous paraît possible de dire qu'il est irrationnel. On ne saurait trop le souligner, en dehors de la condition de cohérence, il n'y a pas de critère de la rationalité des fins considérées en elles-mêmes. Ces fins sont absolument arbitraires<sup>11</sup>.

<sup>8</sup> Les travaux d'Herbert A. Simon portant sur la rationalité limitée sont foisonnants. On se référera prioritairement à : Herbert A. Simon, « A Behavioral Model of Rational Choice », *Quarterly Journal of Economics*, n° 69, 1955, p. 99-118; Herbert A. Simon, « Theories of Decision-Making in Economics and Behavioral Science », *American Economic Review*, vol. 49, n° 1, 1959, p. 253-283; Allen Newell, Herbert A. Simon, *Human Problem Solving*, Englewood Cliffs (NJ), Prentice-Hall, 1972; Herbert A. Simon, « On how to Decide What to Do », *The Bell Journal of Economics*, vol. 9, n° 2, 1978, p. 494-507; Herbert A. Simon, *Models of Thought*, vol. 1, New Haven (CT), Yale University Press, 1979; Herbert A. Simon, « From Substantive to Procedural Rationality », dans Spiro J. Latsis (dir.), *Method and Appraisal in Economics*, Cambridge (MA), Cambridge University Press, 1976, p. 129-148; Herbert A. Simon, *Reason in Human Affairs*, Stanford (CA), Stanford University Press, 1983. On peut se reporter aussi à André Demailly, *Herbert Simon et les sciences de conception*, Paris, L'Harmattan, coll. « Ingénium », 2004.

<sup>9</sup> Daniel Kahneman, *Système 1 / Système 2 : les deux vitesses de la pensée*, Paris, Flammarion, 2012.

<sup>10</sup> L'histoire attribuée à Jean Buridan, philosophe français du XIII<sup>e</sup> siècle, l'expérience de pensée dite de l'âne : cet animal meurt de faim et de soif, étant autant attiré par le picotin d'avoine qui se trouve d'un côté que de l'eau qui se trouve de l'autre : il n'est pas capable de choisir.

<sup>11</sup> Maurice Allais, *op. cit.*, p. 27.

Reste le principe de cohérence, qui s'avère ainsi central :

Nous pensons quant à nous que la conduite d'un individu est rationnelle lorsqu'elle satisfait au principe général de non-contradiction avec soi-même, lequel implique lui-même deux conditions; la première, c'est la cohérence logique des fins poursuivies, la seconde, c'est l'utilisation des moyens effectivement appropriés aux fins poursuivies<sup>12</sup>.

Une question est donc de savoir si, généralement, l'individu est cohérent ou non.

Ainsi, nombre d'auteurs<sup>13</sup> se sont lancés dans une course à la découverte des cas spécifiques qui remettent en cause ces postulats. La psychologie comportementale expérimentale, notamment, a apporté nombre de cas montrant, dans des situations précises, par des protocoles *ad hoc* mis en œuvre, la relative incohérence de l'acteur, mais sans qu'apparaisse pour l'instant du moins une théorie unifiée de l'incohérence et, partant, de la non-rationalité de l'être humain. Au final, les critiques portées précédemment à la théorie standard ne font que l'égratigner : elle est capable de les absorber par des modifications qui sont alors de l'ordre de la

<sup>12</sup> Maurice Allais, *op. cit.*, p. 26.

<sup>13</sup> Établir la liste des parutions scientifiques relatives à cette question n'est pas envisageable dans le cadre d'un article. Nous proposons quelques auteurs et quelques titres dont la vocation d'ensemble est de montrer la très grande variété des approches : Amos Tversky, Daniel Kahneman, « Rational Choice and the Framing of Decisions », dans David E. Bell, Howard Raiffa, Amos Tversky (dir.), *Decision Making, Descriptive, Normative and Prescriptive Interactions*, Cambridge (MA), Cambridge University Press, 1988, p. 167-192; Amos Tversky, « Rational Theory and Constructive Choice », dans Kenneth J. Arrow, Enrico Colombatto, Mark Perlman, Christian Schmidt (dir.), *The Rational Foundations of Economics Behaviour*, New York, Macmillan and St-Martin Press, 1996, p. 185-197; Graham Looms, « Further Evidence on the Impact of Regret and Disappointment in Choice under Uncertainty », *Economica*, London School of Economics and Political Science, vol. 2, n° 1, 1988, p. 47-62; Philippe Gouillou, *Pourquoi les femmes des riches sont belles : programmation génétique et compétition sexuelle*, Bruxelles, De Boeck, 2014; Gérard Bronner, *L'incertitude*, Paris, Presses universitaires de France, coll. « Que sais-je? », 1997; Amos Tversky, Daniel Kahneman et Paul Slovic, *Judgment under Uncertainty: Heuristics and Biases*, Cambridge (MA), Cambridge University Press, 1982; George J. Stigler, « The Economics of Information », *Journal of Political Economy*, vol. 69, n° 3, 1961, p. 213-225; Louis Lévy-Garboua, « L'économie et le rationnel », *L'année sociologique*, no 31, 1981, p. 19-47; Jean Piaget et Bärbel Inhelder, *La genèse de l'idée de hasard chez l'enfant*, Paris, Presses universitaires de France, 1974.

nuance. Nous partageons avec Herbert Simon l'idée que les postulats de la théorie du choix rationnel nous paraissent insuffisants à la fois dans leur capacité à décrire le réel et dans sa vraisemblance quant aux comportements humains. Mais, en accord avec la rupture que propose Victor Vieira da Silva entre pseudo-comportementalistes et comportementalistes<sup>14</sup>, nous nous distinguons radicalement de Herbert Simon par le fait que celui-ci « aménage » cette théorie de l'intérieur pour tendre vers une certaine vraisemblance sans la remettre en cause. Boudon montre que Simon ne fait qu'ajouter un postulat aux autres postulats de cette théorie, postulats qu'il conserve<sup>15</sup>. De notre côté, nous partons de l'idée que, quelle que soit l'extension que l'on confère à la notion de rationalité, elle ne permet pas l'explication de l'ensemble des comportements humains, non pas à la marge de cet ensemble mais en son cœur.

De leur côté, certains philosophes se sont attaqués au cœur même de la notion de cohérence de l'action humaine, tant du côté de l'existentialisme français que de la philosophie analytique nord-américaine.

Ce sont les avancées de ces philosophes que nous allons examiner, mobilisant les notions de volonté, d'intention et d'engagement de façon non à clarifier la notion même de rationalité mais à mettre en évidence ses limites. L'objectif visé, au-delà, est une définition de la complexité de l'action humaine et, par extension, de la complexité. Plus précisément, il s'agit de proposer une définition de la « structure de base » de la complexité, sa « plus petite brique », à l'image de la plus petite « brique de matière » des physiciens.

La démarche est la suivante : prenant pour point de départ que postuler la rationalité de l'action et de l'acteur, faisant de celui-ci un modèle, est nécessairement une simplification dont l'unité de base est le couple fin-moyen, c'est-à-dire en posant que « rationalité → simplicité », nous déduisons de cette proposition

<sup>14</sup> Victor Vieira da Silva, *L'hypothèse de rationalité et ses prémisses : le défi comportementaliste, une investigation aux frontières de l'économie et de la psychologie*, thèse de doctorat de sciences économiques, Lyon, Université de Lyon, 2001.

<sup>15</sup> Raymond Boudon, *Raison. Bonnes raisons*, op. cit.

la contraposée logique « non-simplicité  $\rightarrow$  non-rationalité ». Il s'agit donc de faire des investigations aux limites de la rationalité pour en induire ce qui se situe aux limites de la complexité, c'est-à-dire cette « unité de base » de la complexité.

Nous ne nous attaquons pas à l'ensemble de la théorie ou des approches de la complexité. Ce qui est ici plus particulièrement visé est, dans le couple attitude cognitive-action, la mise en évidence de l'inapplicabilité du principe du tiers-exclu : les propositions A et non-A découlent, au-delà de la logique classique, l'une de l'autre, directement et indirectement et, de ce fait, coexistent.

Deux sources d'information ont été pour nous majeures pour l'examen de ces limites de la rationalité. La première est l'observation que chacun peut faire d'actes qu'il est possible de qualifier de non rationnels. Cependant on peut toujours douter de leur réelle non-rationalité puisque l'on n'a jamais toute l'information nécessaire pour juger du bien-fondé ou non, en termes de rationalité, de l'acte observé. Aussi est-il aisé, ainsi que l'a montré la psychologie, de toujours attribuer une « bonne raison » à tout acte observé, en accord avec la terminologie qu'emploie Raymond Boudon, ce qui évite de se confronter au gouffre de l'irrationalité qui pourrait se cacher derrière. Là où cela devient plus délicat, c'est lorsque ces actes non rationnels que l'on a sous les yeux sont les nôtres. Si je fais un acte non rationnel, il devient urgent de le rationaliser de crainte de passer pour irrationnel, a-rationnel, non-rationnel, voire simplement peu raisonnable, et de se sentir comme tel. Mais si cela devient urgent pour garder l'intégrité de la personne, cela devient plus difficile de rationaliser car je suis parfaitement au courant du contexte dans lequel j'ai agi et de la manière dont j'ai opéré des choix. Pourtant, on effectue souvent des rationalisations *a posteriori* qui nient, par différentes formes de réécriture du passé ou de la réalité présente, tout ou partie de ce contexte. La seconde source d'information, qui fixe le cadre général de référence de notre propos, est la production philosophique avec pour principales références Jon Elster, Donald Davidson, Jean-Paul Sartre, Jean-Pierre Dupuy, Pascal Engel.

C'est essentiellement cette seconde source qui sera mobilisée, et donc présentée, dans cet article même si l'observation, l'échange et l'introspection ont pu orienter nos questionnements et nos analyses.

Notre point de départ est la question de la volonté, qui n'entre pas directement dans la thématique de la rationalité, mais qui est révélatrice de certaines de ses limites. Cette entrée permet d'évacuer la question de l'information et de son accessibilité ainsi que du traitement qui en est fait, nous éloignant en cela des travaux d'Herbert Simon. Ensuite, notre propos sera ancré essentiellement dans la notion de stratégie et notamment de pointer des stratégies individuelles efficaces et qui, pourtant, comportent des éléments résolument contradictoires entre eux et notamment supposent une réécriture du passé. Enfin, nous aborderons les notions de mensonge à soi-même, ou de mauvaise foi, qui, elles, supposent une réécriture du présent.

Notre cheminement logique va de la volonté à la structure de base de la complexité dans le couple attitude cognitive/action. La volonté fait preuve de faiblesse : certains objectifs ne peuvent être voulus sous peine de devenir impossibles à atteindre, et pourtant il nous arrive de les vouloir bien que nous sachions cela voué à l'échec. Par ailleurs, toujours par faiblesse de notre volonté, nous choisissons de faire ce que nous ne voulons ni ne préférons (cas de l'akrasie ou acrasie). Vouloir ce qui ne peut pas, de toute évidence, être atteint ou vouloir ce que nous ne préférons pas met à mal l'idée d'un homme rationnel. Nous mettons alors parfois en œuvre des stratégies d'autocontrainte irrationnelle, mais efficace (exemple du *sunk cost fallacy*, exemple de l'engagement qui, par définition n'est pas rationnel). Ces stratégies, non rationnelles, mais possiblement efficaces, le sont du fait de notre capacité à être de mauvaise foi, au sens de Sartre<sup>16</sup>. La structure de cette mauvaise foi est fondée sur le résultat de la contradiction entre croyances incompatibles et, pourtant, leur renforcement réciproque, malgré le fait que se forcer à croire (et plus encore à

<sup>16</sup> Jean-Paul Sartre, *L'être et le néant*, Paris, Gallimard, coll. « Bibliothèque des idées », 1943.



croire deux propositions contradictoires) est un exemple d'objectif que l'on ne peut vouloir. C'est le mécanisme de cette contradiction que nous décrivons et posons le résultat de ce mécanisme comme la « brique élémentaire » de la complexité<sup>17</sup>. Notre angle d'attaque est la critique du principe du tiers exclu, dans la lignée des travaux de Stéphane Lupasco<sup>18</sup>.

## I – Intention, volonté et rationalité

Les effets essentiellement secondaires

Elster a analysé ce qu'il a nommé « effet essentiellement secondaire », montrant le caractère limité de la volonté humaine<sup>19</sup>. C'est pour nous un préambule puisque la volonté se situe hors de la réflexion sur la rationalité, elle est « en amont ». Cependant ce préambule nous conduit à la notion de « faiblesse de la volonté » qui conduit ou peut conduire à différents types d'action qui n'entrent pas dans le cadre de la rationalité.

Il existe des objectifs que l'on ne peut pas vouloir, sous peine de ne pas parvenir à les atteindre. L'idée n'est pas qu'on ne dispose pas des moyens pour atteindre notre objectif (ce sont alors des « infaisables ») ni qu'on ne se donne pas les moyens de l'atteindre (dans le cas du velléitaire) mais que le seul fait de vouloir atteindre cet objectif entraîne des conséquences telles que l'objectif ne peut pas être atteint : c'est la volonté ici qui pose problème. Quelques exemples peuvent être présentés. Le premier est celui de la personne qui, insomniaque pour une nuit ou depuis longtemps, cherche à s'endormir en s'efforçant, par la volonté (et non par un quelconque artifice chimique) d'y parvenir. Une telle stratégie ne peut amener que son contraire, c'est-à-dire la prolongation de la phase de non-sommeil préalable à l'endormissement. Un deuxième exemple pourrait être la volonté de créer l'apparence du désordre (par exemple pour donner l'idée, fausse, qu'il y a eu

<sup>17</sup> Edgar Morin, *Introduction à la pensée complexe*, Paris, Seuil, 2005.

