

In J. Dubucs & P. Lepage, dir., *Méthodes logiques pour les sciences cognitives*, Paris : Hermes, 1995, pp. 25-75.

Repris dans D. Andler, dir., *Introduction aux sciences cognitives*, Paris : Gallimard, nouv. éd. 2004, pp. 31-405.

LOGIQUE, RAISONNEMENT ET PSYCHOLOGIE

Les rapports entre logique et raisonnement ont longtemps été fort étroits. Ce commerce s'était établi à l'abri de la digue que Frege, Husserl et leurs successeurs avaient érigée pour protéger la logique de la psychologie : discipline normative ayant pour objet la manière dont l'homme doit raisonner, elle était en droit à l'abri de toute espèce d'enquête empirique sur la manière dont les humains raisonnent en fait, ou sur la manière dont ils apprennent à le faire. Pour faire bref, on pourrait parler d'une « ancienne alliance » entre logique, raisonnement et antipsychologisme.

Les psychologues, les pédagogues, les philosophes, les anthropologues et la sagesse commune, qui fournissaient mille témoignages sur la prévalence d'erreurs logiques et de modes de pensée illogiques ou alogiques, loin de faire peser une menace sur cette alliance, ne faisaient que renforcer sa légitimité. Elle était la gardienne d'un domaine particulier, celui de la rationalité.

Plus délicate était apparemment la situation créée par la nécessité, dans les sciences et dans certains domaines de la vie pratique, de recourir à des procédures inductives. Le calcul des probabilités et la recherche d'une logique inductive furent les principales et efficaces ripostes. Il y aurait beaucoup à dire sur les difficultés rencontrées d'une part pour fonder le calcul des probabilités, d'autre part pour découvrir une logique inductive — et il faudra y revenir même dans le cadre restreint du présent article. Mais il est évident que la psychologie « réelle » du savant, de l'assureur, du médecin ou du chasseur primitif n'était pour rien dans ces difficultés : l'« alliance » dont nous parlons n'avait rien à craindre de ce côté-là.

Tout a changé. Non pas précisément sous l'effet des premières recherches en intelligence artificielle et en psychologie cognitive, qui s'attachaient à l'explication et à la simulation des comportements humains en matière de résolution de problèmes : les heuristiques et les stratégies de l'IA première manière et du *human problem solving* dans le style de Herbert Simon et Allen Newell ne faisaient que reprendre, avec — du moins en principe — la rigueur expérimentale et la précision de la programmation en plus, les thèmes familiers de la découverte mathématique et scientifique et de la logique inductive. Mais c'est bien de la même région du savoir, à savoir les sciences cognitives au sens large que sont parties les attaques convergentes qui ont mis à bas l'ancienne alliance.

L'analyse complète de cette convergence nous entraînerait trop loin. Rappelons seulement ses facteurs les plus importants. *Primo*, les difficultés rencontrées par l'intelligence artificielle première manière ont conduit de nombreux chercheurs du domaine à récuser en bloc la logique comme *organon* de la raison. *Secundo*, la prolifération de « logiques » développées pour la modélisation du raisonnement « naturel » ou scientifique, les difficultés rencontrées d'une part pour montrer qu'elles sont à la hauteur de la tâche, d'autre part pour leur donner une base formelle convenable, tout cela a conduit certains logiciens, dont Gilbert Harman, à mettre en cause, sur le plan proprement conceptuel, le lien entre logique et raisonnement, et à s'interroger avec une vigueur nouvelle sur la nature véritable et sur le statut cognitif de la logique. *Tertio*, l'émergence d'une solution de rechange possible à l'approche dite « symbolique » ou « formaliste » dans les sciences cognitives, solution fondée sur l'outil technique et conceptuel des réseaux de neurones formels, a donné du poids à l'idée que certains comportements peuvent sembler être déterminés par des règles alors qu'ils ne le sont pas : si la logique semble commander, en fait ou en droit, le raisonnement, il se pourrait donc qu'il n'en soit rien (ni en fait *ni en droit*). Enfin, un considérable ensemble de recherches relevant de la psychologie expérimentale et destinées à déterminer les aptitudes réelles de l'homme aux tâches déductives et inductives les plus élémentaires a mis en évidence des écarts systématiques entre les performances observées et les réponses dictées par les normes logiques. Ces résultats ont conduit certains auteurs à des conclusions très générales sur les aptitudes de l'homme en matière de rationalité. D'un autre côté, des théories ont été élaborées dans le but de rendre compte de ces résultats à partir des processus cognitifs qui sont

effectivement mis en œuvre dans les tâches de raisonnement. Ces théories sont fortement divergentes et leur confrontation, sur le plan tant conceptuel qu'empirique, a donné naissance à un débat nourri. D'une complexité parfois déroutante, ce débat a un enjeu principal assez clair, qui n'est autre que le rôle que la logique joue dans le comportement humain.

