

L'homme est-il un animal rationnel ?

Stephen Stich

Département de philosophie et Centre de sciences cognitives
Rutgers University

I. Introduction : les approches descriptives et normatives de l'étude du raisonnement humain

Dans l'étude du raisonnement humain, on considère généralement que nous pouvons adopter deux approches très différentes. L'une est descriptive ou empirique. Ceux qui l'adoptent cherchent à caractériser la façon dont les gens s'y prennent pour raisonner ainsi qu'à découvrir les mécanismes et processus psychologiques qui sous-tendent les formes de raisonnement qu'ils observent. Cette approche descriptive est le plus souvent pratiquée par des psychologues, quoique des anthropologues aient aussi fait des travaux très intéressants pour savoir si et dans quelle mesure les gens appartenant à des cultures différentes raisonnent différemment. L'autre approche de l'étude du raisonnement est normative. Ceux qui l'adoptent ne s'intéressent pas à la façon dont les gens raisonnent vraiment, mais plutôt à la façon dont ils devraient raisonner. Leur but est de découvrir des règles ou principes spécifiant ce qu'est raisonner correctement ou rationnellement. Depuis l'Antiquité, l'approche normative du raisonnement a été pratiquée par des logiciens et des philosophes et, plus récemment, des statisticiens ainsi que des théoriciens des probabilités et de la décision ont joint leurs rangs.

Les buts et les méthodes de ces deux types d'étude du raisonnement sont très différents. Quoi qu'il en soit, certaines des thèses les plus intéressantes et les plus chaudement débattues à propos du raisonnement humain ont trait à la question de savoir dans quelle mesure tel ou tel groupe de gens raisonne rationnellement. Ces thèses sont intrinsèquement hybrides ou interdisciplinaires. Pour les évaluer, nous devons savoir à la fois ce qu'est raisonner rationnellement et comment le groupe de personnes en question raisonne vraiment. La thèse la plus célèbre de ce type est celle d'Aristote, selon laquelle l'homme est un animal rationnel. La rationalité, selon Aristote, est une propriété essentielle de l'espèce humaine; c'est ce qui distingue l'homme de la bête. Un des buts centraux de cet essai est de réexaminer la thèse d'Aristote à la lumière des études empiriques récentes sur le raisonnement humain. Aristote avait-il raison ? Les humains sont-ils vraiment des animaux rationnels ?

Pour répondre sérieusement à cette question, nous devons d'abord fournir une interprétation plus précise de la thèse d'Aristote. Il est évident que les humains ne sont pas toujours rationnels. Lorsque les gens sont ivres, épuisés ou sous le coup d'une rage incontrôlable, ils raisonnent en fait médiocrement. Et, bien sûr, Aristote savait cela. Lorsqu'il disait que les humains sont des animaux rationnels, il n'a sûrement jamais voulu nier que les gens peuvent raisonner et raisonnent parfois de façon irrationnelle. Mais que voulait-il dire alors ? À la section II, nous verrons comment nous pouvons commencer à fournir une interprétation assez précise et en fait assez troublante de la thèse d'Aristote en empruntant une idée des sciences cognitives contemporaines - l'idée que les gens ont des «compétences» sous-jacentes dans des domaines variés, même si ces compétences ne se reflètent pas toujours dans leur comportement. Une interprétation séduisante de la thèse d'Aristote est que les humains normaux ont une compétence rationnelle dans le domaine du raisonnement.

Pour expliquer la notion de compétence rationnelle, il nous faudra cependant faire plus qu'expliquer la notion de compétence ; nous devons aussi dire ce qu'est une compétence rationnelle. À la section III, nous examinerons une tentative élégante et influente d'expliquer ce qu'est une forme de raisonnement rationnelle - une approche qui fait appel à la notion d'équilibre réfléchi.

Ayant interprété la thèse d'Aristote comme l'affirmation que les gens normaux possèdent une compétence de raisonnement rationnelle, notre attention se portera, à la section V, sur les études descriptives du raisonnement. Dans cette section, nous examinerons quelques découvertes empiriques fascinantes et intrigantes qui semblent suggérer qu'Aristote avait tort parce que la plupart des gens ne possèdent pas la compétence de raisonner rationnellement sur divers types de questions ! Ces découvertes ne remettent pas uniquement en question la thèse optimiste d'Aristote concernant la rationalité humaine, elles semblent également menacer l'explication de la rationalité en termes d'équilibre réfléchi. La section V commencera par expliquer pourquoi les découvertes empiriques posent un problème à l'explication en termes d'équilibre réfléchi et elle se poursuivra en examinant quelques réponses possibles.

Dans la section VI, nous reviendrons à la littérature empirique à propos du raisonnement humain, cette fois en portant notre attention sur quelques études très récentes des psychologues évolutionnistes qui prennent comme point de départ l'idée que les composantes ou «organes" de l'esprit ont été façonnés par la sélection naturelle tout comme les composantes ou organes des (autres) parties du corps l'ont été. Cette perspective les conduit à s'attendre à ce que nos esprits parviennent à des résultats assez satisfaisants lorsqu'ils raisonnent à propos de problèmes du type de ceux qui auraient été importants dans l'environnement dans lequel notre espèce a évolué, et ils ont produit quelques raisons très intéressantes de penser que c'est en fait bien le cas.

Finalement, à la section VII, nous nous demanderons ce que ces études récentes nous disent à propos de la thèse d'Aristote. Est-ce qu'elles montrent, comme certains le prétendent, qu'Aristote avait après tout raison ? Je soutiendrai alors que la réponse est que ni la position d'Aristote ni celle ses adversaires ne sont justifiées par la recherche empirique. Ce que ces études montrent plutôt est que les questions que nous nous sommes posées à propos de la rationalité humaine - des questions comme «Est-ce que l'homme est un animal rationnel ?" - sont trop simplistes. Si nous voulons des réponses plausibles construite sur une base scientifique solide, nous devrions apprendre à poser des questions meilleures et plus élaborées.

II. Compétence et performance

Tous s'entendent pour dire que les gens sont parfois très irrationnels. Ainsi, par exemple, il arrive que des gens qui sont ivres, hors d'eux-mêmes ou sous l'influence de drogues raisonnent et prennent des décisions en empruntant des voies qui ne sont sanctionnées par aucune théorie sérieuse à propos des principes gouvernant le bon raisonnement et la prise de décision. Puisque c'est un fait si évident et aucunement controversé, comment quiconque peut-il néanmoins prétendre que les humains sont des animaux rationnels ? Qu'est ce que cela peut bien vouloir dire ? Une réponse très séduisante peut être donnée en exploitant la distinction entre compétence et performance.

La distinction compétence/performance a d'abord été introduite en sciences cognitives par Chomsky, qui l'utilisa dans sa présentation des stratégies explicatives des théories linguistiques. En testant les théories linguistiques, une source de données importante provient des «intuitions" ou jugements non réfléchis que les locuteurs d'une langue ont (ou font) à propos de la grammaticalité des phrases et de diverses propriétés linguistiques (par exemple, «Est-ce que la phrase est ambiguë? ", «Est-ce que ce syntagme est le sujet de ce verbe ?"). Pour expliquer ces intuitions, et aussi la façon dont les

locuteurs s'y prennent pour produire et comprendre des phrases de leur langue, Chomsky proposa ce qui est devenu une des hypothèses sur l'esprit qui compte parmi les plus importantes de l'histoire des sciences cognitives. Ce que cette hypothèse stipule est qu'un locuteur d'une langue possède une «grammaire générative" internalisée de cette langue. Une «grammaire générative" est un ensemble intégré de règles et de principes - nous pouvons nous la représenter comme analogue à un système d'axiomes - qui implique un nombre infini d'énoncés à propos de cette langue. Pour chacune des phrases de la langue du locuteur, la grammaire internalisée du locuteur indique qu'elle est grammaticale ; pour chaque phrase ambiguë, la grammaire internalisée indique qu'elle est ambiguë, etc. Lorsque les locuteurs portent un jugement que nous nommons une «intuition linguistique", ils ont accès et se fient à l'information de cette grammaire internalisée, quoique ni le processus ni la grammaire internalisée ne soient accessibles à l'introspection consciente. Puisque la grammaire internalisée joue un rôle central dans la production des intuitions linguistiques, ces intuitions peuvent servir comme une importante source de données pour les linguistes qui veulent spécifier quels sont les règles et principes de la grammaire internalisée.