<sup>18</sup> Stéphane Lupasco, *Le principe d'antagonisme et la logique de l'énergie. Prolégomènes à une science de la contradiction*, Paris, Hermann, 1951.

<sup>19</sup> Jon Elster, *Le laboureur et ses enfants. Deux essais sur les limites de la rationalité*, Paris, Minuit, 1986.

fouille ou cambriolage ou, sur un bureau, l'impression qu'il y a eu une activité intense qui, en fait, n'a pas eu lieu). Un œil extérieur ne sera généralement pas dupe et percevra, souvent s'il est attentif, qu'il y a eu « triche ». On peut aussi citer le cas des personnes qui veulent donner l'apparence de la spontanéité ou de l'aisance dans un milieu donné. Jouer la spontanéité est forcément non spontané. Seul le bon acteur de cinéma ou de théâtre peut y arriver et c'est peut-être un critère permettant de distinguer le bon acteur du mauvais acteur, mais il joue, il n'est pas réellement spontané<sup>20</sup>.

Les situations présentées forment relativement à l'acte de volition une catégorie mais, à l'intérieur de celle-ci, on peut distinguer plusieurs cas. Tout d'abord, les deux exemples de la volonté de s'endormir et de celle de faire du désordre montrent la nécessité de la distinction à faire entre les situations qui ne concernent qu'une personne et celles qui en concernent plusieurs. Elster donne à cette catégorisation de ce qu'il nomme effets essentiellement secondaires l'explication suivante, concernant le premier type (une seule personne) et commence par la nécessité de penser qu'il y a chez l'individu deux volontés qui se contre-carrent. La première est une volonté dont l'objectif est de perdre sa propre importance jusqu'à l'endormissement conçue ici comme réduction ou perte de volonté. La seconde est celle qui volontairement cherche à faire que la première disparaisse. Comme la première ne disparaît pas, la seconde doit devenir de plus en plus importante. Ce faisant, elle maintient la personne attentive, mobilisant des forces, donc éveillée. Il n'existe très certainement qu'une seule volonté, qui ne peut pas répondre à ce double objectif contradictoire, sachant que l'un porte sur l'objectif à atteindre (l'endormissement), l'autre sur le moyen d'atteindre cet objectif (les moyens, méthodes, stratégies pour parvenir à l'endormissement)<sup>21</sup>.

<sup>20</sup> Sans entrer dans le détail d'une discussion portant sur le talent des acteurs, on pourrait dire que le très bon acteur ressent ce qu'il joue et n'a donc pas besoin de chercher à être spontané, il l'est comme il est le personnage qu'il joue.

<sup>21</sup> Jon Elster, *op. cit.*

Donner l'impression de désordre relève du second type. Il s'agit bien de donner à autrui une impression particulière, ce qui dans le cas présent (mais non dans celui de la spontanéité de l'acteur) nécessite que l'objectif soit autant matériel (le désordre au moins sur le plan formel, l'apparence de désordre) que relationnel : il s'agit de donner à autrui cette impression de désordre. Dans ce cas, on peut analyser le fait de la façon suivante : la personne qui souhaite faire du désordre pour en donner l'impression n'a pas de raison de créer du « vrai » désordre alors qu'elle a des raisons d'en donner l'impression. Aussi, soit elle en fait trop, soit elle en fait trop peu, sur un plan « quantitatif ». En fait, il manque du temps pour que le désordre se fasse comme conséquence d'activités normales et non comme résultat d'une volonté orientée vers ce désordre.

Par ailleurs, le désordre n'étant pas homogène, il y a des endroits qui, plus localement, sont plus hétérogènes et d'autres plus homogènes que l'ensemble. Une analyse en termes de chaos est ici entièrement justifiable. Cela sous-entend qu'il y a un ordre sous-jacent dans le désordre, ordre que l'individu n'est pas capable de créer volontairement mais dont l'absence saute aux yeux de celui qui observe le résultat.

Le cas de la spontanéité est intéressant parce qu'il relève en même temps des deux types précédents. En effet, il s'agit, en même temps, d'être spontané (c'est un état comme l'est le sommeil) et de donner l'impression d'être spontané (impression au même titre que l'impression de désordre). L'impression de spontanéité est bien sujette à une évaluation par autrui donc cela relève du second type. En revanche, être spontané relève du premier type puisque l'évaluateur est aussi la personne elle-même (il n'y a donc, en même temps, qu'une personne en cause). Concernant cette unique personne, il ne peut s'agir de la simple impression, mais bien de la réalité de la spontanéité. Or, s'obligeant à être spontané, il ne peut pas l'être et, s'il le croit, il se ment à lui-même. Le phénomène de mensonge à soi-même réfère aussi à cette double volonté : l'une voulant croire alors que l'autre sait le contraire.

Dans ce domaine de la spontanéité, on trouve chez Pierre Bourdieu et Jean-Claude Passeron toute une analyse de la petite bourgeoisie qui montre l'impossibilité de « faire comme si » on appartenait à la (grande) bourgeoisie. En termes de culture, de choix artistiques, on en fait trop ou trop peu, généralement trop. Il en est de même en ce qui concerne le langage comme révélateur d'appartenance à un groupe :

Pour ne prendre qu'un exemple, des traits distinctifs de la langue des classes moyennes, telle que l'hypercorrection fautive et la prolifération des signes de contrôle grammatical, sont des indices parmi d'autres d'un rapport à la langue caractérisé par la référence anxieuse à la norme légitime de la correction académique : l'inquiétude de la bonne manière, manières de table ou manières de langue, que trahit le langage des petits-bourgeois s'exprime encore plus clairement dans la recherche avide des moyens d'acquérir les techniques de sociabilité de la classe d'aspiration, manuels de savoir-vivre ou guides du bon usage. On voit que ce rapport au langage est partie intégrante d'un système des attitudes à l'égard de la culture qui repose sur la volonté pure de respecter une règle culturelle reconnue plutôt que connue et sur le rigorisme de l'attention à la règle, cette bonne volonté culturelle exprimant en dernière analyse les caractéristiques objectives de la condition et de la position des couches moyennes dans la structure des rapports de classes<sup>22</sup>.

La « référence anxieuse » et l'apparente spontanéité sont antinomiques. La « règle culturelle reconnue plutôt que connue » montre l'insistance qui est portée sur la forme (qui, seule, n'est pas spontanéité) plutôt que sur le fond (la connaissance de la règle qui, si elle est parfaitement intégrée, devient application spontanée). Le « rigorisme » et l'« hypercorrection fautive » sont d'autres façons d'exprimer « en faire trop ».

On peut résumer ce type de situation sous le terme d'« inviolable<sup>23</sup> » pour insister sur le fait que c'est la volonté même qui pose problème.

<sup>22</sup> Pierre Bourdieu, Jean-Claude Passeron, *La reproduction, éléments pour une théorie du système d'enseignement*, Paris, Minuit, 1970, p. 146.

<sup>23</sup> Denis Martouzet, « L'urbaniste est-il en mesure de créer les conditions spatiales de l'urbanité », dans Pierre W. Boudreault et Denis Jeffrey (dir.), *Identités en errance. Multi-identité, territoire impermanent et être social*, Québec, Presses de l'Université Laval, 2007, 93-113.

Le caractère problématique de ces actions est bien lié à la personne : ces actions sont bien du ressort de la personne qui souhaite les atteindre et l'impossibilité à les atteindre ne dépend pas de facteurs extérieurs (des lois de la nature qui empêcheraient d'atteindre tel objectif ou des moyens encore actuellement inexistantes) et ne dépend pas de l'irrationalité de l'acteur qui se fixerait des objectifs qu'il saurait être techniquement impossibles à atteindre (comme courir le 100 mètres en moins de 5 secondes).

Ces « involuables » réfèrent à la personne et celle-ci est considérée comme rationnelle, et pourtant, s'il n'est pas déraisonnable de vouloir dormir ou être spontané, ce n'est pas pour autant rationnel :

- soit dormir est un objectif et la rationalité ne porte, selon les formalisateurs de celle-ci, que sur les moyens, et donc c'est hors du champ de la rationalité,
- soit dormir est un moyen et il est rationnel d'utiliser ce moyen pour être reposé. Le problème porte alors sur les moyens nécessaires à la mise en œuvre de ce moyen (qu'est le sommeil) pour dormir et, là, aucun moyen volontaire n'est possible, et donc tout moyen volontaire pour dormir est donc irrationnel puisque inadapté.

Une définition plus complète est donc la suivante : un « involuable » est ce qui ne peut être obtenu par volonté mais qui peut survenir comme conséquence (non voulue, non pensée) d'un objectif voulu, qui est autre. Le désordre est la conséquence non voulue (non recherchée) d'une activité de travail, d'un cambriolage... : il n'est pas recherché en tant que tel, pour lui-même. C'est ce que Elster appelle « effets essentiellement secondaires ». Si l'on peut préférer le terme d'« involuable » pour insister sur le fait de l'inefficacité de la volonté, l'expression d'Elster insiste sur la survenue des situations recherchées. Cependant, elle peut être confondue avec l'ensemble des actions qui ont des effets pervers<sup>24</sup>. Il s'agit des situations dans lesquelles les conséquences négatives non prévues ou mal estimées d'une action remettent en cause

<sup>24</sup> Raymond Boudon, *Effets pervers et ordre social*, Paris, Presses universitaires de France, 1993.

l'efficacité même de l'action relativement à son objectif premier. La mise en œuvre de moyens visant un objectif qu'Elster appelle effets essentiellement secondaires entraîne bien des effets pervers : il s'agit même d'effets pervers radicaux dans la mesure où ils font totalement disparaître la possibilité d'atteindre l'objectif.

Mais on peut considérer qu'il ne s'agit que d'une analogie puisque les effets pervers tels que définis sont de l'ordre de l'action, comme conséquence de moyens mis en œuvre, alors que les effets essentiellement secondaires sont de l'ordre de la volonté et renvoient notamment à des états mentaux<sup>25</sup>, à moins de considérer, d'une part, que l'acte de volonté est une action comme les autres (et la volonté un moyen comme un autre) et, d'autre part, que Raymond Boudon n'exclut aucun type d'action (y compris les actions non matérielles comme dire et vouloir), alors une situation de type « involuable » n'est qu'un cas particulier d'effets pervers. Ce ne serait alors pas une simple analogie.

Dans la terminologie de Forrester, on parlera, plutôt que d'effets pervers, d'effets contre-intuitifs<sup>26</sup>, soulignant ainsi la responsabilité de l'acteur puisque sont ainsi opposées intuition et raison : s'il y a des effets pervers ou des effets contre-intuitifs, c'est du fait du manque d'observation, d'information ou de réflexion de l'acteur, et non du fait même du type d'action (moyen) ou du type d'objectif poursuivi (fin). On peut donc concevoir la possibilité d'éviter les effets pervers, il suffit de trouver les « bons » moyens pour atteindre l'objectif, si ceux-ci existent; s'ils n'existent pas, ils n'entrent pas pour autant dans la

<sup>25</sup> Il semble néanmoins que l'idée d'« état mental » soit récurrente : il s'agit de forcer une personne à adopter un « état mental » spécifique, qu'elle n'a pas naturellement, « état mental » qui permet tel ou tel comportement, telle ou telle croyance. Concernant la spontanéité, le sommeil, c'est évident : le « forçage » est contradictoire avec l'objectif. Concernant l'impression de désordre ou l'impression de spontanéité, celui qui veut donner cette impression sait que ce n'est pas vrai et lorsqu'il évalue le résultat obtenu, il fait comme s'il était quelqu'un d'autre qui découvrirait ce désordre ou croirait à sa propre spontanéité, faisant comme s'il n'en connaissait pas l'origine. Or, il ne peut oublier volontairement cette origine. Il doit donc en même temps croire quelque chose et son contraire, ce qui pose aussi problème en termes d'« état mental ».

<sup>26</sup> Jay Wright Forrester, *Dynamique urbaine*, Paris, Economica, 1979.

catégorie des « involubles », ils peuvent être seulement infaisables. On peut donc éviter dans certains cas les effets pervers (si l'on s'en donne les moyens et si les moyens existent). *A contrario*, on ne peut vouloir atteindre un « involuble » en modifiant la façon de vouloir (considérée ici comme moyen d'obtenir quelque chose) si l'on considère que la volonté est un tout non contrôlable si ce n'est par elle-même. Dans le cas du sommeil et dans celui de la spontanéité, le problème est accru par le fait que la volonté, en même temps, est moyen (de dormir) et fait partie de l'objectif (réduction de la volonté qu'est le sommeil, comme la spontanéité).

On peut aussi montrer que les situations de type « involubles » sont un cas particulier d'injonctions paradoxales<sup>27</sup> qui se caractérisent par le fait que, dans le cadre d'une communication entre deux individus, langage et métalangage réduisent le champ des possibles à tel point qu'il n'existe plus aucune possibilité matérielle opératoire permettant de répondre à la demande formulée. L'archétype est « sois spontané ». La différence avec les injonctions paradoxales est que les « involubles » sont possibles sans qu'il y ait nécessairement deux personnes. On est alors dans la situation d'injonction paradoxale à soi-même (« endors-toi ! »).

Le cas de ses situations de type « involubles » est révélateur du flou de la délimitation de la notion de rationalité. En effet, si on se place dans le cadre des théories conséquentialistes de la rationalité, il ne s'agit pas de discuter des fins. Il suffit de vouloir telle fin, reste à mettre en œuvre les moyens appropriés. Peut-on pour autant, tout en restant rationnel, se donner des objectifs tels qu'il n'en existe pas de moyens? Il faut ici distinguer le cas où l'acteur ne sait pas qu'il n'existe pas de moyens. Dans ce cas, il n'est pas irrationnel, seulement l'information pertinente lui manque. Cependant, cette situation est transitoire car l'expérience même de chercher à vouloir un « involuble » révèle l'impossibilité de l'obtenir, et donc l'inutilité de la recherche. Le second cas est donc celui où l'acteur sait qu'il n'existe pas de

<sup>27</sup> Gregory Bateson, *Vers une écologie de l'esprit*, Paris, Seuil, 2008 [1972]; Paul Watzlawick, John Weakland et Robert Fisch, *Changements, paradoxes et psychothérapie*, Paris, Seuil, 1975.

moyens d'atteindre l'objectif. Dans ce cas, chercher à l'atteindre est déraisonnable et montre une faiblesse de la théorie économique de la rationalité qui ne peut examiner ce cas, ainsi que toutes les théories conséquentialistes de la rationalité. De même, il semble que de telles situations n'entrent pas dans le cadre des théories sociologiques non conséquentialistes de la rationalité. Ce n'est pas une action logique au sens de Vilfredo Pareto<sup>28</sup> dans la mesure où il n'y a pas coïncidence entre les moyens et la fin recherchée. Si l'on reprend la typologie de Weber, on peut éventuellement considérer que dormir ou être spontané, par exemple, sont des fins « bonnes en soi » (wertrationnel), mais non chercher à dormir, qui comme mise en œuvre de moyens, relèverait plutôt de la Zweckrationalität et ferait retomber la question dans le champ des théories conséquentialistes.

#### La faiblesse de la volonté

Les effets essentiellement secondaires sont l'indice que parfois l'être humain pêche ou, en l'occurrence, veut pêcher, par excès de volonté, car il croit pouvoir le faire, il veut, du côté de la fin visée, ce qui n'est pas « voulable ». Mais, en même temps, il pêche, du côté des moyens à mettre en œuvre, par faiblesse de la volonté : il n'est pas capable de mettre en œuvre une volonté suffisamment forte pour atteindre son objectif. L'akrasie peut être définie comme une faiblesse de la volonté de l'individu<sup>29</sup> dans certaines situations. L'exemple donné habituellement est celui d'Ulysse qui se fait lier au mât de son bateau et donne des ordres

<sup>28</sup> Vilfredo Pareto, *Traité de sociologie générale*, Genève, Droz, 1968 [1917].

<sup>29</sup> En fait, l'akrasie comme défaut de volonté est une interprétation parmi d'autres. Aristote, selon Richard Robinson (Richard Robinson, « L'akrasie selon Aristote », *Revue philosophique*, n° 80, 1955, p. 261-280), donne quatre explications possibles de l'akrasie, quand il soutient contre Socrate que l'akrasie peut exister, fondée sur le savoir (connaissance et mobilisation de la connaissance, oubli) ou sur la volonté. Sans entrer dans les détails, nous choisissons l'interprétation par la volonté plus que par la connaissance car elle ouvre des possibilités de réflexion sur la rationalité qui ne se rapportent pas strictement à l'information, au manque d'information ou au traitement de celle-ci, ce qui renverrait à Herbert Simon. Il ne s'agit pas d'une rationalité limitée mais d'une irrationalité.



pour qu'on ne l'en détache pas, de peur de ne pouvoir résister au chant des sirènes<sup>30</sup>. Cet exemple a plusieurs caractéristiques représentatives de l'akrasie :

- face à deux (ou plus dans d'autres exemples) possibilités (aller vers les sirènes ou ne pas y aller), l'individu a à choisir;
- l'individu évalue les deux possibilités, les compare et les classe, l'une étant meilleure que l'autre. Dans l'exemple cité, on pose que ne pas aller vers les sirènes est mieux qu'y aller;
- malgré ce classement, fait consciemment et sans ambiguïté, il peut arriver que l'individu fasse ce qu'il n'a pas classé en premier.

Mais, nous verrons que cet exemple est plus représentatif de la possibilité pour l'être humain de mettre en place des stratégies de manière à se contraindre soi-même. Il est donc plutôt l'indice de situations face auxquelles la personne peut être irrationnelle par akrasie. Elle n'est pas rationnelle car elle choisit de faire ce qu'elle-même a considéré comme moins bien par rapport à la meilleure possibilité, qu'elle connaît (on suppose que l'évaluation qu'elle a faite des possibles l'a été dans des conditions correctes, notamment en termes d'information quant aux conséquences de chacune des options possibles).

Cet exemple a néanmoins un inconvénient : il laisse entendre que, d'un côté, il y a la raison (sauver l'équipage) et, de l'autre côté, la passion ou l'élan physique (aller vers les sirènes car leur chant est irrésistible). Dans cet exemple, si Ulysse ne s'était pas contraint physiquement lui-même par avance, il n'aurait pas été possible que sa raison lui fasse entendre raison puisqu'elles sont irrésistibles. Il est difficile de le qualifier d'irrationnel puisque la force qui le pousse à faire autre chose est extérieure et infiniment plus grande que les capacités de volonté (cette force s'apparente à un moyen technique qui « tracte » l'individu). Il y a bien faiblesse de la volonté (ou limite), mais non irrationalité. On

<sup>30</sup> Le côté exemplaire de ce cas provient du choix fait par Jon Elster d'intituler l'un de ses essais *Imperfect Rationality: Ulysses and the Sirens* (Jon Elster, op. cit.).

pourrait même estimer qu'il est très rationnel puisqu'il anticipe non sur son changement interne mais sur l'évolution des contraintes externes (la survenue du chant). En revanche de nombreux cas existent où la cause du choix de l'option moins intéressante est, au moins en partie, non extérieure à l'individu et pas aussi irrésistible. Il y a alors faiblesse de la volonté *et* irrationalité.

Le principe de Médée consiste dans le fait que, « selon cette doctrine, une personne peut agir à l'encontre de son meilleur jugement, mais seulement quand une force étrangère l'emporte, elle ou sa volonté<sup>31</sup> ». À l'opposé, se situe le principe de Platon<sup>32</sup>, c'est-à-dire la doctrine de la rationalité pure. L'akrasie se situe entre les deux et, finalement, il peut sembler que l'exemple d'Ulysse tient plus du principe de Médée que de l'akrasie tant la force des sirènes est importante et extérieure.

La définition la plus simple de l'akrasie est donnée par Davidson : « considérez le cas d'une action où l'agent agit à l'encontre de ce qu'il croit, toutes choses considérées, être le meilleur (Aristote appelait un tel comportement "*akrasia*"; on parle aussi d'"incontinence" ou de "faiblesse de la volonté")<sup>33</sup> ». Le « toutes choses considérées » invoqué ici réfère à l'information disponible mais surtout au traitement qui en est fait, c'est-à-dire au calcul coût-avantage, au sens large (y compris éthique) qui caractérise l'homme dans ce qu'il a de rationnel. L'irrationalité provient lorsque la décision porte sur la moins bonne des deux possibilités à envisager dans l'analyse : par simplification de l'analyse, sans que pour autant la réalité soit simplifiée, il suffit de considérer le système de deux actions, celle qui est faite et celle qui est la meilleure. Il n'y pas besoin, dans l'analyse, de mobiliser les autres. Il a jugé du pour et du contre de l'action que, irrrationnellement, il va commettre (soit que le contre est supérieur au pour, soit que le pour et le contre sont globalement inférieurs au pour et au contre de l'action la meilleure). On constate donc qu'il y a du pour à faire l'action moins bonne. Cela signifie qu'il y a

<sup>31</sup> Donald Davidson, *Paradoxes de l'irrationalité*, Combas, Éditions de l'Éclat, 1991, p. 28.

<sup>32</sup> Selon la terminologie de Donald Davidson, *ibid.*, p. 28.

<sup>33</sup> Donald Davidson, *ibid.*, p. 27.

une ou des bonnes raisons de faire telle action moins bonne. Alors que les philosophes analytiques de l'action, à la suite d'Aristote, concluent à l'irrationalité, Raymond Boudon, lui, y voit plutôt remplie la condition suffisante pour qu'il y ait rationalité.

L'origine de l'akrasie est ainsi exposée par Davidson :

la faiblesse de la volonté est analogue à une certaine erreur cognitive, que j'appellerai *faiblesse de la garantie*. La faiblesse de la garantie peut se produire quand une personne a des données à la fois pour confirmer et pour infirmer une hypothèse. La personne juge que, relativement à toutes les données qui lui sont disponibles, l'hypothèse est plus probable que non probable, et pourtant elle n'accepte pas l'hypothèse (ou la force de sa croyance en la vérité de l'hypothèse est plus forte que la force de sa croyance en la vérité de la négation de l'hypothèse [...])<sup>34</sup>. La faiblesse de la garantie a, de toute évidence, la même structure logique (ou plutôt la même structure illogique) que la faiblesse de la volonté; la première implique l'existence d'une croyance irrationnelle formée en présence de données entrant en conflit avec cette croyance, la seconde une intention irrationnelle (et peut-être aussi une action irrationnelle) formée en présence de valeurs entrant en conflit avec cette intention (ou cette action)<sup>35</sup>.

Il y a donc conflit interne. L'origine de ce conflit peut provenir d'une différence entre les cadres de références respectifs dans lesquels est élaboré chacun des raisonnements visant à l'évaluation *ex ante* de l'action : ainsi, j'ai de bonnes raisons d'être généreux dans le cadre d'actions humanitaires par exemple, mais j'ai aussi de bonnes raisons de ne pas l'être si je pense qu'il est mieux de maximiser ma propre utilité. Mais je suis irrationnel à partir du moment où je compare les deux possibilités et choisis

<sup>34</sup> Donald Davidson précise que « le principe normatif contre lequel l'agent a péché (lorsqu'il fait cette erreur cognitive) est ce que Hempel et Carnap appellent le réquisit de la totalité des données disponibles pour le raisonnement inductif [*requirement of total evidence*] : quand nous sommes en train de choisir au sein d'un ensemble d'hypothèses mutuellement exclusives, ce réquisit nous enjoint de donner créance à l'hypothèse qui est la mieux étayée par toutes les données pertinentes disponibles » (Donald Davidson, *ibid.*, p. 48.

<sup>35</sup> Donald Davidson, *ibid.*, p. 48.

celle que, au final, je considère comme la moins bonne<sup>36</sup> des deux, c'est-à-dire que je me laisse aller à mes penchants qui peuvent être *étayés par la raison* plutôt que de *suire ma raison* : le fait d'avoir une ou de bonnes raisons de faire X n'empêche pas d'avoir aussi, en même temps, une ou de bonnes raisons de faire Y à la place de X. La difficulté provient de ce que la raison intervient dans les deux cas, dans les deux processus amenant des décisions opposées alors que la comparaison effectuée suppose un critère (ou une série de critères) unique. On peut voir ainsi le fait que la comparaison de deux actions (ou plus) dans deux cadres (ou plus) différents suppose un méta-critère. Ce méta-critère est donné ou élaboré par la raison. Néanmoins, choisissant la moins bonne solution, c'est comme si je récusais ce méta-critère, et donc ma raison qui l'a élaboré, par un méta-méta-critère qui, lui, récusé ma raison quand elle a évalué la meilleure solution comme étant justement la meilleure possible.

Ces cadres différents peuvent être des cadres temporels : tout bien considéré, il est mieux de réviser des cours pour des épreuves d'examen qui auront lieu dans quelques semaines et finalement choisir d'aller ce soir au cinéma et de repousser au lendemain les révisions, soir après soir. Chez Aristote l'enjeu éthique principal consiste en la dévalorisation du plaisir immédiat pris au détriment d'un profit (plaisir ou autre) plus grand mais survenant plus tard. Nordhaus, ainsi que le souligne Elster, montre l'importance de cette question du temps, notamment en politique : l'idée centrale de son analyse est que les électeurs et les planificateurs accordent toujours une certaine importance au bien-être des générations futures, alors que les politiciens, qui essayent de maximiser le nombre de voix qu'ils obtiendront lors de la prochaine élection, ne s'intéressent qu'au présent<sup>37</sup>.

<sup>36</sup> Il ne s'agit pas d'entrer ici dans le détail de ce qui amène à évaluer la qualité relative à chacune des possibilités : chaque acteur à ses propres critères et nous nous positionnons dans l'examen de l'auto-cohérence de la personne qui ne dépend pas de la nature des critères d'évaluation choisis.

<sup>37</sup> Jon Elster, *op. cit.*, p. 131.