Les intuitions d'un locuteur ne sont cependant pas une source d'informations infaillible à propos de la grammaire de sa langue, parce que les règles internalisées de la grammaire ne peuvent à elles seules produire les intuitions linguistiques. La production des intuitions est un processus complexe dans lequel la grammaire internalisée doit interagir avec divers autres mécanismes cognitifs, incluant ceux qui sont responsables de la perception, de la motivation, de l'attention, de la mémoire à court terme et peut-être une foule d'autres. Dans certaines circonstances, l'activité de n'importe lequel de ces mécanismes peut donner lieu à un jugement sur une phrase qui ne s'accorde pas avec ce que la grammaire indique à son propos. Le mécanisme d'attention offre un exemple clair de ce phénomène. Il est très probable que la grammaire internalisée des locuteurs typiques du français indique qu'un nombre indéfini de phrases de la forme :

A dit que B pense que C croit que D soupçonne que E pense ... que p.

sont grammaticales dans la langue du locuteur. Cependant, si on vous demandait de juger la grammaticalité d'une phrase contenant quelques centaines d'enchâssements de ce type ou même quelques douzaines, il y a de bonnes chances que vos jugements ne reflèteraient pas ce que votre grammaire indique, puisque dans des cas comme ceux-là, l'attention s'égarerait facilement. La mémoire à court terme fournit un exemple plus intéressant de la façon dont un jugement grammatical peut échouer à représenter ce que contient vraiment la grammaire. Il existe de solides raisons de penser que les mécanismes de la mémoire à court terme ont du mal à manipuler des structures enchâssées par le centre. Ainsi, il peut bien se trouver qu'une grammaire internalisée indique que la phrase anglaise suivante est grammaticale:

What What What he wanted cost would buy in Germany was amazing.

en dépit du fait que les intuitions de la plupart des gens suggèrent, crient en fait, qu'elle ne l'est pas. Dans la terminologie que Chomsky a introduite, les règles et les principes de la grammaire générative internalisée d'un locuteur constituent la compétence linguistique du locuteur; les jugements qu'un locuteur fait à propos des phrases, de même que les phrases produites par ce locuteur, font partie de sa performance linguistique. De plus, comme nous venons tout juste de le voir, quelques-unes des phrases produites par le locuteur et quelques-uns des jugements que celui-ci fait à propos de ces phrases ne reflètent pas fidèlement sa compétence linguistique. Dans ces cas, le locuteur commet des erreurs de performance.

Il y a des analogies évidentes entre les phénomènes étudiés par la linguistique et ceux qui sont étudiés par les cognitivistes qui s'intéressent au raisonnement. Dans les deux cas, les gens sont capables de traiter spontanément et inconsciemment une classe ouverte d'«inputs» - ils sont capables de comprendre un nombre indéterminé de phrases et de tirer des inférences à partir d'un nombre indéfini de

prémises. À la lumière de cette analogie, on peut explorer l'idée que le mécanisme sous-jacent à notre habileté à raisonner est similaire au mécanisme sous-jacent à notre capacité à traiter la langue. Si Chomsky a raison à propos de la langue, alors l'hypothèse analogue à propos du raisonnement suggérerait que les gens ont un ensemble intégré de règles et de principes internalisés - une «psycho-logique" comme on l'a appelée - auquel ils ont accès et auquel il se fient habituellement lorsqu'ils tirent des inférences ou portent des jugements à leur propos. Comme dans le cas de la langue, nous devrions nous attendre à ce que ni les processus impliqués ni les principes de la psycho-logique internalisés ne soient accessibles immédiatement à la conscience. Nous devrions également nous attendre à ce que les inférences et les jugements des gens ne soient pas des guides infaillibles au sujet de ce que la psycho-logique sous-jacente indique à propos de la validité ou de la plausibilité d'une inférence donnée, parce qu'ici, comme dans le cas de la langue, les règles et principes internalisés doivent interagir avec beaucoup d'autres mécanismes. L'activité de ces mécanismes peut donner lieu à des erreurs de performance - des inférences ou des jugements qui ne reflètent pas la psycho-logique qui constitue la compétence à raisonner d'une personne.

Nous sommes finalement en position d'expliquer une interprétation de la thèse d'Aristote qui est compatible avec le fait indubitable que, parfois du moins, les gens raisonnent de façon très irrationnelle. Ce que cette thèse stipule, c'est que les gens normaux ont une compétence rationnelle à raisonner. Les règles ou les principes de raisonnement qui constituent leur psycho-logique sont rationnels ou normativement appropriés ; ils spécifient la façon de raisonner correctement. Selon cette interprétation, lorsque les gens font des erreurs dans le raisonnement ou lorsqu'ils raisonnent irrationnellement, leurs erreurs sont des erreurs de performance qui peuvent être dues à la fatigue, au manque d'attention, à la confusion ou à une foule d'autres facteurs. Mais aussi communes qu'elles puissent être, les erreurs de performance ne reflètent pas les règles de raisonnement qui constituent la compétence de raisonnement d'une personne normale. Dire que l'homme est un animal rationnel, selon cette interprétation, c'est dire que la compétence de raisonnement des gens normaux est rationnelle même si leur performance en cette matière ne l'est parfois pas.

III. Qu'est-ce que la rationalité ? Une explication en termes d'équilibre réfléchi

Qu'est-ce qui justifie un ensemble de règles ou de principes pour le raisonnement ? Qu'est-ce qui rend les règles de raisonnement rationnelles ? Il y a environ quarante ans, dans un des passages les plus influents de la philosophie analytique du vingtième siècle, Nelson Goodman suggérait d'élégantes réponses à ces questions. Dans ce passage, Goodman décrivait le processus d'harmonisation entre les jugements à propos des inférences particulières et les principes généraux de raisonnement. Dans l'harmonie ainsi établie, soutient Goodman, se trouve toute la justification désirée, et toute la justification possible, pour les principes inférentiels qui sont apparus. D'autres philosophes, notamment John Rawls, ont adopté une version modifiée du processus de Goodman comme procédure pour savoir si des principes moraux sont corrects. Nous devons à Rawls le terme d'équilibre réfléchi, qui a été largement utilisé pour caractériser un système de principes et de jugements qui ont été rendus cohérents les uns avec les autres de la façon décrite par Goodman.

Il est difficile d'imaginer la notion d'équilibre réfléchi expliquée de façon plus éloquente que dans les termes de Goodman. Aussi, laissez-moi citer un peu longuement ce qu'il en dit :

Comment justifier une *déduction*? Tout simplement, en illustrant sa conformité aux règles générales des inférences déductives. À partir du moment où un argument se conforme à ces règles, il est justifié et valide, en dépit de la fausseté éventuelle de sa conclusion. S'il viole une de ces règles, il est faux, même si sa conclusion est vraie. Aucune connaissance des faits n'est donc nécessaire à la justification d'un argument déductif. En outre, lorsqu'il appert qu'un argument est conforme aux règles de

l'inférence déductive, on le considère habituellement comme justifié et valide, sans se demander ce qui justifie ces règles. De la même manière, le but premier de la justification de l'inférence inductive consiste à montrer qu'elle se conforme aux règles générales de l'*induction*. Cela reconnu, une grande étape est franchie dans la clarification du problème.

Il faut tout de même justifier les règles. La validité d'une déduction dépend en effet non pas de sa conformité avec n'importe quelles règles arbitraires, mais avec un ensemble de règles valides. Quand on parle *des* règles d'inférence, on parle des règles valides, ou, mieux encore, de *certaines* règles valides, puisqu'il peut exister plusieurs ensembles de règles aussi valides les unes que les autres. Comment alors déterminer la validité des règles? Ici encore, certains philosophes vous diront que ces règles proviennent d'un axiome évident, alors que d'autres tenteront de démontrer leur enracinement dans la nature même de l'esprit humain. Je ne crois pas qu'on ait à chercher la réponse si loin. Les principes de l'inférence déductive sont justifiés par leur conformité à la pratique déductive courante. Leur validité dépend de leur compatibilité avec les inférences deductives particulières que nous faisons effectivement et que nous ratifions. Si une règle engendre des inférences inacceptables, nous la considérons comme invalide et nous l'abandonnons. La justification de règles générales se dégage donc des jugements d'acceptabilité des inférences deductives particulières.

Nous semblons tourner en rond de façon flagrante. Les inférences deductives sont justifiées par leur conformité aux règles générales valides, qui sont elles-mêmes justifiées par leur conformité aux inférences valides. Il ne s'agit pas cette fois-ci d'un cercle vicieux: les règles, comme les inférences particulières, sont toutes deux justifiées par leur conformité réciproque. *On modifie une règle si elle engendre une inférence que nous ne sommes pas prêts à accepter; on rejette une inférence si elle viole une règle que nous ne sommes pas prêts à modifier.* La justification est un délicat processus d'ajustement mutuel entre les règles et les inférences acceptées, et cet accord constitue la seule justification dont chacun ait besoin.

Tout cela s'applique aussi à l'induction. Une inférence inductive est justifiée par sa conformité aux règles générales de l'induction, qui sont elles-mêmes justifiées par leur conformité à des inférences inductives reconnues ». (1984, 79-80)

Selon l'explication de Goodman, du moins selon la lecture que j'en propose, passer le test de l'équilibre réfléchi est (comme les philosophes le disent parfois) constitutif de la justification ou de la validité des règles d'inférence. Etre rationnel pour un système de règles inférentielles et d'inférences qui s'accordent avec ces dernières consiste simplement à être en équilibre réfléchi. Mais quel est le statut de cette assertion ? Pourquoi passer le test de l'équilibre réfléchi est-il constitutif de la justification ou de la rationalité des règles inférentielles ? La réponse, je crois, est que Goodman la prend pour une vérité conceptuelle - elle découle de la signification de termes comme «justifié» ou «rationnel» ou de l'analyse du concept de rationalité. On peut soutenir que ce concept est un peu (plus qu'un peu) vague, et que l'analyse en termes d'équilibre réfléchi ne fait aucun effort pour saisir ce caractère vague. Il cherche plutôt à mettre de l'ordre dans le flou et de préciser (spécifier) le concept. Il est ainsi peut-être préférable de voir chez Goodman l'idée que l'explication en termes d'équilibre réfléchi saisit quelque chose comme notre concept ordinaire de rationalité, et que c'est la meilleure façon de rendre ce concept précis.