## II – Contrainte à soi-même, engagement et réécriture du passé

La contrainte à soi-même selon Elster

L'exemple d'Ulysse montre la capacité de l'être humain de s'imposer soi-même des contraintes. De peur de faire quelque chose, je m'arrange préalablement pour que ce quelque chose ne puisse se faire, ou bien soit moins intéressant ou bien que je me trouve dans une situation telle que mon état d'esprit soit moins porté à faire ce quelque chose. Cette contrainte à soi-même révèle là encore soit un problème lié à la volonté, soit lié à la rationalité humaine. Pourquoi se contraindre soi-même, physiquement par exemple, comme Ulysse, si ce n'est parce que ma volonté ne sera pas assez forte pour résister? Mais, en même temps, pourquoi se contraindre face à quelque chose que finalement on souhaite (et on sait qu'on le souhaite puisque préalablement on se contraint) : on souhaite donc quelque chose que ne souhaite pas notre raison. Il y a donc irrationalité, telle que définie dans l'akrasie.

De cette constatation, certains, dont Jean-Jacques Rousseau, ont tiré des principes moraux ou éthiques, consistant, par avance, à ne pas se mettre dans des situations, même intéressantes, telles que je pourrais avoir envie de faire quelque chose que ma morale personnelle réprouverait.

J'en ai tiré cette grande maxime de morale, la seule peut-être d'usage dans la pratique, d'éviter les situations qui mettent nos devoirs en opposition avec nos intérêts, et qui nous montre notre bien dans le mal d'autrui, sûr que dans de telles situations, quelque sincère amour de la vertu qu'on y porte, on faiblit tôt ou tard sans s'en apercevoir, et l'on devient injuste et méchant dans le fait, sans avoir cessé d'être juste et bon dans l'âme<sup>38</sup>.

D'une façon générale, les contraintes à soi-même entrent dans la catégorie des actions qui supposent de la part de l'acteur, mais de façon non nécessairement consciente, une capacité d'anticipation et, plus particulièrement, d'anticipation sur lui-même et de stratégie (sa conception et sa mise en œuvre). Mais on peut préciser encore cela, et Elster énonce cinq conditions pour qu'une

<sup>38</sup> Rousseau Jean-Jacques, *Les confessions*, Paris, Gallimard et Librairie Générale Française, 1963, tome 1, p. 96.

action soit une contrainte à soi-même volontairement consentie. Il considère que ces cinq conditions ne sont peut-être ni nécessaires ni suffisantes pour définir la contrainte à soi-même, mais comme première approximations elles peuvent convenir. Nous les discutons ensuite. Premièrement : « s'imposer une contrainte à soi-même, c'est prendre une certaine décision à un instant  $t_1$  afin d'accroître la probabilité que l'on prendra une certaine décision à un instant  $t_2$ <sup>39</sup> ». Cela signifie que la prise de la première décision doit être volontaire et consciente, d'une part, et, d'autre part, orientée vers une certaine finalité relative à la deuxième décision, qui est prise par la suite. Cela exclut donc les contraintes imposées par d'autres ou par les événements que l'on ne maîtrise pas de même que des contraintes que l'on se serait imposées sans savoir que ce sont des contraintes ou que leurs conséquences en soient. Prendre une décision contraignante pour soi suppose :

- de savoir que l'instant  $t_2$  de la deuxième décision interviendra quel que soit le contenu de  $t_1$ <sup>40</sup>. Autrement dit, la décision prise ou non en  $t_1$  ne modifie pas radicalement la situation en  $t_2$  puisque celle-ci nécessite toujours de prendre une décision mais la modifie suffisamment pour que le choix effectué dans cette deuxième décision soit différent de ce qu'il aurait été sans cela.
- que l'on sache, au moins approximativement, les possibilités offertes en  $t_2$ , qu'il y ait ou non décision en  $t_1$ .
- que l'on sache dès  $t_1$  les différences au moins approximatives entre le choix des possibles en  $t_2$  avec et sans contrainte à soi-même en  $t_1$ .

Il s'agit nécessairement de probabilités subjectives : il suffit de ne pas savoir avec certitude qu'une occurrence sûre soit sûre pour qu'elle ne soit que, subjectivement parlant, plus ou moins probable. Cela signifie par ailleurs qu'il peut y avoir des contraintes

<sup>39</sup> Jon Elster, *op. cit.*, p. 105.

<sup>40</sup> Par instant  $t_2$ , il ne faut pas voir juste un moment sur un axe de mesure temporel, mais bien le contenu de ce moment, la situation qu'il est, c'est-à-dire une situation caractérisable qui nécessite une prise de décision même si aucune décision n'a été prise en  $t_1$ .

à soi-même inefficaces, c'est-à-dire soit qu'elles n'augmentent pas la probabilité de faire X en  $t_2$ , soit qu'elles ne l'augmentent pas assez. Finalement le choix porte sur Y alors que la contrainte à soi-même avait pour but de ne pas amener Y, en amenant X.

La deuxième condition énoncée par Elster est la suivante : « si l'action à l'instant  $t_1$  a pour effet d'induire un changement dans l'ensemble des options qui seront disponibles à l'instant  $t_2$ , alors, si le nouvel ensemble contient l'ancien, il ne s'agit pas de contrainte imposée à soi-même<sup>41</sup> ». Elster précise que « la condition 2 dit que l'élargissement de cet ensemble n'est pas un cas d'application de cette stratégie. Il doit y avoir *exclusion* de certaines options pour que l'on puisse parler d'engagement préalable<sup>42</sup> » ou de contrainte à soi-même.

La condition 3 précise que « la décision prise à l'instant  $t_1$  doit avoir pour effet d'établir un processus causal dans le monde extérieur<sup>43</sup> ». « Ce dont il est question ici, c'est de modifier l'environnement *dans le but* de maîtriser son propre comportement<sup>44</sup>. » Une contrainte à soi-même ne peut être une modification de l'acteur lui-même comme pourrait l'être une prise de résolution ou une volonté de fermeté vis-à-vis de la volonté, qui, de toute façon, peuvent disparaître en  $t_2$  ou même avant. La contrainte à soi-même est différente d'une autocontrainte immédiate : la contrainte à soi-même passe par l'extérieur et revient par une sorte d'effet boomerang. L'individu s'arrange en  $t_1$  pour que l'environnement, dans la durée entre  $t_1$  et  $t_2$ , lui ôte une partie de son pouvoir d'action sur l'environnement de manière à ce que  $t_2$  présente une situation telle que le choix qui aurait été fait en  $t_2$  en l'absence de contrainte à soi-même soit celui à cause duquel la contrainte à soi-même a été mise en œuvre.

<sup>41</sup> Jon Elster, *op. cit.*, p. 108.

<sup>42</sup> Jon Elster, *ibid.*, p. 108. C'est Elster qui souligne. Le terme d'engagement serait ici inapproprié s'il n'était suivi de l'adjectif préalable et correspond alors à *precommitment* et non à *commitment* qui est l'engagement dans l'action et non préalable à l'action.

<sup>43</sup> Jon Elster, *ibid.*, p. 109.

<sup>44</sup> Jon Elster *ibid.*, p. 110. C'est Elster qui souligne.

La condition 4 semble aller de soi, il faudrait peut-être la nuancer en fonction des capacités d'ordre psychologique des moments de prises de décision : « La résistance qu'il faut vaincre pour s'imposer à l'instant  $t_1$  une contrainte visant à maîtriser son propre comportement à l'instant  $t_2$  doit être moindre que la résistance qu'il aurait fallu vaincre pour maîtriser son comportement à l'instant  $t_2$  si l'on ne s'était pas imposé à l'avance cette contrainte à l'instant  $t_1$ <sup>45</sup> ».

En revanche la cinquième condition reste discutable, mais elle intervient peu en ce qui concerne la question de la rationalité : « une contrainte imposée à soi-même doit résulter d'un engagement et non d'une abstention<sup>46</sup> ».

### Interprétation

Ce sont surtout les première et deuxième conditions qui posent problème. Tout d'abord, la formulation de la première condition contient une définition implicite de la contrainte à soi-même et, par suite, exclut une autre définition possible. En effet, pour Elster, s'imposer une contrainte à soi-même a pour objectif d'accroître la probabilité de l'option qui est, tout bien considéré, la meilleure. On peut au contraire soutenir qu'une contrainte à soi-même a pour objectif de diminuer la probabilité de l'option qui, bien que n'étant pas la meilleure, risque d'être choisie en  $t_2$  s'il n'y a pas cette prise de décision en  $t_1$ . Or, il semble que :

- s'imposer une contrainte à soi-même se fait par un raisonnement relevant de la deuxième définition plus que de la première, comme l'indique par exemple, la morale rousseauiste;
- malgré l'apparence des formules, il n'y pas symétrie entre elles : l'une n'est pas strictement le négatif de l'autre dès que le choix des options en  $t_2$  ne se réduit pas à deux options mutuellement exclusives.
- Si l'on considère la définition implicite d'Elster, alors effectivement la condition 2 est valide : l'exclusion d'un

<sup>45</sup> Jon Elster *ibid.*, p. 111.

<sup>46</sup> Jon Elster *ibid.*, p. 113.



certain nombre d'options suffit, sous certaines conditions, à augmenter la probabilité d'une option en particulier, celle qui est tout bien considéré la meilleure, notamment si la contrainte à soi-même empêche ou réduit la probabilité de choisir en  $t_2$  l'option qu'on aurait prise sans la contrainte découlant de  $t_1$ .

Si on prend le point de vue de l'économie standard, la contrainte à soi-même n'exclut pas au sens propre une possibilité parmi d'autres mais la seule possibilité de l'*homo œconomicus*, celle qu'il aurait choisi en  $t_2$  s'il n'y avait pas eu contrainte à soi-même en  $t_1$ , par une autre et une seule possible, celle que l'*homo œconomicus* va choisir étant donné qu'il y a eu contrainte à soi-même en  $t_1$ . Dans ce cas, la condition  $t_2$  n'a pas de sens puisque le problème essentiel de l'*homo œconomicus* est que, même en état d'incertitude, son action est prévisible : il n'y a donc pas choix puisqu'il n'y a qu'une option envisageable. Cependant, d'une part, un *homo œconomicus*, dans sa version stricte, ne peut et n'a pas besoin de se donner de contrainte à soi-même et, d'autre part, Elster ne se positionne pas dans le cadre de la rationalité standard, il doit donc modifier la portée de sa condition 2 ou choisir explicitement une définition de la contrainte à soi-même telle que nous la donnons.

Il semble que le fait de prendre cette deuxième définition permette plus de précision quant aux cas possibles de contrainte à soi-même. Considérons que A est l'option choisie (et en même temps crainte) s'il n'y a pas décision contraignante en  $t_1$ , que B, C, D... sont des options plus ou moins intéressantes mais non rejetées comme A peut l'être par exemple pour des raisons morales. Les probabilités que soient choisies A, B, C et D sont, s'il n'y a pas décision en  $t_1$  :  $P(A) = 40$ ,  $P(B) = 35$ ,  $P(C) = 25$ ,  $P(D) = 0$  (D n'apparaît pas s'il n'y a pas décision en  $t_1$ ). On peut alors établir le tableau 1 suivant, présentant les cas possibles en ce qui concerne les conséquences d'une prise ou d'une non-prise de décision en  $t_1$ . Dans le tableau, les deux colonnes de droite indiquent, selon la définition choisie, la présence ou non d'une contrainte à soi-même.

Tableau 1

Conséquences de la prise de décision en $t_1$		Commentaires	CSM selon Elster	CSM selon l'autre définition
Exclusion de A. Diminution du champ des possibles	$P(B)=1$ ou $P(C)=1$	La situation en $t_2$ n'est plus un choix, $t_1$ n'amène donc pas une contrainte à soi-même mais <u>est</u> la décision qui aurait dû avoir lieu en $t_2$ .	oui	non
	$P(B)=0,45$ $P(C)=0,55$	La situation en $t_2$ reste un choix entre B et C. La disparition de A augmente les probabilités de B et de C, de façon non égale mais en changeant l'ordre des probabilités entre B et C.	oui	oui
	$P(B)=0,55$ $P(C)=0,45$	La situation en $t_2$ reste un choix entre B et C. La disparition de A augmente les probabilités de B et de C, de façon non égale mais sans changer l'ordre des probabilités entre B et C.	oui	oui
Exclusion de A. Diminution et augmentation du champ des possibles	$P(B)=0,45$ $P(C)=0,35$ $P(D)=0,20$	La situation en $t_2$ reste un choix mais entre B, C et D, sans changer l'ordre entre B et C. D reste peu probable.	?	oui
	$P(B)=0,35$ $P(C)=0,15$ $P(D)=0,50$	La situation en $t_2$ reste un choix mais entre B, C et D, et D devient la plus probable des options.	?	oui
Maintien de A. Diminution du champ des possibles	$P(A)=0,55$ $P(B)=0,45$	A reste la plus probable, la contrainte est inefficace et la décision prise en $t_1$ augmente même la probabilité de choisir A.	non	non
	$P(A)=0,40$ $P(B)=0,60$	La probabilité de A reste constante mais l'action faite en $t_1$ rend B plus probable, du fait de la disparition de C.	oui	oui
	$P(A)=0,30$ $P(B)=0,70$	Même cas que le précédent, mais la probabilité de A diminue. L'action faite en $t_1$ rend plus B probable, du fait de la disparition de C.	oui	oui
	$P(A)=0,45$ $P(B)=0,55$	Même cas que le précédent, mais la probabilité de A augmente. Cependant, l'action faite en $t_1$ rend plus B probable que A.	oui	oui
Maintien de A. Augmentation du champ des possibles	$P(A)=0,25$ $P(B)=0,35$ $P(C)=0,20$ $P(D)=0,20$	L'apparition de D provoque une redistribution des probabilités : B devient plus probable que A.	non	oui
	$P(A)=0,35$ $P(B)=0,25$ $P(C)=0,25$ $P(D)=0,15$	A reste la plus probable, bien que sa probabilité diminue. La contrainte est inefficace.	non	non
Maintien de A. Constance du champ des possibles	$P(A)=0,10$ $P(B)=0,40$ $P(C)=0,50$	La décision en $t_1$ modifie les probabilités de A, B et C. Ici celle de A diminue suffisamment pour qu'il y ait contrainte.	non	oui
	$P(A)=0,40$ $P(B)=0,50$ $P(C)=0,10$	La décision en $t_1$ modifie les probabilités de A, B et C. Ici celle de A ne diminue pas mais celle de B augmente suffisamment pour qu'il y ait contrainte.	non	oui
	$P(A)=0,50$ $P(B)=0,40$ $P(C)=0,10$	La décision en $t_1$ modifie les probabilités de A, B et C. Ici celle de A augmente et celle des autres n'augmentent pas assez pour qu'il y ait contrainte.	non	non
	Maintien de A Diminution et augmentation du champ des possibles	$P(A)=0,20$ $P(C)=0,50$ $P(D)=0,30$	La décision en $t_1$ provoque l'apparition de D, ce qui augmente la probabilité de C et diminue celle de A, suffisamment pour qu'il y ait contrainte.	?
$P(A)=0,20$ $P(C)=0,30$ $P(D)=0,50$		La décision en $t_1$ provoque l'apparition de D, ce qui diminue la probabilité de A, suffisamment pour qu'il y ait contrainte.	?	oui
$P(A)=0,40$ $P(C)=0,30$ $P(D)=0,30$		La décision en $t_1$ provoque l'apparition de D, mais A reste l'option la plus probable.	non	non

La formulation d'Elster ne semble pas adéquate dans la mesure où elle est trop restrictive. En effet, contrairement à ce qu'il énonce, il se peut que le nouvel ensemble contienne l'ancien, si le changement induit par la décision en  $t_1$  modifie les probabilités d'occurrence des différentes possibilités et la valeur relative de ces différentes possibilités et qu'ainsi le choix final soit différent de

ce qu'il aurait été s'il n'y avait pas eu décision de se contraindre soi-même en  $t_1$ . L'ambiguïté réside dans le terme d'option : au sens strict il s'agirait des possibilités et au sens large des possibilités affectées de leur valeur en  $t_2$  et de la probabilité d'occurrence. C'est sans doute en ce second sens que doit être comprise la seconde condition. De plus, non seulement le nouvel ensemble peut être égal, plus restreint ou plus large que l'ancien, mais il peut être en même temps plus large (« d'un côté ») et moins large (« d'un autre côté ») : il exclut des possibilités mais en ajoute d'autres.

### L'engagement

La théorie du choix rationnel repose sur l'idée somme toute raisonnable que l'homme ne peut changer que le futur. En tout cas, s'il croit pouvoir changer des situations, par rapport à ce qu'elles seraient s'il ne cherchait pas à les changer, ces situations ne peuvent être que futures, avec comme cas limite les situations présentes en admettant l'immédiateté des effets d'une cause présente, comme dans le cas des énoncés performatifs<sup>47</sup>. Que l'on adhère à cette croyance d'un pouvoir de changement, qu'elle soit vraie ou fausse, suffit par contrepoint, pour adhérer à l'idée que le passé ne peut pas être changé.

Pour que l'homme soit rationnel, il est nécessaire d'adhérer au principe de fixité du passé. Ce principe ne signifie pas que l'homme ne puisse être déterminé au moins en partie par le passé, mais celui-ci reste fixe. La détermination de l'acteur actuel par le passé passe par l'information, les habitudes, les routines... Mais, le calcul coûts/avantages face à une situation de choix ne doit pas intégrer les coûts et les avantages déjà réalisés. Les coûts et avantages passés peuvent donner une idée de stratégie ou une information sur les coûts et avantages à venir mais n'entrent pas dans le calcul de ceux-ci. Seul compte ce que je peux attendre, dans le futur de mon choix présent pour l'avenir.

Tout cela semble aller de soi. Or, l'être humain n'envisage pas toujours les choses de cette façon, comme en témoignent toutes

<sup>47</sup> John Langshaw Austin, *Quand dire, c'est faire*, Paris, Seuil, 1970.

les actions qui relèvent de la vengeance. En effet, l'acte vengeur entraîne des coûts qui peuvent être, qui sont généralement, plus élevés que les gains attendus, y compris si on y ajoute la satisfaction directe liée à la vengeance. Il semblerait même que le gain attendu soit finalement moins élevé et la vengeance décevante. La littérature occidentale regorge de ces exemples de vengeances qui ne satisfont pas, qui déçoivent le vengeur et dont l'archétype est donné par le *Colomba* de Mérimée. Cependant, même s'il y a erreur dans le calcul du gain attendu de la vengeance, on doit bien convenir qu'il y a risque d'addition du coût lié à l'acte initial et ses conséquences et coût lié à la vengeance envers celui qui a réalisé l'acte initial et dont on a subi les conséquences. L'acte de vengeance est coûteux : si on suppose un acte de meurtre pour cause de vengeance, le seul gain est la satisfaction liée au fait de laver l'affront précédent et, à l'opposé, il y a risque important de punition lourde et, de plus, l'acte censé venger est coûteux puisque ce même acte n'aurait généralement pas été fait pour un autre motif, y compris pour un gain direct, financier par exemple. Il y a ici une forme d'akrasie puisque, tout bien considéré, le vengeur sait qu'il vaut mieux en certaines circonstances ne pas se venger et le fait quand même. Si ce risque est important, pourquoi n'est-il pas pris en compte? D'après le sophisme de l'amortissement, cela provient du fait de l'idée que l'acte vengeur va réduire le coût lié à l'acte initial. Ces coûts peuvent être scindés en deux catégories : ceux que l'on a subis et ceux que l'on va subir. Or, les premiers étant passés, ils ne sont pas modifiables et, généralement, les seconds ne sont pas modifiés par l'acte vengeur. Si l'on prend l'exemple de quelqu'un qui veut tuer une personne qui a tué quelqu'un de cher, la peine de la perte de l'être cher jusqu'à présent est un coût inchangeable et la peine à venir de la perte de l'être cher ne sera pas changée par l'acte vengeur de l'élimination de la personne qui a été cause de la peine. On peut analyser cela de deux façons : soit la personne qui se venge pense pouvoir modifier le passé (par un acte réduisant le coût de la peine), soit la personne qui se venge utilise un moyen (la vengeance) totalement

inadéquat par rapport à la fin (réduire sa peine). Non seulement, le moyen est inadéquat mais il est aussi coûteux.

Est-ce pour autant irrationnel? Au sens de la théorie du choix rationnel, oui mais pas nécessairement dans un sens plus large de la rationalité. En effet on peut penser que l'erreur relative au calcul gain/coût de l'acte vengeur est liée au manque d'information que l'on peut attendre vraiment de cet acte vengeur. Il n'y aurait donc pas d'irrationalité mais manque d'information.

On peut, par ailleurs, inclure dans ce calcul les coûts et les gains relatifs à la notoriété. Une société où l'honneur, dans le sens classique du terme, est une valeur socialement partagée, peut conduire à l'idée que ne pas se venger ajoute un coût au coût de l'acte initial, sans que pour autant se venger puisse réduire le coût de cet acte initial. Cependant, la vengeance existe dans les sociétés d'honneur comme dans les autres.

Enfin, on peut inclure dans le calcul des coûts et des gains le coût psychologique de l'incohérence qu'entraîne l'idée d'avoir été victime et de ne pas se venger qui pourrait être interprétée comme une possibilité donnée aux autres que je peux être, voire que je souhaite être encore victime sans risque pour l'autre. Il y a là à la fois l'idée de réduction de dissonance cognitive (personnelle et vis-à-vis des autres) et l'idée de donner une image de soi telle que je ne puisse être plusieurs fois victimes : je me forge une notoriété dans le cas où une situation semblable se reproduirait. Ce faisant, dans ce cas, ce serait complètement ignorer les phénomènes d'escalade de la violence, dont l'archétype est la *vendetta* qui résulte du fait que la vengeance dans le but d'arrêter la situation-problème contribue à donner des raisons à d'autres de la faire prolonger. En d'autres termes, au lieu de s'annuler, les coûts s'additionnent et demandent donc une suite de plus en plus importante.

La vengeance n'est qu'un cas particulier de sophisme de l'amortissement incluant plusieurs personnes et dont la charge émotionnelle des situations est très forte, trop peut-être pour pouvoir être analysée en termes de rationalité.

Il faut considérer, en dehors de tout aspect comptable, qu'un bien ou un service une fois payé est d'ores et déjà amorti et que si l'achat s'est révélé être une erreur on ne peut rien y faire (si ce n'est limiter les dégâts en tentant de revendre l'objet). Cependant, l'exemple que donne Dupuy de l'achat d'un objet coûteux qui se révèle décevant (en l'occurrence une télévision), qui ne sera pas revendu (en tout cas pas rapidement) ni relégué à la cave et qui sera malgré tout utilisé<sup>48</sup>. On voit dans cet exemple que la charge émotionnelle est moins importante qu'en ce qui concerne la vengeance. Selon Dupuy, l'acheteur déçu s'efforcera d'utiliser quand même l'objet, malgré le désagrément que cette utilisation amène, désagrément supérieur au plaisir lié à l'utilisation (il peut y avoir en même temps plaisir relatif et désagrément relatif, notamment du fait que le temps passé à l'utilisation de l'objet pourrait être mieux utilisé en une activité procurant une satisfaction plus grande ou un désagrément plus faible). Se forçant à utiliser l'objet, il cherche à se convaincre qu'il a bien fait de l'acquérir. La preuve en est qu'il l'utilise. Ce qui est le plus surprenant est qu'en même temps il sait avoir fait un mauvais choix *et* il veut se faire croire qu'il a fait un bon choix, mais surtout qu'il veut justement se faire croire qu'il a fait un bon choix *parce qu'il* sait qu'il en a fait un mauvais. Nous verrons plus loin ce que signifie cette incohérence, en termes de mauvaise foi.

Ici on peut considérer qu'à chaque fois qu'il utilise l'objet, c'est pour se convaincre qu'il avait bien fait de l'utiliser la fois précédente. Le raisonnement mis en œuvre ici est du type de celui d'une démonstration par récurrence (la propriété P<sub>n</sub> est vraie pour tout n si, en admettant qu'elle est vraie pour n'importe quel n en particulier, on peut montrer qu'elle est vraie pour n + 1). Mais si cette démonstration est valide en mathématique, elle pose problème pour l'analyse d'une pratique répétitive. En effet, en mathématiques, la démonstration « va » de n - 1 à n, à n + 1, à n + 2... Dans le cas pratique non mathématique, c'est n - 1 qui

<sup>48</sup> Jean-Pierre Dupuy, « Temps et rationalité », dans René Frydman (dir), *Quelles hypothèses de rationalité pour la théorie économique?*, Cahiers d'économie politique, n<sup>os</sup> 24-25, 1994, p. 69-104.

devient vrai parce que n « l'est »<sup>49</sup> même si globalement on avance dans le temps, de n vers p (p étant après n) à chaque étape du raisonnement, la démarche est inverse et « va » de n à n - 1. La première utilisation de l'objet ne permet pas de justifier *a posteriori* l'achat, au contraire elle révèle l'erreur qu'est cet achat. Autrement dit on a un « escalier » très bien construit mais dont la première marche ne tient absolument pas et révèle la faiblesse et l'inutilité de cet « escalier ». Cependant, sur un plan plus psychologique, il faut y ajouter le facteur temps : lorsque n est très grand, validant n - 1 qui valide n - 2, n = 1 est très éloigné et peut ne plus être à l'esprit.

Cette prise en compte du passé alors qu'elle ne devrait pas être apparaît dans de nombreux cas. Ainsi toute forme de pré-engagement relève de cette catégorie et s'avère être un sophisme. En effet, le fait que je m'engage maintenant pour plus tard ne m'oblige que moralement mais non rationnellement à faire, nécessairement plus tard, après le moment de l'engagement, ce à quoi je m'étais engagé, notamment dans le cas où finalement je pense avoir fait une erreur en m'engageant. De ce fait, on peut déduire qu'un engagement n'est qu'une coquille vide, si on lui ôte le caractère moral, car si ce à quoi je me pré-engage est « bien », étant bien je le ferai et n'ai donc pas besoin de me pré-engager à le faire car j'ai de bonnes raisons de le faire et si ce à quoi je me pré-engage est « mal » alors je ne le ferai pas (ayant de bonnes raisons de ne pas le faire) puisque je ne suis pas lié par mon engagement.