Toutes les pièces sont maintenant en place pour interpréter la thèse d'Aristote. L'homme est un animal rationnel, selon l'interprétation proposée, signifie que les humains normaux ont une compétence pour le raisonnement - un ensemble mentalement représenté de règles ou de principes pour le

raisonnement - et que ces règles sont rationnelles - elles passeraient le test de l'équilibre réfléchi. Interrogeons-nous maintenant sur la plausibilité de cette thèse. Pour ce faire, il nous faut tourner notre attention vers les études empiriques du raisonnement humain.

IV. Quelques faits troublants à propos de la façon dont les humains raisonnent

Nous commencerons notre exploration de l'aspect psychologique du raisonnement humain en portant notre attention sur des études dont certains auteurs pensent qu'elles présentent un sombre tableau de la rationalité des gens ordinaires. Toutes ces études concernent des sujets normaux (souvent des étudiants universitaires) qui ne sont ni épuisés ni stressés émotionnellement. Néanmoins, plusieurs d'entre eux parviennent à de piètres résultats dans les tâches de raisonnement qu'on leur demande de résoudre.

La tâche de sélection

En 1966, Peter Wason fit état des premières expériences concernant un ensemble de problèmes de raisonnement que l'on en est venu à nommer «la tâche de sélection». Un recueil de textes récent sur le raisonnement décrit ces tâches comme «le problème sur lequel on a fait les recherches les plus intenses de toute l'histoire de la psychologie du raisonnement» (Evens, Newstead & Byrne, 1993, p. 99). Un exemple typique du problème de la tâche de sélection est le suivant:

Voici quatre cartes. Chacune porte une lettre sur un côté et un nombre sur l'autre. Deux de ces cartes apparaissent du côté lettre et deux du côté nombre

E C 5 4

Indiquer quelle carte il vous faut tourner pour vérifier l'affirmation suivante :

Si une carte porte une voyelle d'un côté, alors elle porte un nombre impair de l'autre côté.

Ce que Wason et de nombreux autres chercheurs ont découvert, c'est que les gens ordinaires parviennent en général à des résultats médiocres pour des questions comme celles-là. La plupart des sujets répondent, correctement, que la carte E doit être retournée, mais plusieurs jugent que la carte 5 doit être retournée aussi, en dépit du fait que la carte 5 ne peut falsifier l'énoncé, quoi qu'il y ait de l'autre côté. Une grande majorité des sujets jugent également qu'il ne faut pas retourner la carte 4, même s'il n'y a pas moyen de savoir s'il y a une voyelle de l'autre côté sans la retourner. Les sujets ne sont cependant pas toujours aussi médiocres dans les problèmes de la tâche de sélection. Un grand nombre de variations dans le motif de base ont été essayées, et dans quelques versions du problème un grand pourcentage de sujets répond correctement. Ces résultats forment un motif déroutant, puisqu'il n'y a pas de caractéristiques ou de groupes de caractéristiques évidentes qui séparent les versions où les sujets parviennent à des résultats satisfaisants de celles où les résultats sont médiocres. Comme nous le verrons à la section VI, quelques psychologues évolutionnistes ont soutenu que ces résultats peuvent être expliqués si nous nous concentrons sur les sortes de mécanismes mentaux qui auraient été cruciaux pour le raisonnement à propos des échanges sociaux (ou de «l'altruisme réciproque») dans l'environnement de nos ancêtres hominidés. Les versions de la tâche de sélection pour lesquelles nous sommes bons, selon ce que soutiennent ces théoriciens, sont seulement celles pour lesquelles nos mécanismes auraient été conçus.

L'erreur de la conjonction

En novembre 1980, Ronald Reagan est élu président des États-Unis. Au cours du mois suivant, Amos Tversky et Daniel Kahneman ont administré un questionnaire à 93 sujets qui n'avaient jamais eu de formation en statistiques. Les instructions du questionnaire étaient les suivantes :

Dans ce questionnaire, on vous demande d'évaluer la probabilité d'événements variés qui pourraient se produire pendant 1981. Chaque problème comprend quatre événements possibles. Votre tâche est d'ordonner ces événements selon leur degré de probabilité, en inscrivant 1 pour l'événement le plus probable, 2 pour le second, 3 pour le troisième et 4 pour l'événement le moins probable.

Voici une des questions présentées aux sujets :

Ordonnez les événements suivant selon la probabilité qu'ils se produisent en 1981 :

- (a) Reagan coupera l'aide fédérale aux gouvernements locaux.
- (b) Reagan fournira une aide fédérale aux mères célibataires.
- (c) Reagan va augmenter le budget de la défense de moins de 5%.
- (d) Reagan fournira une aide fédérale aux mères célibataires et coupera l'aide fédérale aux gouvernements locaux.

Le résultat déconcertant fut que 68% des sujets estimèrent que (d) était plus probable que (b), en dépit du fait que (d) ne peut pas se produire sans que (b) ne se produise (Tversky et Kahneman 1982). Dans une autre expérience, qui est depuis lors devenue célèbre, Tversky et Kahneman (1982) présentèrent aux sujets la tâche suivante :

Linda a 31 ans, elle est célibataire, franche et très brillante. Elle a une majeure en philosophie. Quand elle était étudiante, elle était très préoccupée par les questions de discrimination et de justice sociale, et elle a aussi participé à des manifestations antinucléaires.

Ordonnez les énoncés suivants selon leur probabilité, en inscrivant 1 pour le plus probable et 8 pour le moins probable.

- (a) Linda enseigne dans une école primaire.
- (b) Linda travaille dans une librairie et prend des leçons de yoga.
- (c) Linda est active dans le mouvement féministe.
- (d) Linda est une travailleuse sociale en milieu psychiatrique.
- (e) Linda est membre de la ligue des électrices.
- (f) Linda est guichetière dans une banque.
- (g) Linda est vendeuse d'assurances.
- (h) Linda est guichetière dans une banque et active dans le mouvement féministe.

Dans un groupe de sujets naïfs sans formation en probabilités et statistiques, 89% ont jugé que l'énoncé (h) était plus probable que l'énoncé (f). Lorsque la même question fut présentée à des sujets habitués aux statistiques - des étudiants des cycles supérieurs du programme de science de la décision de l'École de commerce de Stanford- 85% ont porté le même jugement ! Des résultats de cette nature, dans lesquels les sujets jugent qu'un événement ou un état de choses composé est plus probable qu'un de ses composants, se sont produits de façon répétée depuis les études fondatrices de Kahneman et Tversky.

L'oubli de la fréquence de base

Selon une explication bayésienne familière, la probabilité d'une hypothèse à propos d'une évidence

particulière dépend, en partie, de la probabilité antérieure de l'hypothèse¹. Cependant, dans une série d'expériences, Kahneman et Tversky (1973) ont montré que les sujets sous-évaluaient souvent sérieusement l'importance des probabilités antérieures. Une de ces expériences présente à la moitié des sujets l'histoire suivante :

Un comité de psychologues a interviewé et administré des tests de personnalité à 30 ingénieurs et 70 avocats ayant tous du succès dans leur domaine respectif. Sur la base de cette information, des vignettes décrivant les 30 ingénieurs et les 70 avocats ont été écrites. Vous trouverez sur vos formulaires cinq descriptions, choisies au hasard parmi les 100 descriptions disponibles. Pour chaque description, veuillez indiquer la probabilité, sur une échelle allant de 0 à 100, que la personne décrite soit un ingénieur.

On présentait à l'autre moitié des sujets le même texte, à l'exception du fait que les fréquences de base étaient inversées. On leur disait que les tests de personnalité avaient été administrés à 70 ingénieurs et 30 avocats. Quelques-unes des descriptions qu'on leur offrait étaient conçues pour être compatibles avec le stéréotype de l'ingénieur que les sujets pouvaient avoir, mais pas avec leur stéréotype de l'avocat. D'autres étaient conçues pour concorder avec leur stéréotype de l'avocat, mais pas avec celui de l'ingénieur. Enfin, une autre série de descriptions était neutre, ne donnant aux sujets aucune information pouvant être utilisée pour prendre leur décision. Voici deux exemples, le premier conçu pour dénoter plutôt un ingénieur, le second conçu en termes neutres :

Jack est un homme de 45 ans. Il est marié et a quatre enfants. Il est conservateur, consciencieux et ambitieux. Il ne montre aucun intérêt pour les problèmes politiques ou sociaux et consacre la majeure partie de ses temps libres à plusieurs passes-temps dont la menuiserie, la voile et les casses-têtes mathématiques.

Dick est un homme de 30 ans. Il est marié, sans enfant. C'est un homme de grande habileté et hautement motivé, il est promis au succès dans son domaine. Il est très aimé de ses collègues.