On peut considérer que la valeur morale de l'engagement suffit, en référence à la Wertrationalität de Weber pour qui tenir un engagement préalablement pris est bien en soi, quel que soit la nature et le contenu de l'engagement. Cependant, si en droit français et dans la très grande majorité des autres systèmes juridiques, la rupture de l'engagement ou rupture de contrat est assortie d'un surcoût (paiement des dommages et intérêts par exemple, ou versement d'une prime de licenciement lors de la

<sup>49</sup> En fait, n *devient* vrai dans l'action.

rupture d'un contrat de travail), ceci montre bien que la seule valeur morale n'est pas suffisante pour faire tenir un engagement.

Là où c'est plus intéressant, c'est lorsque l'on suit Éric Brousseau pour qui les contrats de type interentreprises sont nécessairement insuffisants pour que l'engagement contenu dans et par le contrat soit tenu. Or, ils le sont très généralement et, en tout cas, pendant un certain temps.

Parce que sa signification réelle – la manifestation de la volonté de coopérer – dépasse sa signification apparente – l'aménagement d'un mécanisme incitatif imparfait – le contrat a la force d'un symbole. Les agents vont respecter l'esprit de l'engagement qu'ils ont pris les uns envers les autres alors même que le mécanisme formel qui est décrit ne les y oblige pas. D'une part, la structure d'incitation est imparfaite : elle ne prévoit pas de sanctionner certains comportements pourtant collectivement sous-optimaux. D'autre part, les engagements ne sont pas nécessairement auto-exécutoires car le système judiciaire ne permet pas de faire exécuter l'intégralité du contrat qui est en grande partie tacite. Mais les agents vont faire « comme si » le contrat qui a été passé était à la fois incitatif et exécutoire. Autrement dit, *ils prennent le contrat pour ce qu'il n'est pas – un mécanisme garantissant l'élimination des comportements opportunistes – tout en sachant que ce n'est pas le cas*<sup>50</sup>.

Tenir un engagement pour raison morale n'est pas rationnel au sens de la théorie du choix rationnel mais il ne l'est pas non plus, même lorsqu'il s'agit d'un contrat pour des raisons relatives au coût et au gain. Un engagement est donc non rationnel si on y enlève les coûts et les gains « en valeur », de notoriété, de prestige, de réputation. L'engagement est un performatif qui entraîne une croyance autovérifiante dans la valeur de l'engagement et cette croyance est autovérifiante par des effets d'agrégation : les coûts en notoriété, prestige, valeur n'existent pour moi que parce que les autres y croient. Si les autres n'accordaient aucune valeur à ces coûts, je n'aurais pas à y accorder de valeur. Les autres accordant de la valeur à la notoriété, je suis obligé de lui en accorder. On remarquera l'utilité sociale de cette croyance

<sup>50</sup> Éric Brousseau, « Contrats et comportements coopératifs : le cas des relations interentreprises », dans Jacques-Laurent Ravix (dir.), *Coopération entre les entreprises et organisation industrielle*, Paris, Éditions du CNRS, 1996, p. 38. C'est nous qui soulignons.



autovérifiante : si aucun engagement n'était possible, aucune forme de contrat social ne pourrait exister, ni donc aucune forme de vie commune.

Une autre forme d'engagement est l'abonnement. Dupuy<sup>51</sup> rapporte l'exemple donné par Robert Nozick<sup>52</sup> qui s'oblige à aller au théâtre malgré les coûts que cela entraîne et parce qu'il est convaincu que, tout bien considéré, il a raison d'aller au théâtre, en prenant un abonnement. L'ayant pris, il se sent obligé d'y aller pour rentabiliser son abonnement. Ce raisonnement irrationnel est néanmoins fréquent : le calcul permettant de dire que je dois aller cinq fois au théâtre dans l'année pour que l'abonnement qui coûte X et permet des réductions de Y % sur le billet d'entrée normal est vrai. Il est vrai avant le paiement de l'abonnement et intervient dans la décision de s'abonner à condition de savoir que la probabilité que j'y aille cinq fois ou plus dans l'année à venir est suffisamment élevée (soit que les années précédentes j'y sois allé plus ou beaucoup plus de cinq fois, soit que, bien que j'y aille en moyenne de quatre à six fois, j'estime que j'irai plus souvent parce que cette année la programmation m'apparaît meilleure). Ce raisonnement reste nécessairement vrai après la décision mais il est devenu inutile car inadéquat pour décider si ce soir je vais au théâtre. Que j'aie payé l'abonnement où non n'intervient pas si ce n'est que je peux estimer que la pièce jouée (dans le contexte général de la situation) « mérite » le prix du billet réduit ou le prix du billet normal. L'abonnement peut donc inciter à aller au théâtre dans la mesure de la réduction de prix ainsi permise, non y obliger : il ne m'engage pas à y aller.

Il y a dans ce cas une forme d'akrasie double : d'abord, je sais qu'il est mieux d'aller au théâtre et pourtant je n'y vais pas. Je me forge alors une contrainte à moi-même (l'abonnement). Si je crois sincèrement que l'abonnement me contraint, alors l'erreur est efficace puisque finalement je fais ce que j'estime le mieux de faire, c'est-à-dire aller au théâtre. Cependant, si je sais que

<sup>51</sup> Jean-Pierre Dupuy, *op. cit.*, p. 74-75.

<sup>52</sup> Robert Nozick, *The Nature of Rationality*, Princeton (NJ), Princeton University Press, 1993.

l'abonnement ne me contraint pas vraiment, je peux, par une seconde forme d'irrationalité<sup>53</sup> me sentir obligé de me sentir contraint. Je crois (ou je fais semblant de croire) que la contrainte est efficace alors qu'elle ne l'est pas mais le simple fait de la croire efficace la rend efficace, puisque je la suis.

La question primordiale posée derrière ces exemples et analyses, et relativement à la question de la prise en compte du temps, est de savoir si un engagement ou, d'une façon générale, une situation passée, nous lie les mains. Il semble, en tout cas à travers le sophisme de l'amortissement, que le passé ne nous lie pas les mains mais que nous croyons que le passé nous lie les mains et que, cette croyance étant autovérifiante, le passé nous lie les mains.

Le sophisme de l'amortissement est ici analysé en termes de réécriture du passé qui pourrait découler d'une sorte d'oubli de ce qu'il a été, à l'instar de l'interprétation aristotélicienne de l'akrasie. Cependant, on peut y voir aussi une analyse en termes d'irrationalité pour laquelle le sophisme de l'amortissement ne serait qu'un cas particulier de la mauvaise foi.

### III – Mauvaise foi et structure de base de la complexité

La structure logique de la mauvaise foi

Selon la définition donnée par Sartre, la mauvaise foi est une catégorie très particulière de mensonge : elle a « en apparence la structure du mensonge. Seulement ce qui change tout, c'est que dans la mauvaise foi, c'est à moi-même que je masque la vérité<sup>54</sup> ». Il ne s'agit donc pas de conférer à cette expression le sens habituel du langage courant (« malhonnêteté de quelqu'un qui affirme des choses qu'il sait fausses ou qui feint l'ignorance » qui s'oppose en partie à la bonne foi considérée comme « attitude de quelqu'un qui parle ou qui agit avec la conviction d'être honnête, de respecter la vérité », ce qui n'exclut pas l'erreur, mais seulement

<sup>53</sup> Cette seconde irrationalité relève plus de la mauvaise foi que de l'akrasie puisqu'elle porte sur des croyances et non sur des actions. Ce n'est pas une faiblesse de la volonté mais plutôt une faiblesse de la garantie.

<sup>54</sup> Jean-Paul Sartre, *op. cit.*, p. 84.

l'erreur faite exprès, comme l'indique la définition juridique de la bonne foi : « croyance erronée en l'existence d'un droit ou d'une règle juridique, par ignorance ou suite à une tromperie<sup>55</sup> ». La définition de Sartre ne correspond donc pas à celle du langage commun et même s'y oppose puisque celui qui fait preuve de cette mauvaise foi sartrienne le fait en toute bonne foi. « Le véritable problème de la mauvaise foi vient évidemment de ce que la mauvaise foi est foi<sup>56</sup>. » En effet, « la dualité du trompeur et du trompé n'existe pas ici<sup>57</sup> ».

Il s'ensuit d'abord que celui à qui l'on ment et celui qui ment sont une seule et même personne, ce qui signifie que je dois savoir en tant que trompeur la vérité qui m'est masquée en tant que je suis trompé. Mieux encore, je dois savoir très précisément cette vérité *pour* me la cacher plus soigneusement – et ceci non pas à deux moments différents de la temporalité – ce qui permettrait à la rigueur de rétablir un semblant de dualité – mais dans la structure unitaire d'un même projet<sup>58</sup>.

Des exemples paraissent ici nécessaires pour mieux comprendre ce dont il s'agit. Sartre donne celui de la femme qui est en train de se laisser séduire par un homme dont elle connaît très bien l'objectif ou au moins la dimension physique de cet objectif et parvient en même temps à croire ce qu'il dit et qui est très respectueux dans la forme<sup>59</sup>. La mauvaise foi se structure, selon cet exemple, par le fait qu'une personne croit une proposition (« ce monsieur a des intentions pures envers moi-même ») et une proposition opposée (« ce monsieur a un désir physique dont la réalisation est son objectif envers moi »), propositions nécessairement incompatibles entre elles.

Jean-Pierre Dupuy, reprenant Amos Tversky et George A. Quattrone, donne un autre exemple<sup>60</sup>, celui qui est contenu dans *L'éthique protestante et l'esprit du capitalisme* de Weber. Selon Weber, les calvinistes, presque par définition, croient en la prédestination : au moment de la naissance, peut-être même avant ou même de

<sup>55</sup> *Petit Larousse*, édition 2003.

<sup>56</sup> Jean-Paul Sartre, *op. cit.*, p. 104.

<sup>57</sup> Jean-Paul Sartre, *ibid.*, p. 84.

<sup>58</sup> Jean-Paul Sartre, *ibid.*, p. 84-85.

<sup>59</sup> Jean-Paul Sartre, *ibid.*, p. 91-92.

<sup>60</sup> Jean-Pierre Dupuy, *op. cit.*, p. 80 et s.

toute éternité, la personne est prédéterminée, c'est-à-dire qu'après sa vie terrestre, elle ira – schématiquement – au paradis ou en enfer, sans qu'il soit possible de faire quoi que ce soit pour modifier cet état de fait : « c'est par la grâce que vous êtes sauvés, par le moyen de la foi. Cela ne vient pas de vous, c'est le don de Dieu. Ce n'est pas par les œuvres, afin que personne ne se glorifie » (Épître aux Ephésiens, II, 8). Cette phrase, fondement du calvinisme et notamment de la prédestination, ne peut, rationnellement, que conduire à une vie terrestre maximisant le plaisir puisque cela n'a aucune répercussion sur la vie dans l'au-delà. Weber constate cependant, chez les calvinistes, une vie plus ascétique que chez d'autres, notamment les catholiques, à l'encontre de la rationalité. Ils font donc un choix, dont l'objectif est d'acquérir le signe de leur élection, c'est-à-dire que la prédestination les envoie vers le salut, mais sans jamais pouvoir en acquérir la certitude. Mais ils ne peuvent pas faire un choix puisqu'ils sont prédestinés. Ils ne peuvent que croire avoir fait un choix, alors qu'en même temps ils croient ou savent ne pas pouvoir en faire. Dupuy résume ainsi les deux propositions :

- (1) Les Calvinistes croient qu'ils ont procédé eux-mêmes à leur élection en choisissant x;
- (2) Les Calvinistes croient qu'ils n'ont pas procédé eux-mêmes à leur élection.

(1) et (2) expriment des croyances contradictoires. On peut de plus supposer que :

Les Calvinistes s'arrangent pour se cacher (1)  
parce qu'ils veulent croire qu'ils ont été élus par Dieu<sup>61</sup>.

Plus loin, Dupuy donne une autre version des croyances contradictoires :

- (3) Les Calvinistes croient qu'ils n'ont pas procédé eux-mêmes à leur élection, parce qu'ils croient que c'est Dieu qui

<sup>61</sup> Jean-Pierre Dupuy, *ibid.*, p. 80.

les a élus; (4) Les Calvinistes croient qu'ils étaient libres de choisir  $x$  ou non- $x$  lorsqu'ils ont choisi  $x$ <sup>62</sup>.

Davidson donne un troisième exemple, celui d'un homme qui se voit chauve et se croit tel et qui croit qu'il n'est pas chauve<sup>63</sup>. Ces trois exemples donnent une indication sur la structure de la mauvaise foi. En effet, il est complètement irrationnel de croire à la fois  $x$  et non- $x$ . Pourtant, on trouve des cas où il y a ces deux croyances chez un même individu au même moment et sur une durée non négligeable (pour le calviniste, au moins toute la durée de la vie terrestre). Il faut donc qu'il y ait quelque chose qui fasse que les deux croyances contradictoires puissent coexister durablement. Selon Davidson, suivi en cela par Dupuy, alors que ce n'est qu'implicite chez Sartre, il faut que l'une des croyances découle de l'autre, non sur un plan strictement logique mais plutôt d'ordres psychologique et affectif. Il semble qu'il faille, non seulement, que non- $X$  découle de  $X$  mais qu'en plus elle soit renforcée par  $X$  pour qu'elle puisse coexister dans la durée. Cela va à l'encontre de l'idée de David Pears selon qui celui qui fait preuve de mauvaise foi doit, d'une façon ou d'une autre, « oublier » un élément de la contradiction ou de la construction de celle-ci<sup>64</sup>, reprenant le fondement de la thèse aristotélicienne de la connaissance. Ainsi, c'est parce que  $D$  est chauve (et qu'il le sait, bien qu'il suffise qu'il le croie) qu'il met en œuvre des stratégies « cosmétiques et surtout psychologiques<sup>65</sup> » pour se persuader et finir par croire qu'il n'est pas chauve; mais s'il ne l'était pas, il n'aurait pas besoin de ces stratégies. Des phrases comme « je ne suis pas si chauve que cela », au-delà de leur ambiguïté sémantique, montrent bien la coexistence contradictoire des deux propositions.

Nous utilisons le terme de mauvaise foi, dans le sens sartrien du terme malgré la confusion possible avec le sens courant, sans approfondir les nuances qui peuvent apparaître dans les

<sup>62</sup> Jean-Pierre Dupuy, *ibid.*, p. 81.

<sup>63</sup> Donald Davidson, *op. cit.*, p. 45-46.

<sup>64</sup> David Pears, *Motivated Rationality*, Oxford, Oxford University Press, 1984.

<sup>65</sup> Jean-Pierre Dupuy, *op. cit.*, p. 80.

différentes situations et surtout selon les différents auteurs traitant ce thème.

La self-déception. Le terme est malaisé à traduire. Comme sur ce point les philosophies anglo-américaines, malgré leur aversion habituelle pour la pensée française, n'hésitent pas à convoquer Sartre, je me rabattraï sur « mauvaise foi » au sens que l'auteur de *L'Être et le Néant* donne à cette expression, c'est-à-dire le mensonge à soi-même<sup>66</sup>.

Pascal Engel, lui, écrit :

On ne peut pas traduire par « mensonge à soi-même » pour des raisons qu'indique Davidson. (Peut-être le menteur qui croit à ses propres mensonges est-il un *self deceive*, mais c'est un cas trop spécifique). « Mauvaise foi » pourrait convenir, mais, outre qu'il n'y a pas de forme verbale transitive correspondante, son sens sartrien, trop particulier pour être assimilé à la *self deception*, est trop prégnante<sup>67</sup>.

La mauvaise foi est un état mental et, en tant que tel, est un « involuable ». En effet, si croire est un « involuable » (je ne peux pas vouloir croire), croire non-X alors que je crois X est « encore plus » un « involuable » puisqu'il s'agit de se forcer à croire (ce qui n'est pas possible) quelque chose qui n'est pas vrai. Par conséquent, une partie du couple composé de croire X et croire non-X étant « involuable », ce couple lui-même l'est. La mauvaise foi n'existe donc qu'en toute bonne foi, parce qu'elle n'existe qu'involontairement. Sinon on retombe dans le simple mensonge. De même, sortir d'une situation caractérisée par la mauvaise foi ne peut se faire volontairement car alors il faut respecter le principe d'exclusion des contraires et le principe du tiers exclu qui consistent à opter pour la croyance en X ou la croyance en non-X. Mais opter pour une l'une des deux croyances la renforce et risque de renforcer l'autre : la mauvaise foi se caractérise donc aussi, sur un plan dynamique, par un cercle vicieux.

Da Silva propose une autre possibilité d'explication de la mauvaise foi qui ne serait qu'un leurre sémantique provenant du fait que le langage ne permet que de croire ou de ne pas croire alors que, selon cet auteur, on doit envisager la possibilité de croire plus ou moins que la structure de la croyance n'est pas

<sup>66</sup> Jean-Pierre Dupuy, *op. cit.*, p. 80.

<sup>67</sup> Pascal Engel, dans Donald Davidson, *op. cit.*, p. 10, note de bas de page.

bivalente ou manichéenne mais strictement continue, ce qu'il ne s'agit pas de confondre ni avec l'incertitude relative ni avec la probabilité<sup>68</sup>.

Davidson conclut à l'irrationalité de celui qui fait preuve de mauvaise foi en donnant les éléments suivants :

- La mauvaise foi survient plus souvent lorsqu'il y a croyance que lorsqu'il y a certitude (il ne donne d'ailleurs pas d'exemple d'existence de tels cas) :

il est probable qu'il arrive rarement qu'une personne soit certaine qu'une proposition donnée est vraie et en même temps certaine que sa négation est vraie.

- La mauvaise foi survient lorsque les propositions X et non-X sont établies en termes de croyances excluant la certitude absolue;
- Alors que l'akrasie porte sur l'action ou au moins l'intention d'agir, la mauvaise foi porte sur des croyances c'est-à-dire des attitudes cognitives;
- La mauvaise foi inclut la faiblesse de la garantie;
- La mauvaise foi dépasse la faiblesse de la garantie dans la mesure où l'agent a une raison d'avoir cette faiblesse : « c'est l'agent lui-même qui induit la faiblesse de la garantie (il l'a produite lui-même) »<sup>69</sup>.

L'une de ces raisons peut relever du fait de prendre ses désirs pour des réalités, mais pas de façon exclusive ni nécessaire. Avoir le désir que non-X soit vrai alors que X est fortement probable tend à faire croire que non-X est plus probable qu'il ne le serait s'il n'y avait pas ce désir et plus probable que X. Or, l'acte de prendre ses désirs pour des réalités provient bien du fait que X est vraie ou plus probable que ce que l'on désire. Par exemple, le désir d'être aimé de quelqu'un peut engendrer la croyance qu'il en est ainsi. On a donc une « bonne raison » de croire et, notamment si la personne en question (celle de laquelle on souhaite être aimé) n'a jamais donné de preuves ou d'indices dans

<sup>68</sup> Victor Vieira da Silva, *op. cit.*

<sup>69</sup> Donald Davidson, *op. cit.*, p. 52.

un sens ou dans l'autre. Cela entraîne une raison d'un autre type qui consiste à avoir des données qui confirment ces croyances : de faux indices sont alors érigés en certitudes ou en indices vrais alors que « tout » indique le contraire, notamment le réquisit de la totalité des données disponibles.

Enfin, et c'est ce qui fait de l'action de se duper soi-même quelque chose de problématique, l'état qui motive la duperie de soi et l'état qu'elle produit coexistent; au sens le plus fort, la croyance que  $p$  non seulement cause une croyance en la négation de  $p$ , mais aussi l'étaye. Se duper soi-même est donc une forme de faiblesse de la garantie induite par le sujet lui-même, où le motif qu'il a d'induire cette croyance est une croyance contradictoire (ou ce qu'il tient comme confirmant la croyance contradictoire). Dans certains cas, mais pas dans tous, le motif prend sa source dans le fait que l'agent souhaite que la proposition, la croyance qu'il provoque en lui, soit vraie, ou dans la peur qu'elle ne le soit pas. C'est pourquoi se duper soi-même implique souvent que l'on prenne aussi ses désirs pour des réalités<sup>70</sup>,

ce qui est un « involuable ». Le sophisme de l'amortissement est une manifestation de mauvaise foi. En effet, le faire reposer sur l'ignorance (non volontaire) du principe de fixité du passé n'est pas satisfaisant, bien qu'il permette d'en faire quelque chose de rationnel dans une perspective de rationalité limitée. Si l'on considère que, au moins intuitivement, l'agent sait qu'il ne peut changer le passé et qu'il ne peut maximiser son utilité que dans le présent et l'avenir, alors le fait d'essayer de réécrire le passé suppose, qu'en même temps, il sache que le passé est fixe et qu'il croit qu'il peut le changer. Il y a bien mauvaise foi, reposant sur une faiblesse de la garantie.

Reste la question de la fréquence de la mauvaise foi. Aucune réponse directe n'est proposée, la question de la mauvaise foi étant, une fois montrée son existence, d'en analyser les conditions d'émergence et d'existence et les ressorts logiques et psychologiques qui la sous-tendent. Des textes de Davidson, on ne peut guère uniquement conclure que la mauvaise foi existe « plus souvent que rarement », mais il ne donne pas plus de précision et prend toutes les précautions au point que l'on doute que la

<sup>70</sup> Donald Davidson, *ibid.*, p. 56.



formulation choisie l'ait été pour sous-entendre un certain niveau de fréquence. De façon plus explicite, Sartre laisse penser que la mauvaise foi est la règle ou l'une des principales règles d'attitude de l'être humain, envers les autres mais surtout envers soi-même : « elle peut même être l'aspect normal de la vie pour un très grand nombre de personne<sup>71</sup> ». Il le laisse aussi entendre par la simplicité, la banalité des exemples qu'il prend : le garçon de café qui joue à être garçon de café, l'épicier qui joue à être épicier, l'élève attentif, le tailleur, le commissaire-priseur, le beau parleur, celui qui est triste, celui qui est sincère : « Qu'est-ce donc que la sincérité, sinon précisément un phénomène de mauvaise foi?<sup>72</sup> ».

### La structure de base de la complexité

Décortiquons un exemple donné par Sartre, celui du garçon de café parisien. Celui-ci est « sûr »<sup>73</sup> d'être garçon de café (c'est la proposition A et il peut en être sûr, il a toutes les raisons d'en être sûr, ne serait-ce que par son contrat de travail). Sartre constate cependant que ce garçon de café joue au garçon de café (il en fait un peu trop : non seulement il joue un rôle mais en plus il sur-joue). Il en déduit que si ce garçon de café joue au garçon de café, c'est qu'il cherche à convaincre qu'il l'est bien. Mais qui cherche-t-il à convaincre sinon lui-même? S'il cherchait à convaincre autrui, il le ferait plutôt sous la forme argumentative et, en tout cas, ne jouerait pas aussi systématiquement au garçon de café car l'enjeu est, vis-à-vis d'autrui, assez minime et circonstanciel. Or, il joue son rôle tout le temps (du moins le temps dédié au travail) ou presque. C'est donc lui-même qu'il cherche à convaincre et s'il cherche cette conviction, c'est parce que, justement, il n'en est pas sûr (c'est la proposition non-A). On a donc bien coexistence de deux propositions contradictoires.

<sup>71</sup> Jean-Paul Sartre, *op. cit.*, p. 85.

<sup>72</sup> Jean-Paul Sartre, *ibid.*, p. 99.

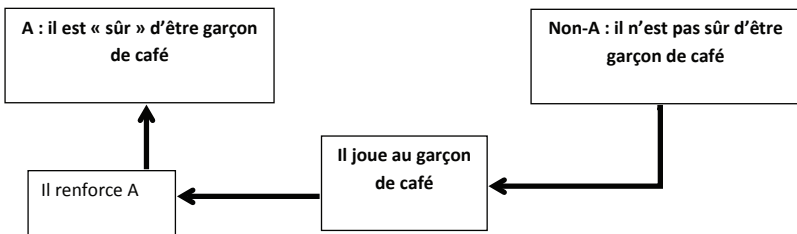
<sup>73</sup> On place « sûr » entre guillemets car la suite de la démonstration montre qu'il n'en est pas si sûr que cela.



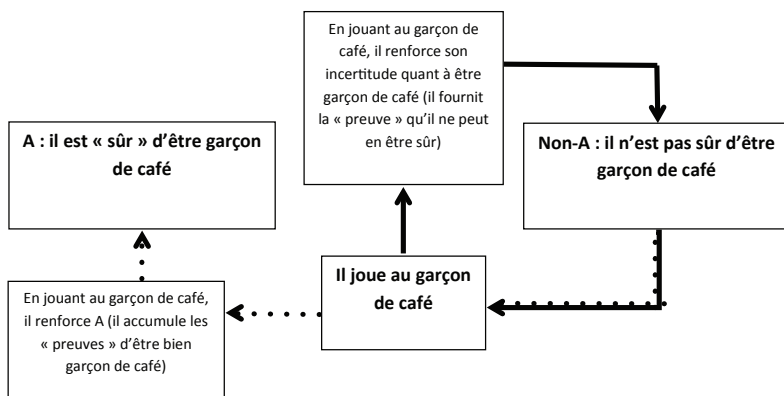
Le fait qu'il joue au garçon de café pour se convaincre qu'il l'est bien découle du fait qu'il n'est pas sûr de l'être (c'est-à-dire de non-A).



Ce faisant, il renforce sa certitude d'être un garçon de café :



Mais, en renforçant A par le rôle qu'il joue, le garçon de café se donne aussi pour preuve que A n'est pas si sûr que cela, que A n'est pas aussi bien établi (sinon, il ne chercherait pas à renforcer sa certitude : s'il a besoin de la renforcer c'est qu'il n'en est pas sûr, c'est-à-dire que cette « certitude » n'en est pas une). On a alors le schéma complet suivant :



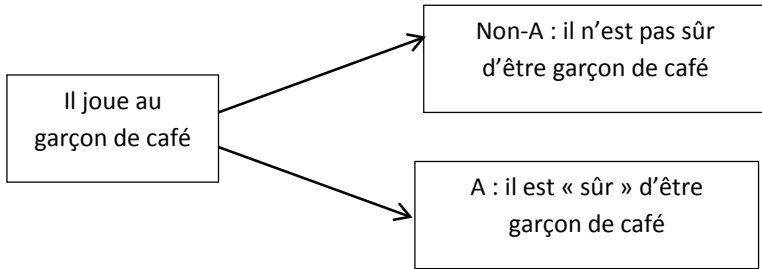
Par le jeu des flèches pointillées, il est de plus en plus sûr de A et, par le jeu des flèches autres, il est de moins en moins sûr de A. Le chemin matérialisé par les flèches pointillées est rationnel puisque, partant d'une incertitude, il se donne des moyens pour la lever et de fait est de plus en plus sûr (de moins en moins incertain). Mais en même temps, la mise en œuvre de moyens donne aussi des indices comme quoi il a raison de ne pas être sûr. La « tension » entre les deux propositions contradictoires grandit mais, en même temps, de façon dynamique, elles se contraignent l'une l'autre, allant jusqu'à donner une certaine stabilité à l'ensemble : la tension est réduite par le jeu de rôle que s'impose le garçon de café et qui s'impose à lui.