Comme on s'y attendait, les sujets des deux groupes ont pensé que la probabilité que Jack soit un ingénieur était assez élevée. De plus, dans ce qui semble une violation claire des principes bayesiens, la différence dans les histoires entre les deux groupes de sujets n'a eu presque aucun effet. L'oubli de l'information concernant la fréquence de base était encore plus frappant dans le cas de Dick. Cette description était construite pour être totalement dénuée d'informations eu égard à la profession de Dick. Ainsi, la seule véritable information utile que les sujets avaient était celle touchant la fréquence de base contenue dans l'histoire. Mais cette information était totalement négligée. L'estimé de probabilité médian dans les deux groupes de sujets était de 50%. Les sujets de Kahneman et Tversky n'étaient cependant pas complètement insensibles à l'information concernant la fréquence de base. À la suite des cinq descriptions de leur formulaire, les sujets pouvaient trouver la description «nulle» suivante :

Supposons maintenant qu'on ne vous donne aucune information concernant un individu choisi au hasard dans l'échantillon. La probabilité que cet homme soit un des 30 ingénieurs [ou pour l'autre groupe de sujets: un des 70 ingénieurs] dans un échantillon de 100 est de _____%

Dans ce cas, les sujets s'appuient entièrement sur la fréquence de base; l'estimé médian fut de 30% pour le premier groupe de sujet et de 70% pour le second. Dans leur discussion de ces expériences, Nisbett et Ross présentent l'interprétation suivante :

L'implication de ce contraste entre les conditions «aucune information» et les conditions «une

information totalement non-diagnostic» semble claire. Lorsqu'aucune évidence spécifique à propos d'un cas cible n'est fournie, les probabilités antécédentes sont utilisées de façon appropriée ; lorsque des évidences spécifiques sans valeur sont données, les probabilités antécédentes sont en général ignorées, et les gens répondent comme s'il n'y avait pas de base pour estimer qu'il y a des différences entre les probabilités relatives. La compréhension qu'ont les gens de la pertinence de l'information concernant la fréquence de base doit être très faible s'ils peuvent devenir réticents à l'utiliser du fait de la présence d'informations inutiles dans les cas cibles. (Nisbett et Ross 1980, 145-6)

Avant de quitter le sujet de l'oubli de la fréquence de base, je voudrais présenter un exemple supplémentaire illustrant le genre de conséquences pratiques sérieuses que le phénomène peut avoir. Voici un problème que Casscells et al. (1978) ont présenté à un groupe de professeurs, d'employés et d'étudiants de quatrième année de l'Ecole de médecine de Harvard.

Si un test pour détecter une maladie dont la prévalence est 1/1000 a un taux de faux-positif de 5%, quelle est la probabilité qu'une personne qui a un résultat positif ait vraiment la maladie, étant entendu que vous ne savez rien à propos des symptômes de cette personne ? _____%

Selon l'interprétation la plus plausible du problème, la réponse bayésienne correcte est 2%. Mais seulement dix-huit pourcent du groupe de Harvard donna une réponse de l'ordre de 2%. Quarante-cinq pourcent de ce groupe distingué ignora complètement l'information concernant la fréquence de base et affirma que la réponse était 95%. (Si, comme la plupart des médecins de Harvard, vous ne voyez pas pourquoi 2% est la bonne réponse, continuez votre lecture. Après avoir lu la section VI, la raison pour laquelle c'est la bonne réponse deviendra beaucoup plus claire). Il est un peu alarmant, pour ne pas dire plus, que les mêmes sujets expérimentaux aient posé des diagnostics concernant de vrais patients et leur offraient des conseils sur des questions telles que les traitements à suivre et la pertinence d'une opération.

La confiance excessive

Un des groupes de phénomènes les plus abondamment étudiés et les plus inquiétants qui aient été explorés par les psychologues intéressés par le raisonnement et le jugement met en cause le degré de confiance que les gens ont dans leurs réponses à des questions factuelles - des questions comme :

Dans chacune des paires suivantes, quelle ville a le plus d'habitants ?

- | | |
|--------------|---------------|
| a) Las Vegas | b) Miami |
| a) Sidney | b) Melbourne |
| a) Hyderabad | b) Islamabad |
| a) Bonn | b) Heidelberg |

Dans chacune des paires suivantes, quel événement s'est produit le premier ?

- | | |
|-----------------------------------|--------------------------------------|
| a) La signature de la Magna Carta | b) La naissance de Mahomet |
| a) La mort de Napoléon | b) L'achat de la Louisiane |
| a) L'assassinat de Lincoln | b) La naissance de la reine Victoria |

Après chaque réponse, on demandait également aux sujets :

À quel point êtes-vous confiant que votre réponse est correcte?
50% 60% 70% 80% 90% 100%

Dans une expérience utilisant des questions relativement difficiles, on découvre souvent que pour des cas où les sujets disent être confiants à 100%, seulement 80% de leurs réponses sont correctes; pour les cas où ils se disent confiants à 90%, seulement 70% de leurs réponses sont correctes; dans les cas où ils se disent confiants à 80%, seulement 60% de leurs réponses sont correctes. Cette tendance à la confiance excessive semble très forte. Le fait d'avertir les sujets que les gens sont parfois excessivement confiants n'a pas d'effet significatif, non plus que le fait de leur offrir de l'argent (ou des bouteilles de champagne) en récompense. Qui plus est, le phénomène a été démontré pour un grand nombre de groupes différents, comprenant des étudiants de premier cycle, des étudiants des cycles supérieurs, des médecins et même des analystes de la C.I.A. (pour un survol de la littérature, voir Lichtenstein, Fischhoff et Philips 1982).

Les découvertes empiriques dont nous venons de rendre compte ne sont qu'un petit échantillon de l'immense littérature empirique qui a été publiée au cours des 25 dernières années sur les défauts dans le raisonnement humain (pour des aperçus plus détaillés de la littérature de ce qui est parfois appelé la tradition des «heuristiques et préjugés», voir Nisbett et Ross 1980; Baron 1988; Piattelli-Palmarini 1994; Dawes 1988; Sutherland 1994. Kahneman, Slovic et Tversky 1982 constitue une anthologie très utile). Une conséquence apparemment inévitable de ce vaste corps de découvertes expérimentales est que la performance des gens dans un vaste domaine de problèmes inférentiels laisse beaucoup à désirer. Les réponses données par plusieurs sujets expérimentaux s'éloignent substantiellement et systématiquement des réponses qui seraient en accord avec un ensemble rationnel de principes inférentiels. Bien sûr, il se pourrait que toutes ces erreurs soient simplement des erreurs de performance et qu'elles ne reflètent pas exactement les principes de raisonnement qui forment la compétence sous-jacente du raisonnement ou la «psycho-logique» des sujets. Mais de nombreux auteurs ont proposé une interprétation beaucoup plus troublante de ces résultats expérimentaux. Ces auteurs soutiennent que dans des expériences comme celles décrites dans cette section, les gens raisonnent de fait en suivant des formes qui reflètent leur psycho-logique. Ces sujets n'utilisent pas les principes corrects simplement parce qu'ils n'y ont pas accès ; ces principes ne font pas partie de la compétence au raisonnement internalisée des sujets. Ce dont ils disposent plutôt, selon ce point de vue, est une collection de principes plus simples, ou «heuristiques», qui peuvent souvent donner lieu à la bonne réponse, mais souvent aussi conduire à une réponse erronée. Selon cette hypothèse, les sujets font des erreurs parce que leur psycho-logique est normativement défectueuse ; leurs principes de raisonnement internalisés ne sont pas des principes rationnels.

Daniel Kahneman et Amos Tversky, qui sont largement reconnus comme les fondateurs et des chercheurs de pointe dans la tradition des heuristiques et préjugés, expliquent cela de la manière suivante :

En faisant des prédictions et des jugements en situation d'incertitude, les gens ne semblent pas suivre le calcul des chances ou la théorie statistique de la prédiction. Au contraire, ils s'en remettent à un nombre limité d'heuristiques, qui parfois conduisent à des jugements raisonnables, et parfois à des erreurs graves et systématiques. (1973, 237)

Slovic, Fishhoff et Lichtenstein, d'importants chercheurs dans cette veine expérimentale, sont encore plus catégoriques. «Il semble, écrivent-ils, que les gens n'ont pas les bons programmes pour plusieurs tâches de jugement importantes [...] Nous n'avons pas eu l'opportunité d'acquérir à travers le processus de l'évolution un intellect capable de faire face conceptuellement à l'incertitude». (1976, 174). Stephen J. Gould, théoricien de l'évolution bien connu et auteur de nombreux ouvrages réputés sur la science, est du même avis. Après avoir décrit l'expérience de la «guichetière de banque», il demande: «Pourquoi faisons-nous constamment cette simple erreur de logique ?» Sa réponse est la suivante : «Tversky et Kahneman ont soutenu, correctement selon moi, que nos esprits ne sont pas construits (pour quelque raison que ce soit) pour utiliser les règles de la probabilité» (1992, 469). Si ces auteurs ont raison, alors Aristote a tort. L'homme n'est pas un animal rationnel.

V. Un défi pour l'explication de la rationalité en termes d'équilibre réfléchi

À la section précédente, nous avons examiné quelques résultats expérimentaux qui mettent en péril l'évaluation optimiste d'Aristote au sujet de la rationalité humaine. Un certain nombre de philosophes et de psychologues pensent aussi que de telles découvertes posent un défi majeur à l'explication par l'équilibre réfléchi, que nous avons ébauchée à la section III. L'argument justifiant cette conclusion est assez simple. Il commence par reconnaître que certaines des formes de raisonnement douteuses qu'on trouve dans la littérature peuvent être en équilibre réfléchi pour nombre de gens. Lorsque les principes sous-tendant les inférences sont rendus explicites et que les sujets peuvent réfléchir sur eux et sur leur propre pratique inférentielle, ils peuvent bien accepter à la fois les inférences et les principes. (Cette thèse trouve une illustration étonnante dans certains textes de logique au XIX^e siècle dans lesquels des principes problématiques sont explicitement adoptés². Les auteurs de ces textes acceptaient sans doute réflexivement à la fois les principes et les inférences qui s'accordaient avec ceux-ci.). Si tel est le cas, si les principes sous-tendant des formes d'inférences douteuses mentionnées dans la littérature psychologique sont vraiment en équilibre réfléchi pour de nombreuses personnes, alors, soutiennent les critiques, l'explication de la rationalité en termes d'équilibre réfléchi est erronée. Parce que selon cette explication, être rationnel consiste justement à passer le test de l'équilibre réfléchi. Si cette explication est correcte, alors l'erreur de la conjonction, l'oubli de la fréquence de base ou d'autres formes de raisonnement problématiques seraient rationnels pour ces gens.

Bien entendu, chaque exemple d'un principe inférentiel malheureux qui pourrait apparemment passer le test de l'équilibre réfléchi peut être remis en question. La question de savoir si les principes douteux qui semblent guider les pratiques inférentielles survivent ou non à la minutie réflexive qu'exige le test de Goodman est empirique. Et pour toute règle donnée, un tenant de l'explication en termes d'équilibre réfléchi pourrait soutenir que le cas empirique n'a pas été défini adéquatement. Je suis enclin à penser que ceux qui construisent leur défense de cette façon risquent d'être dépassés par l'abondance croissante des découvertes empiriques. Mais l'argument ne dépend pas nécessairement du fait que ce soupçon s'avérera fondé, parce que même la possibilité que les faits se présentent comme je soupçonne qu'ils le feront pose un sérieux problème à la thèse de l'équilibre réfléchi. Ce n'est assurément pas une vérité nécessaire que les principes inférentiels étranges échoueront le test de l'équilibre réfléchi pour tous les sujets. Et si on nous accorde, comme on doit clairement le faire, que la fréquence de base ou l'erreur de la conjonction (ou n'importe quel autre principe inférentiel qui a retenu l'attention des psychologues ces dernières années) peut passer le test de l'équilibre réfléchi pour un groupe de sujets quelconque, c'est suffisant pour jeter un doute sur l'idée que l'équilibre réfléchi est constitutif de la rationalité. Sûrement, la plupart d'entre nous ne sommes pas portés à dire qu'il est rationnel d'utiliser n'importe quel principe inférentiel - si bizarre qu'il soit - simplement parce qu'il s'accorde avec nos pratiques inférentielles réflexives.

Cela n'est pas, je m'empresse de le préciser, un argument décisif contre l'explication de la rationalité en termes d'équilibre réfléchi ; il est difficile de produire des arguments décisifs dans ce domaine. Quand ils rencontrent le fait que l'erreur de la conjonction, celle de l'oubli de la fréquence de base ou quelque autre principe que désapprouvent la plupart des théoriciens normatifs pourraient passer le test de l'équilibre réfléchi pour un quelconque groupe de gens réels ou hypothétiques, certains philosophes durcissent leur position et insistent sur le fait que si le principe en question est en équilibre réfléchi pour un groupe de gens, alors ce principe est en fait justifié ou rationnel pour eux. Cette évaluation, insistent-ils, s'accorde assez bien avec au moins un sens des notions de justification ou de rationalité pour être une « précification » de celles-ci plutôt que des notions protéiformes.

Bien que s'entêter et insister sur le fait que l'explication en termes d'équilibre réfléchi saisit (ou

précisifie) un sens de la rationalité ne constituent peut-être pas une position intenable, celle-ci présente manifestement des désavantages sérieux, le plus évident étant la conséquence très contre-intuitive qu'à peu près n'importe quelle règle de raisonnement, si loufoque qu'elle soit, pourrait être rationnelle pour une personne, en autant qu'elle s'accorde avec sa pratique inférentielle réflexive. Cela a amené de nombreux philosophes à construire une explication assez différente de ce qu'est un principe de raisonnement rationnel. Dans grand nombre d'explications, la notion de vérité joue un rôle central. Les tenants de ces explications considèrent d'entrée de jeu que le but véritable de la pensée et du raisonnement est de construire une explication exacte de la façon dont les choses sont dans le monde. Ce que nous voulons vraiment, ce sont des croyances vraies. Et si cela est juste, nous devrions employer des principes de raisonnement et de formation de croyances qui peuvent produire de vraies croyances. Ainsi, selon ces explications, un principe inférentiel est rationnel ou justifié si la personne qui l'utilise peut avoir des croyances vraies.

Malheureusement, de nombreuses explications de la rationalité articulées autour de la notion de vérité ont elles aussi des conséquences contre-intuitives. Peut-être la façon la plus facile de les voir est-elle d'invoquer une variation sur le thème du Malin Génie de Descartes. Imaginez deux personnes qui sont soudainement victimes du Malin Génie et qui reçoivent à partir de ce moment des inputs perceptuels systématiquement trompeurs ou illusoire. Supposons de plus qu'une de ces victimes a utilisé des processus inférentiels assez semblables aux nôtres et que ceux-ci ont produit des croyances vraies, alors que les processus inférentiels de l'autre victime sont assez fous (selon ce que nous pouvons juger) et n'ont pas réussi à produire des croyances vraies. Nous pouvons même imaginer que la seconde victime est internée dans un asile de fous et qu'elle est tourmentée par des hallucinations. Dans leur nouvel environnement créé par le Malin Génie, cependant, le système de principes inférentiels sain - celui qui est comme le nôtre - conduit à un tissu croissant de fausses croyances. L'autre système, par contre, fera un bien meilleur travail, produisant des croyances vraies et évitant les fausses croyances, puisque ce que le Malin Génie fait, c'est donner à la victime des apparences radicalement trompeuses - des apparences que seul un lunatique pourrait prendre comme preuve de ce qui se produit vraiment. Selon l'explication de la rationalité en fonction de la vérité, les principes inférentiels du lunatique seraient rationnels dans cet environnement, alors que les nôtres seraient plutôt irrationnels. Cela semble à plusieurs être un résultat sérieusement contre-intuitif. Les tenants de l'explication par la vérité ont proposé diverses stratégies pour éviter des cas comme ceux-ci, quoiqu'on débâte encore beaucoup du taux de succès de ces stratégies.

Où cela nous laisse-t-il? Le seul point sur lequel je crois qu'il y a un large consensus, c'est qu'il n'y a pas de réponses non problématique et généralement acceptée à la question par laquelle nous avons ouvert la section III : «Qu'est-ce qui justifie un ensemble de règles ou de principes pour le raisonnement ? Qu'est-ce qui rend les règles de raisonnement rationnelles ?» La nature de la rationalité est encore très discutée. De nombreux philosophes s'entendent également pour dire que les études empiriques sur le raisonnement (comme celles dont nous avons rendu compte à la section IV) imposent des contraintes importantes au genre de réponses qui peut être apporté à ces questions, bien qu'on ne sache encore trop quelles sont ces contraintes. Dans la section qui suit, nous retournerons à l'étude empirique du raisonnement et examinerons quelques résultats récents de la psychologie évolutionniste qui mettent en question la conclusion pessimiste à propos de la rationalité humaine que certains pourraient tirer des études dont nous avons rendu compte à la section IV. En interprétant ces résultats, nous n'avons d'autre choix de nous en remettre à nos intuitions sur la rationalité, puisqu'il n'y a pas de théorie généralement acceptée sur ce qu'est la rationalité.