Les exemples que donnent Sartre, Davidson ou Dupuy reposent, au-delà des interprétations philosophiques ou des hypothèses d'ordre psychologique, non pas sur l'idée que A entraîne non-A et, de façon simultanée, que non-A entraîne A<sup>74</sup>, mais sur l'idée d'une cause commune à A et non-A.

74

Cette idée que  $[A \rightarrow \text{non-A}]$  et en même temps  $[\text{non-A} \rightarrow A]$  renvoie à l'idée de boucle de rétroaction, centrale dans l'examen de la complexité des systèmes, notamment quand celui-ci est représenté via la théorie des graphes.

Cette cause commune est le fait de jouer et cette cause a des conséquences ainsi corrélées mais surtout contradictoires. On a (en simplifiant le schéma et en gardant une partie de sa structure) :

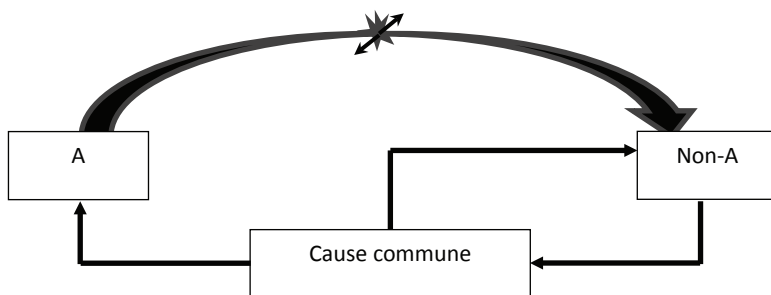


Cependant, il est clair que de façon dynamique on a bien un système qui évolue, éventuellement s'entretient ou se dirige vers une situation de crise, mais rien n'indique pourquoi cette dynamique s'enclenche : pourquoi le garçon de café, à un moment donné, se met à douter de ce qu'il n'a aucune raison de douter d'être?

### Conclusion

À partir de ce qui précède, on peut compléter par le fait que, en ce qui concerne les systèmes complexes, un certain niveau de complexité est atteint lorsqu'il existe un ou plusieurs phénomènes de rétroaction entre les différents éléments du système, ce qui rejait sur le fonctionnement de chacun de ces éléments et le fonctionnement global, et que cette rétroaction est l'une des conditions de la complexité. Le dépassement que nous proposons est le suivant : on atteint un autre niveau de complexité – et ce faisant d'imprédictibilité de l'évolution du système considéré – lorsqu'il y a, outre ce bouclage rétroactif, dans les éléments du circuit qui compose ce bouclage, une cause commune à deux éléments contradictoires et néanmoins coexistants, cette cause commune renforçant chacun des deux (ou plus) éléments et l'un des deux éléments découlant, contre toute logique, de l'élément qui lui est contradictoire. On a :

- une boucle de rétroaction;
- la coexistence de deux éléments contradictoires;
- l'existence d'une cause commune à ces deux éléments contradictoires;
- l'« action » de cette cause commune renforce chacun des deux éléments (elle les accentue);
- l'un des deux éléments découle (indirectement, via la cause commune) de l'autre, mais il n'y a pas réciprocité.



Qu'est-ce que A : une proposition (au sens de la logique), une situation, une croyance? Dans les exemples donnés, Dupuy et Davidson se situent clairement du côté des croyances et l'essentiel des exemples qu'ils décrivent est de cet ordre-là, mais Davidson n'exclut pas la possibilité d'une situation semblable fondée non sur une croyance mais sur une certitude. Pour Sartre, si la mauvaise foi renvoie clairement, pour lui aussi, à des attitudes cognitives ou plus largement des états mentaux, il les présente, néanmoins, par la description d'actions, d'attitudes et de comportements (physiques, gestuels, relationnels et non simplement des attitudes cognitives), ou plus largement de situations matérielles : il induit de l'observation de ces situations matérielles ou comportements l'existence de situation de mauvaise foi.

Un système complexe n'a pas forcément capacité à avoir des attitudes cognitives. Ainsi, si la structure de base de la complexité dans le champ spécifique qu'est le couple attitude cognitive-action chez l'acteur humain, il apparaît nécessaire d'élargir cette

définition, pour couvrir d'autres champs dans lesquels existent des systèmes complexes. On retrouve ce type de situation d'une cause qui conduit à des phénomènes opposés, hors du champ de ce couple. Par exemple, la réponse aux besoins alimentaires de régions touchées par une famine qui consiste à envoyer une aide strictement alimentaire (envoi de denrées) contribue à rendre plus fréquente ce type d'aide : en effet, l'aide en nourriture arrive dans les grandes villes (portuaires généralement) provoquant la venue des populations non spécifiquement urbaines, parmi lesquelles ceux qui cultivent la terre pour les besoins locaux et qui, pour beaucoup, n'ont pas les moyens de retourner ensuite sur les terres qu'ils cultivaient, pérennisant la situation d'une production agricole locale insuffisante, requérant ainsi l'aide internationale. Cet exemple semble montrer qu'en passant des attitudes mentales (où ce sont finalement les « raisons » même implicites, même inconscientes qui font agir d'une certaine façon) à des actions qui ont des conséquences (donc dans un rapport cause-conséquence) s'intercale le temps : dans les exemples donnés par Sartre les « phases » sont concomitantes (« dans la structure unitaire d'un même projet ») tandis que, dans l'exemple de l'aide alimentaire, il y a d'une part l'urgence de la réponse et, d'autre part, les conséquences à long terme de cette réponse.

## Bibliographie

- Allais, Maurice, « Fondements d'une théorie positive des choix comportant un risque et critique des postulats et axiomes de l'école américaine », *Annales des Mines*, vol. 144, n° spécial, 1955.
- Austin, John Langshaw, *Quand dire, c'est faire*, Paris, Seuil, 1970.
- Bateson, Gregory, *Vers une écologie de l'esprit*, Paris, Seuil, 2008 [1972].
- Becker, Gary, *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education*, Chicago, University of Chicago Press, 1964.
- Becker, Gary, *A Treatise on the Family*, Cambridge (MA), Harvard University Press, 1991 [1981].
- Bentham, Jeremy, *An Introduction to the Principals of Morals and Legislation*, Oxford, Oxford University Press, 1996.
- Boudon, Raymond, *Effets pervers et ordre social*, Paris, Presses universitaires de France, 1993.
- Boudon, Raymond, *Raison. Bonnes raisons*, Paris, Presses universitaires de France, coll. « Philosophe en sciences humaines », 2007.
- Bourdieu, Pierre et Jean-Claude Passeron, *La reproduction, éléments pour une théorie du système d'enseignement*, Paris, Minuit, 1970.
- Bronner, Gérald, *L'incertitude*, Paris, Presses universitaires de France, coll. « Que sais-je? », 1997.
- Brousseau, Éric, « Contrats et comportements coopératifs : le cas des relations interentreprises », dans Jacques-Laurent Ravix (dir.), *Coopération entre les entreprises et organisation industrielle*, Paris, Éditions du CNRS, 1996, p. 23-51.
- Davidson, Donald, *Paradoxes de l'irrationalité*, Combas, Éditions de l'Éclat, 1991.
- Demailly, André, *Herbert Simon et les sciences de conception*, Paris, L'Harmattan, coll. « Ingénium », 2004.
- Dupuy, Jean-Pierre « Temps et rationalité », dans René Frydman (dir.), *Quelles hypothèses de rationalité pour la théorie économique?*, Cahiers d'économie politique, 1994, n° , p. 69-104.
- Elster, Jon, *Le laboureur et ses enfants, Deux essais sur les limites de la rationalité*, Paris, Minuit, 1986.
- Forrester, Jay Wright, *Dynamique urbaine*, Paris, Economica, 1979.
- Gouillou, Philippe, *Pourquoi les femmes des riches sont belles : programmation génétique et compétition sexuelle*, Bruxelles, De Boeck, 2014.

- Kahneman, Daniel, *Système 1 / Système 2 : les deux vitesses de la pensée*, Paris, Flammarion, 2012.
- Levitt, Steven D. et Dubner Stephen J., *Freakonomics. A Rogue Economist Explores the Hidden Side of Everything*, New York, William Morrow, 2005.
- Lévy-Garboua, Louis, « L'économie et le rationnel », *L'année sociologique*, n° 31, 1981, p. 19-47.
- Looms, Graham, « Further Evidence on the Impact of Regret and Disappointment in Choice under Uncertainty », *Economica*, London School of Economics and Political Science, vol. 2, n° 1, 1988, p. 47-62.
- Lupasco Stéphane, *Le principe d'antagonisme et la logique de l'énergie. Prolégomènes à une science de la contradiction*, Paris, Hermann, 1951.
- Martouzet, Denis, « L'urbaniste est-il en mesure de créer les conditions spatiales de l'urbanité », dans Pierre W. Boudreault et Denis Jeffrey (dir.), *Identités en errance. Multi-identité, territoire impermanent et être social*, Québec, Presses de l'Université Laval, 2007, 93-113.
- Morin Edgar, *Introduction à la pensée complexe*, Paris, Seuil, 2005.
- Newell, Allen et Herbert A. Simon, *Human Problem Solving*, Englewood Cliffs (NJ), Prentice-Hall, 1972.
- Nozick, Robert, *The Nature of Rationality*, Princeton (NJ), Princeton University Press, 1993.
- Pareto, Vilfredo, *Traité de sociologie générale*, Genève, Droz, 1968 [1917].
- Pears, David, *Motivated Rationality*, Oxford, Oxford University Press, 1984.
- Piaget, Jean et Bärbel Inhelder, *La genèse de l'idée de hasard chez l'enfant*, Paris, Presses universitaires de France, 1974.
- Robinson, Richard, « L'acrasie selon Aristote », *Revue philosophique*, n° 80, 1955, p. 261-280.
- Rousseau, Jean-Jacques, *Les confessions*, tomes 1 et 2, Paris, Gallimard et Librairie Générale Française, 1963.
- Sartre, Jean-Paul, *L'être et le néant*, Paris, Gallimard, coll. « Bibliothèque des idées », 1943.
- Simon, Herbert A., « A Behavioral Model of Rational Choice », *Quarterly Journal of Economics*, n° 69, 1955, p. 99-118.
- Simon, Herbert A., « From Substantive to Procedural Rationality », dans Spiro J. Latsis (dir.), *Method and Appraisal in Economics*, Cambridge (MA), Cambridge University Press, 1976, p.129-148.
- Simon, Herbert A., *Models of thought*, vol. 1, New Haven (CT), Yale University Press, 1979.



- Simon, Herbert A., « On how to Decide What to Do », *The Bell Journal of Economics*, vol. 9, n° 2, 1978, p. 494-507.
- Simon, Herbert A., *Reason in Human Affairs*, Stanford (CA), Stanford University Press, 1983.
- Simon, Herbert A., « Theories of Decision-Making in Economics and Behavioral Science », *American Economic Review*, vol. 49, n° 1, 1959, p. 253-283.
- Stigler, George J., « The Economics of Information », *Journal of Political Economy*, vol. 69, n° 3, 1961, p. 213-225.
- Tversky, Amos, « Rational Theory and Constructive Choice », dans Kenneth J. Arrow, Enrico Colombatto, Mark Perlman et Christian Schmidt (dir.), *The Rational Foundations of Economics Behaviour*, New York, Macmillan and St-Martin Press, 1996, p. 185-197.
- Tversky, Amos, Daniel Kahneman et Paul Slovic, *Judgment under Uncertainty: Heuristics and Biases*, Cambridge (MA), Cambridge University Press, 1982.
- Tversky, Amos et Daniel Kahneman, « Rational Choice and the Framing of Decisions », dans David E. Bell, Howard Raiffa et Amos Tversky (dir.), *Decision Making, Descriptive, Normative and Prescriptive interactions*, Cambridge (MA), Cambridge University Press, 1988, p. 167-192.
- Vieira da Silva, Victor, *L'hypothèse de rationalité et ses prémisses : le défi comportementaliste, une investigation aux frontières de l'économie et de la psychologie*, thèse de doctorat en sciences économiques, Lyon, Université de Lyon, 2001.
- Watzlawick, Paul, John Weakland et Robert Fisch, *Changements, paradoxes et psychothérapie*, Paris, Seuil, 1975.
- Weber, Max, *Economie et Société*, 2 tomes, Paris, Pocket, coll « Agora », 1995.