VI. Les humains sont-ils des animaux rationnels après tout? Les données de la psychologie évolutionniste

Même si le champ interdisciplinaire de la psychologie évolutionniste est trop nouveau pour avoir développé un corps de doctrine précis et largement accepté, il y a trois thèses fondamentales qui sont clairement reçues : la première est que l'esprit humain contient un grand nombre de systèmes destinés à des fins particulières - souvent appelés «modules» ou «organes mentaux». Ces modules sont invariablement conçus comme un type de mécanisme computationnel spécialisé ou spécifique à un domaine. La seconde thèse centrale de la psychologie évolutionniste est que, contrairement à ce qui a été maintenu par certains éminents chercheurs en sciences cognitives (plus particulièrement Jerry Fodor 1983), la structure modulaire de l'esprit n'est pas limitée aux «systèmes d'inputs» (ceux qui sont responsables de la perception et du traitement du langage) et aux «systèmes d'outputs» (ceux qui sont responsables de la production des mouvements du corps). Selon les psychologues évolutionnistes, les modules servent plusieurs de celles que l'on nomme «capacités centrales», comme le raisonnement et la formation des croyances. La troisième thèse est que les modules mentaux sont ce que les biologistes évolutionnistes ont appelé des adaptations - ils ont été, comme l'écrivent Tooby et Cosmides, «inventés par la sélection naturelle pendant l'histoire évolutionniste de notre espèce pour produire des fins adaptatives dans l'environnement naturel de l'espèce» (Tooby et Cosmides 1995, xiii). Voici un passage dans lequel Tooby et Cosmides offrent une image particulièrement vive de ces thèses centrales de la psychologie évolutionniste :

Notre architecture cognitive ressemble à une confédération de centaines de milliers d'ordinateurs fonctionnels (souvent appelés modules) conçus pour résoudre des problèmes adaptatifs endémiques pour nos ancêtres chasseurs-cueilleurs. Chacun de ces mécanismes possède son propre ordre du jour et impose sa propre organisation à différents fragments du monde. Il y a des systèmes spécialisés dans l'induction de la grammaire, dans la reconnaissance des visages, dans le positionnement dans l'espace, dans la construction des objets et dans la reconnaissance des émotions à partir des expressions faciales. Il y a des mécanismes pour détecter le caractère animé des choses, la direction des yeux et la tricherie. Il y a un module de la «théorie de l'esprit» [...] une variété de modules d'inférence sociale [...] et une multitude d'autres élégantes machines.» (Tooby et Cosmides 1995, xiv)

Si la plus grande partie de la cognition centrale est en fait servie par des modules cognitifs qui ont été conçus pour faire face aux problèmes adaptatifs posés par l'environnement dans lequel nos ancêtres primates ont vécu, il faut s'attendre à ce que les modules responsables du raisonnement accomplissent leur meilleur travail lorsque l'information qui leur est acheminée se trouve dans un format similaire à celui dans lequel l'information était accessible dans l'environnement ancestral. Et, comme l'a soutenu Gerd Gigerenzer, même s'il y avait une quantité d'information probabiliste disponible dans cet environnement, cette information aurait été représentée «en termes de fréquences d'événements, encodée séquentiellement au fur et à mesure des rencontres - par exemple, 3 sur 20 par opposition à 15% ...» (Gigerenzer 1994, 142). Cosmides et Tooby soutiennent à peu près la même chose :

Nos ancêtres hominidés étaient plongés dans un flot de fréquences observables qui pouvaient être utilisées pour améliorer la prise de décision et ils développaient les procédures qui pouvaient en profiter. Ainsi, si nous avons des adaptations pour le raisonnement inductif, elles doivent prendre l'information en termes de fréquence comme input. (1996, 15-16)

Sur la base de ces considérations évolutionnistes, Gigerenzer, Cosmides et Tooby ont proposé et défendu une hypothèse psychologique à laquelle ils réfèrent comme l'hypothèse fréquentiste : «Certains de nos mécanismes de raisonnement inductif incorporent des aspects du calcul des probabilités, mais ils sont conçus pour prendre l'information en termes de fréquence comme input et produire des fréquences comme output» (Cosmides et Tooby 1996, 3).

Cette idée a conduit Cosmides et Tooby à poursuivre une intéressante série d'expériences dans laquelle le «problème de l'Ecole de médecine de Harvard» que nous avons discuté à la section IV a été

systématiquement transformé en un problème dans lequel à la fois la question et la réponse demandent d'être formulées en termes de fréquences. Voici un exemple de leur étude dans lequel l'information concernant la fréquence est rendue particulièrement manifeste :

1 Américain sur 1000 a une maladie X. Un test a été mis au point pour détecter cette maladie. Chaque fois que le test est administré à une personne qui a la maladie, le test est positif. Mais parfois le test est aussi positif lorsqu'il est administré à une personne qui est en parfaite santé. Plus précisément, dans chaque groupe de 1000 personnes en parfaite santé, 50 recevront un diagnostic positif.

Imaginez que nous ayons réuni au hasard un échantillon de 1000 Américains, choisis par tirage au sort. Ceux qui ont fait le tirage n'avaient aucune information sur l'état de santé de ces gens.

Étant donné les informations mentionnées, en moyenne, combien de gens dont le diagnostic est positif pour la maladie l'ont-ils vraiment ? _____ sur _____.

À la différence de l'expérience originale de Casscells et al., dans laquelle seulement 18% des sujets avaient la réponse bayésienne correcte, 76% des sujets de Cosmides et Tooby ont cette fois donné la bonne.

Cela n'est pas seulement un cas isolé dans lequel une «version fréquentiste» des problèmes de raisonnement probabiliste permet des niveaux de performance élevés. Au contraire, il s'avère que dans de nombreux cas, lorsque les problèmes sont formulés en termes de fréquences plutôt que de probabilités, les sujets tendent à raisonner plus rationnellement. Dans une étude, Fiedler (1988) a montré que le pourcentage de sujets qui commettaient l'erreur de la conjonction pouvait être radicalement réduit si le problème était formulé en termes fréquentistes. Utilisant le problème de la «guichetière féministe», Fiedler contraste le libellé rapporté à la section IV avec un problème qui se lit comme suit :

Linda a 31 ans, elle est célibataire, franche et très brillante. Elle a une majeure en philosophie. Quand elle était étudiante, elle était très préoccupée par les questions de discrimination et de justice sociale, et elle a aussi participé à des manifestations antinucléaires.

Il y a deux cents personnes à qui cette description convient. Combien seront :

guichetières de banque

.....

guichetières de banque et actives dans le mouvement féministe?

.....

Dans la formulation originale du problème, 91% des sujets ont jugé que le cas de la guichetière féministe est plus probable que le cas de la guichetière. Toutefois, dans la version fréquentiste présentée ci-haut, seulement 22% des sujets jugent qu'il y aura plus de guichetières féministes que de guichetières.

Des études sur la confiance excessive ont également été conduites pour étayer l'hypothèse fréquentiste. Dans une de celles-ci, Gigerenzer, Hoffrage et Kleinbölting (1991) rapportent que la sorte de confiance excessive décrite à la section IV peut «disparaître» si on fait répondre les sujets à des questions formulées en termes de fréquences. Gigerenzer et ses collègues donnèrent une liste de 50 questions similaires à celles que nous avons vues à la section IV, à ceci près qu'en plus d'évaluer leur confiance après chaque réponse (ce qui, en effet, leur demande de juger la probabilité d'un événement singulier), on

posait également au sujet des questions à propos de la fréquence de réponses correctes : «À combien de ces 50 questions pensez-vous avoir donné une bonne réponse ?». Dans deux expériences, la moyenne de confiance excessive était environ de 15% lorsque la confiance en des événements singuliers était comparée à la fréquence relative réelle de réponses correctes, reproduisant le genre de données présenté à la section IV. Cependant, comparer «la fréquence estimée avec la fréquence réelle de bonnes réponses» des sujets «faisait disparaître la «confiance excessive» [...] Les fréquences estimées étaient pratiquement identiques aux fréquences actuelles ... «L'illusion cognitive» avait disparu» (Gigerenzer et al. 1991a, 89).

À la section IV, nous avons vu une version de la tâche de sélection des quatre cartes de Wason, pour laquelle la plupart des sujets parvenaient à des résultats médiocre. On avait noté qu'alors que les résultats étaient tout aussi médiocres dans de nombreuses autres versions de la tâche de sélection, il y avait des versions où la performance des sujets s'améliorait radicalement. Voici un exemple de Griggs et Cox (1982).

Dans leur lutte contre l'ivresse au volant, les officiers de police du Massachusetts ne cessent de retirer des permis de conduire. Vous êtes videur dans un bar de Boston et vous perdrez votre emploi si vous n'appliquez pas la loi suivante :

« Pour boire de la bière, il faut avoir plus de 20 ans »

Les cartes ci-après contiennent des renseignements sur quatre personnes assises à une table dans votre bar. Sur un côté de chaque carte se trouve la nature de la consommation et sur l'autre, l'âge de la personne. N'identifiez que les cartes que vous devez absolument retourner pour savoir si l'une de ces personnes enfreint la loi.

bière coca 25 ans 16 ans

D'un point de vue logique, ce problème semble structurellement identique au problème de la section IV, mais le contenu des problèmes a clairement un effet majeur sur la correction des réponses données. A peu près 75% des étudiants de collège ont donné la bonne réponse à cette version de la tâche de sélection, alors que 25% ont eu la bonne réponse dans l'autre version. Quoiqu'il y ait eu des douzaines d'études explorant cet «effet de contenu» sur la tâche de sélection, jusqu'à tout récemment ces résultats étaient relativement incompréhensibles puisqu'il n'y avait pas de propriété évidente ou un groupe de propriétés partagées par ces versions de la tâche de sélection pour lesquelles le taux de réponses correctes était élevé. Cependant, dans plusieurs articles récents, Cosmides et Tooby ont soutenu que l'analyse évolutionniste nous permet de déceler un motif surprenant dans ce qui autrement constitue des résultats déroutants (Cosmides 1989, Cosmides et Tooby 1992).

Le point de départ de leur analyse évolutionniste est l'observation que dans l'environnement dans lequel nos ancêtres ont évolué (mais dans le monde moderne également) il arrivait souvent que des individus non apparentés participaient à ce que les théoriciens des jeux ont appelé des échanges à «somme non nulle», dans lesquels les bénéfices du récipiendaire (mesurés en termes d'adaptation reproductive) étaient beaucoup plus grands que les coûts pour le donneur. Dans une société de chasseurs-cueilleurs, par exemple, il arrivera parfois qu'un chasseur a été chanceux un jour donné et qu'il a de la nourriture en abondance, alors qu'un autre chasseur a été malchanceux et meurt de faim. Si le chasseur chanceux donne un peu de sa viande au chasseur malchanceux plutôt que s'en gaver, cela peut avoir un petit effet négatif sur l'adaptation du donneur puisque ces quelques grammes supplémentaires de gras corporel qu'il aurait pu consommer auraient pu être utiles dans le futur, mais le bénéfice pour le récipiendaire est beaucoup plus grand. Il y a quand même un quelconque coût pour le donneur ; il se

porterait un peu mieux s'il n'avait pas aidé ces individus non apparentés. Il est quand même clair que les gens aident parfois ceux qui ne leur sont pas apparentés, et il y a des raisons de penser que les primates non humains (ainsi que les chauves-souris) font la même chose. À première vue, cette sorte d'«altruisme» semble poser un casse-tête évolutionniste puisque si un gène qui rend un organisme moins enclin à aider les individus non apparentés apparaît dans une population, ceux qui auront ce gène seront un petit peu plus adaptés, de sorte que ce gène se répandra graduellement dans la population.

Une solution a été proposée par Robert Trivers (1971), qui avait noté que, alors que l'altruisme unidirectionnel était une mauvaise idée d'un point de vue évolutionniste, l'altruisme réciproque était une autre histoire. Si deux chasseurs (qu'ils soient des humains ou des chauves-souris) peuvent compter l'un sur l'autre lorsque l'un a de la nourriture en abondance et que l'autre n'en a pas, alors cela devrait profiter à tous deux à long terme. Ainsi, les organismes avec un gène ou une suite de gènes qui les rendent enclins à participer à des échanges réciproques avec des non-parents (des «échanges sociaux», comme on les nomme parfois) seraient mieux adaptés que les membres de la même espèce qui n'ont pas ces gènes. Mais, bien sûr, les arrangements d'échange réciproque sont exposés à la tricherie. Dans le cours de la maximisation de l'adaptation, les individus s'en tireront mieux si on leur offre régulièrement et s'ils acceptent de l'aide quand ils en ont besoin, mais ne rendent jamais la pareille lorsque les autres ont besoin d'aide. Cela suggère que si des arrangements d'échanges sociaux peuvent exister, les organismes impliqués doivent avoir des mécanismes qui leur permettent de détecter les tricheurs et d'éviter de les aider dans le futur. Et puisque les humains sont apparemment capables de participer à des relations d'échanges sociaux stables, cette analyse a conduit Cosmides et Tooby à faire l'hypothèse que nous avons un ou plusieurs modules ou organes mentaux dont le travail consiste à reconnaître les arrangements d'échanges réciproques et à détecter les tricheurs qui acceptent les bénéfices de tels arrangements, mais qui n'en paient pas les coûts. Bref, l'analyse évolutionniste conduit Cosmides et Tooby à faire l'hypothèse de l'existence d'un ou plusieurs modules de détection des tricheurs. J'appellerai cela l'hypothèse de la détection des tricheurs.

Si cette hypothèse est correcte, nous devrions découvrir quelques traces de l'existence de ces modules dans la pensée des humains contemporains. C'est ici que la tâche de sélection fait son entrée. En fait, selon Cosmides et Tooby, quelques versions de la tâche de la sélection utilisent un ou plusieurs modules mentaux conçus pour détecter les tricheurs dans les situations d'échange social. D'autres versions de la tâche n'activent pas les modules d'échanges sociaux et de détection des tricheurs. Puisque nous n'avons pas de modules mentaux pour venir à bout de ces problèmes, les gens les trouveront plus difficiles, et leurs performances seront beaucoup moins bonnes. Le problème du videur du bar de Boston présenté plus tôt est un exemple d'une tâche de sélection qui déclenche un mécanisme de détection des tricheurs. Le problème des voyelles et des nombres impairs présenté à la section IV est un exemple de tâche de sélection qui ne déclenche pas de module de détection de tricheurs.

Afin d'étayer leur théorie, Cosmides et Tooby ont réuni un ensemble de données impressionnant. L'hypothèse de la détection des tricheurs soutient que les échanges sociaux ou les «contrats sociaux» donneront lieu à une bonne performance aux tâches de sélection, et cela nous permet de déceler un motif clair dans cette littérature expérimentale déconcertante qui s'est développée avant que leur hypothèse ne soit formulée.

Lorsque nous avons entrepris notre recherche en 1983, la littérature sur la tâche de sélection de Wason était remplie de rapports au sujet d'une grande variété d'effets de contenu, et il n'y avait pas de théorie satisfaisante ou de généralisation empirique qui pouvait rendre compte de ces effets. Quand nous avons catégorisé ces effets de contenu selon qu'ils se conforment ou non aux contrats sociaux, un motif frappant est apparu. On n'a trouvé des effets de contenu solides et qu'on peut reproduire que pour les règles qui lient des termes qui sont reconnaissables comme bénéfique et coût/obligation dans le format des contrats sociaux standard [...] Aucune règle thématique qui n'était pas un contrat social n'a jamais produit

un effet de contenu aussi solide ... Étant donné ce que nous savons, pour les problèmes thématiques de contrats non sociaux, 3 expériences ont produit un substantiel effet de contenu, 2 ont produit un effet de contenu faible, et 14 n'ont produit aucun effet de contenu. En revanche, 16 expériences sur 16 qui se conformaient au critère pour les contrats sociaux standard [...] ont provoqué des effets de contenu substantiels. (Cosmides et Tooby 1992, 183)

Depuis la formulation de l'hypothèse de la détection des tricheurs, de nombreuses autres expériences ont été élaborées pour tester l'hypothèse et éliminer les autres possibilités. Et même si l'hypothèse a toujours de nombreux critiques, il n'y a pas de doute que certaines de ces expériences impressionnent par leur caractère concluant.

De tels résultats ont favorisé le retour de l'optimisme aristotélicien. Selon le point de vue des optimistes, les théories et découvertes de la psychologie évolutionniste suggèrent que le raisonnement humain est servie par «d'élégantes machines» qui ont été conçues et raffinées par la sélection naturelle pendant des millions d'années, et c'est pourquoi toute inquiétude quant à une quelconque irrationalité systématique serait dénuée de fondement. Le titre que Cosmides et Tooby ont choisi pour le texte dans lequel ils font état de leurs données concernant le problème de l'Ecole de médecine de Harvard est un bon indicateur de cet optimisme: «Est-ce que les humains sont de bons statisticiens intuitifs après tout? Repenser certaines conclusions de la littérature sur le jugement en situation d'incertitude». Cinq ans plus tôt, alors que la recherche de Cosmides et Tooby était toujours en cours, Gigerenzer avait rapportés quelques-unes de leurs premières découvertes dans un article au titre provocateur: «Comment faire disparaître les illusions cognitives: au-delà des «heuristiques et préjugés»». Chacun de ces deux titres suggère clairement que les découvertes dont il est fait état attaquent de front le pessimisme caractéristique de la tradition des heuristiques et préjugés. Et il ne s'agit pas que des titres. Article après article, Gigerenzer a fait des déclarations du type : « il nous faut être plus optimistes » (1991b, 245) ou « nous ne devons pas forcément nous en faire à propos de la rationalité humaine » (1998, 280) et il a soutenu qu'on peut considérer « l'intuition comme fondamentalement rationnelle » (1991b, 242). Au vu de commentaires de ce genre, il n'est guère étonnant qu'un commentateur ait dit de Gigerenzer et de ses collègues qu'ils ont « adopté une position empirique contre l'idée entretenue par certains psychologues que les gens sont plutôt stupides » (Lopes, cité par Bower, 1996).

VII. La psychologie évolutionniste montre-t-elle qu'Aristote avait raison?

Les théories préconisées par les psychologues évolutionnistes sont très controversées (si ce n'est plus), et même leurs expériences n'ont pas été sans soulever des contestations. Mais supposez qu'il s'avère que les psychologues évolutionnistes aient raison à propos des mécanismes mentaux qui sous-tendent le raisonnement humain. Est-ce que cela montrerait vraiment que la thèse d'Aristote était correcte? La réponse est, je crois, pas du tout. Pour voir pourquoi, commençons par rappeler comment nous interprétons la thèse d'Aristote. L'affirmation selon laquelle les humains sont des animaux, comme nous l'avons expliqué à la section II, signifie que les gens normaux ont une compétence de raisonnement rationnelle. Cette compétence est un ensemble de règles ou de principes de raisonnement mentalement représenté - une psycho-logique - et la thèse d'Aristote est que ces règles sont rationnelles ou normativement appropriées; elles spécifient la façon correcte de raisonner. Ainsi lorsque les gens font des erreurs de raisonnement ou qu'ils raisonnent irrationnellement, les erreurs sont des erreurs de performance qui ne reflètent pas les principes représentés mentalement qui sous-tendent le raisonnement.

La première chose à dire au sujet de la relation entre la psychologie évolutionniste et la thèse d'Aristote est que si la psychologie évolutionniste a raison, alors notre interprétation de la thèse d'Aristote est trop simpliste pour s'accorder avec les faits. Selon le modèle de l'esprit qu'ont les psychologues

évolutionnistes, les gens n'ont pas une compétence de raisonnement, ils en ont beaucoup, et chacun de ces modules mentaux « à tâche spécifique » possède son propre ensemble de règles. Ainsi il n'y a pas une seule psycho-logique non plus; il y en a beaucoup. On pourrait penser que cela ne pose qu'un problème mineur aux défenseurs de l'optimisme aristotélicien. Plutôt qu'affirmer qu'il n'y a qu'un mécanisme sous-tendant le raisonnement et qu'il comprend un ensemble de règles rationnellement ou normativement appropriées, ils pourraient soutenir qu'il y a de nombreux mécanismes sous-tendant le raisonnement et que tous utilisent des règles normativement appropriées. Mais, et cela est un fait déterminant, la psychologie évolutionniste n'étaye pas cette affirmation. La psychologie évolutionniste maintient que la sélection naturelle nous a équipés d'un nombre de mécanismes de raisonnement bien conçus qui emploient des principes rationnellement ou normativement appropriés pour le genre de problèmes qui étaient importants dans l'environnement des chasseurs/cueilleurs qu'étaient nos ancêtres. Il y a toutefois de nombreuses sortes de problèmes de raisonnement qui sont importants dans le monde moderne - des problèmes impliquant la probabilité d'événements uniques, par exemple - que ces mécanismes ne sont pas conçus pour résoudre. Dans plusieurs cas, suggèrent les psychologues évolutionnistes, les élégants mécanismes de raisonnement spécifique à une tâche conçus par la sélection naturelle ne pourront même pas traiter ces problèmes. Plusieurs des problèmes étudiés par la littérature de «heuristiques et préjugés» étaient de ce genre. Et la psychologie évolutionniste ne nous donne aucune raison de supposer que les gens ont des règles rationnelles représentées mentalement pour faire face à des problèmes comme ceux-là.

Selon l'interprétation de la littérature expérimentale sur le raisonnement que je défends, rien ne justifie ni l'optimisme aristotélicien ni le pessimisme de ceux qui ont suggéré que nos esprits n'étaient équipés que d'un «programme de mauvaise qualité». Nous avons et utilisons quelques programmes remarquablement bons pour venir à bout des problèmes qui étaient importants dans l'environnement où notre espèce a évolué. Mais il y a également des lacunes importantes dans le genre de problèmes dont notre esprit peut venir à bout. Le défi pour les philosophes, psychologues et éducateurs des prochaines décennies sera de concevoir de meilleures façons pour nos esprits de l'âge de pierre de traiter les problèmes de raisonnement que nous rencontrons à l'âge des voyages spatiaux, des réseaux d'ordinateurs globaux et des armes nucléaires.

(Traduit de l'anglais par Luc Faucher et François Latraverse)

Bibliographie

Baron, J. (1988). *Thinking and Deciding*, Cambridge, Cambridge University Press.

Baron-Cohen, S. (1995), *Mindblindness: An Essay on Autism and Theory of Mind*, Cambridge, MA, MIT Press.

Bower, B. (1996), « Rational mind design: research into the ecology of thought treads on contested terrain », *Science News*, 150, 24-25.

Casscells, W., Schoenberger, A. and Grayboys, T. (1978), « Interpretation by physicians of clinical laboratory results », *New England Journal of Medicine*, 299, 999-1000.

Chomsky, N. (1965), *Aspects of the Theory of Syntax*, Cambridge, MA, MIT Press.

Chomsky, N. (1980), *Rules and Representations*, New York, Columbia University Press.

- Coppée, H. (1974), *Elements of Logic*, Revised Edition, Philadelphia, H. Butler & Co.
- Cosmides, L. (1989), « The logic of social exchange: Has natural selection shaped how humans reason? Studies with Wason Selection Task », *Cognition*, 31, 187-276.
- Cosmides, L. and Tooby, J. (1992), « Cognitive adaptations for social exchange », in Barkow, Cosmides and Tooby (1992), 163-228.
- Cosmides, L. and Tooby, J. (1996), « Are humans good intuitive statisticians after all? Rethinking some conclusions from the literature on judgment under uncertainty », *Cognition*, 58, 1, 1-73.
- Dawes, R. (1988), *Rational Choice in an Uncertain World*, Orlando, Harcourt Brace Jovanovich.
- Evans, J. Newstead, S. & Byrne, R. (1993), *Human Reasoning: The Psychology of Deduction*, Hove, Erlbaum.
- Fiedler, K. (1988), « The dependence of the conjunction fallacy on subtle linguistic factors », *Psychological Research*, 50, 123-129.
- Fodor, J. (1983), *The Modularity of Mind*, Cambridge, MA, MIT Press.
- Gigerenzer, G. (1991a), « How to make cognitive illusions disappear: Beyond 'heuristics and biases' », *European Review of Social Psychology*, 2, 83-115.
- Gigerenzer, G. (1991b), « On Cognitive Illusions and Rationality », *Poznan Studies in the Philosophy of the Sciences and the Humanities*, Vol. 21, 225-249.
- Gigerenzer, G. (1994), « Why the distinction between single-event probabilities and frequencies is important for psychology (and vice versa) », in G. Wright and P. Ayton, *Subjective Probability*, New York: John Wiley.
- Gigerenzer, G. (1998), « The modularity of social intelligence », in A. Whiten and R. Byrne, *Machiavellian Intelligence II*, Cambridge: Cambridge University Press.
- Gigerenzer, G., Hoffrage, U., and Kleinbölting, H. (1991), « Probabilistic mental models: A Brunswikean theory of confidence », *Psychological Review*, 98, 506-528.
- Goodman, N. (1984), *Faits, fictions et prédictions*, Paris, Minuit.
- Gould, S. (1992), *Bully for Brontosaurus. Further Reflections in Natural History*, London, Penguin Books.
- Griggs, R. and Cox, J. (1982), « The elusive thematic-materials effect in Wason's selection task », *British Journal of Psychology*, 73, 407-420.
- Kahneman, D. and Tversky, A. (1973), « On the psychology of prediction », *Psychological Review*, 80, 237-251. Repris dans Kahneman, Slovic and Tversky (1982), 48-68.
- Kahneman, D., Slovic, P. and Tversky, A. (dir. de la publ.). (1982), *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge, Cambridge University Press.

Lichtenstein, S., Fischhoff, B. and Phillips, L. (1982), « Calibration of probabilities: The state of the art to 1980 », in Kahneman, Slovic and Tversky (1982), 306-334.

Nisbett, R. and Ross, L. (1980), *Human Inference: Strategies and Shortcomings of Social Judgment*, Englewood Cliffs, Prentice-Hall.

Piattelli-Palmarini, M. (1994), *Inevitable Illusions: How Mistakes of Reason Rule Our Minds*, New York, John Wiley & Sons.

Slovic, P., Fischhoff, B., and Lichtenstein, S. (1976), « Cognitive processes and societal risk taking », in J. S. Carol and J. W. Payne (dir. de la publ.), *Cognition and Social Behavior*, Hillsdale, Erlbaum.

Sutherland, S. (1994), *Irrationality: Why We Don't Think Straight!*, New Brunswick, Rutgers University Press.

Tooby, J. and Cosmides, L. (1995), « Foreword », in Baron-Cohen (1995), xi-xviii.

Trivers, R. (1971), « The evolution of reciprocal altruism », *Quarterly Review of Biology*, 46, 35-56.

Tversky, A. and Kahneman, D. (1982), « Judgments of and by representativeness », in Kahneman, Slovic and Tversky (1982), 84-98.

¹ Le théorème de Bayes, nommé en l'honneur du Révérent Thomas Bayes, stipule que $p(H/D) = p(D/H) \times p(H) / p(D)$ où $p(H/D)$ est la «probabilité conditionnelle » de l'hypothèse, H, étant donné que D (les données) est vraie, $p(D/H)$ est la probabilité conditionnelle de D, étant donné que H est vrai, $p(H)$ est la probabilité «antérieure" que H soit vraie et $p(D)$ est la probabilité que D soit vraie.

² Par exemple, dans *Elements of Logic*, publié en 1974, Henry Coppée approuve explicitement la fameuse «erreur du parieur" (Gambler's Fallacy). Voici ce qu'il dit: « Ainsi, lorsque nous lançons un dé, nous ne pouvons pas être certains qu'aucune face ou combinaison de faces apparaîtra; mais si, à la suite de plusieurs lancers, une face particulière n'est pas apparue, ses chances d'apparaître deviennent de plus en plus fortes, jusqu'à ce qu'elles approchent très près de la certitude. Cela doit se produire; et à chaque lancer où elle n'apparaît pas, la certitude de son occurrence devient de plus en plus grande ».