Nous nous intéresserons dans le présent article principalement au dernier de ces quatre courants, même si tous sont au moins indirectement concernés ; le deuxième le sera plus particulièrement. Mais avant d'entreprendre l'examen de la psychologie du raisonnement, il peut être utile de la situer dans le contexte de ce qu'on pourrait appeler, par symétrie, la « nouvelle alliance ». Il s'agit du rapprochement de courants hétéroclites et encore mal soudés qui ont en commun de rejeter la logique comme ressource ou capacité cognitive fondamentale, soit dans les faits, soit même en droit. La « nouvelle alliance » est antilogiciste de deux manières : elle nie que la logique soit à la base d'une vaste famille de comportements, ceux que la tradition qualifie de rationnels ; et elle nie que la logique soit la norme ou le patron d'une catégorie — ou d'une dimension — importante de ces comportements ; beaucoup plus modestement, la logique prendrait place dans un répertoire varié de techniques cognitives. « La » logique, à supposer qu'il convienne encore de s'y intéresser, serait d'ailleurs, de l'aveu même des logiciens vraiment modernes, mal nommée : il n'y a que des logiques, en quantité indéfinie, qui sont comme les outils d'une trousse de bricoleur (v. par exemple Gamut 1991). Encore n'est-on par certain que ces logiques remplissent collectivement les fonctions dévolues, dans l'ancienne alliance, à la logique : nombre d'auteurs placent au cœur de la cognition des capacités telles que la manipulation de répertoires d'exemples, la reconnaissance de configurations, la construction d'images ou de modèles — bref, dans l'avant-garde de la nouvelle alliance on cherche à donner corps, de diverses manières, au rejet du *formel* : les mécanismes ou processus postulés ont en commun de faire intervenir, en un sens ou en un autre, mais toujours de manière décisive, un *contenu*¹.

Nous ne procéderons pas à l'évaluation systématique des positions de la nouvelle alliance, encore moins chercherons-nous à restaurer l'ancienne, puisque nous sommes acquis à certains des arguments qui ont été invoqués pour la discréditer. Notre tâche sera plus limitée, nous l'avons dit, et plus descriptive ; toutefois, nous serons amenés à insister sur ce qui nous apparaît comme des confusions de nature à compromettre une bonne partie des arguments de portée générale qu'on a cru pouvoir tirer de l'étude expérimentale du raisonnement quant aux aptitudes logiques et à la rationalité de l'homme. Cette attitude critique n'a rien d'original, et pour une bonne part le travail a été fait par d'autres. Notons cependant au passage qu'aucun des camps en présence n'a le monopole des confusions : ceux qui en dénoncent certaines à juste titre en commettent d'autres parfois aussi préoccupantes. C'est un véritable mystère que cette capacité que semble avoir la logique de faire tourner les têtes les plus solides. Puisse l'auteur n'avoir pas commis à son tour d'erreur plus grave que celles qu'il reproche aux autres !

Nous commencerons par examiner les résultats et les théories en présence dans le domaine de la déduction (§§1-2), en mettant l'accent sur la tâche de sélection de Wason. Puis nous procéderons à un examen critique de ces travaux, en nous en tenant en général à des arguments déjà présentés par divers auteurs ; cependant, nous nous appuierons sur un point de vue personnel sur les rapports entre l'implication logique, la conséquence logique et le « si » du langage naturel pour achever de rejeter complètement la thèse de l'incompétence logique des sujets naïfs (§3). Nous passerons alors à l'examen des expériences de Kahneman et Tversky sur le jugement en situation d'incertitude, paradigme de l'étude empirique du raisonnement inductif et nous évoquerons successivement les explications de Kahneman et Tversky fondées sur la notion d'heuristique, puis les critiques, par Gigerenzer et par Cohen, de leurs travaux (§4). En posant alors aux théories psychologiques du jugement en situation d'incertitude les questions que nous avons posées aux théories du raisonnement déductif, nous mettons en évidence d'abord les insuffisances des premières, ensuite la douteuse pertinence de la théorie des probabilités pour l'explication du jugement sur l'incertain, enfin l'erreur qui consiste à mettre les deux sortes de phénomène sur le même plan, à l'image du couple philosophique déduction / induction : tout raisonnement fait intervenir, mais à des niveaux différents, inférences démonstratives et non démonstratives ; une perspective

¹ Même des auteurs proches à beaucoup d'égards de positions classiques militent en faveur d'un certain retour au contenu ; v. Barwise 1986, 1987 et tout récemment Peacocke (1994).

modulariste nous apparaît alors comme susceptible de fournir une bonne heuristique pour délimiter les aptitudes sous-tendant le raisonnement (§5). Nous concluons sur une réhabilitation de la logique classique qui laisse pourtant le champ libre à toutes sortes de processus non logiques.

1. ETUDE EXPERIMENTALE ET THEORIES DU RAISONNEMENT DEDUCTIF ELEMENTAIRE

1. 1 Le paradigme de Wason

Depuis qu'un psychologue anglais, Philip C. Wason, se proposa il y a près de trente ans de revenir sur la question des aptitudes humaines fondamentales en matière de déduction, les anciennes certitudes en la matière vacillent. Des milliers d'études expérimentales ont été rapportées, des centaines d'articles et de livres s'interrogent sur les « comportements déductifs » de l'homme et sur leur explication en tant qu'ils sont le produit de certaines capacités cognitives. Or cette question était classique, et, semble-t-il, solidement prise en main, sinon résolue complètement, par l'école piagétienne. Mais Piaget est le représentant par excellence de l'ancienne alliance : il ne met pas en doute l'idée que la logique classique constitue la structure de tout raisonnement déductif correct, et que la psychologie n'a rien à y redire. Ce qu'il veut expliquer, c'est comment l'enfant en vient à *se saisir* de la logique. Wason est plus empiriste : il veut s'assurer que l'adulte possède bien les aptitudes déductives élémentaires que la tradition lui attribue, et pour cela il doit s'éloigner d'une autre tradition, celle de la psychologie expérimentale dans son ensemble qui, à l'instar de Piaget, tient pour acquise l'existence d'une aptitude logique fondamentale et se demande seulement comment elle est mise en œuvre dans la résolution de problèmes simples. Autrement dit, Wason porte un premier coup à l'ancienne alliance en osant descendre d'un cran dans l'élémentaire.

Pour cela, il imagine un dispositif qui ne tardera pas à constituer un paradigme, appelé la « tâche de sélection ». L'expérience originale consiste à présenter aux sujets un jeu de cartes dont un côté porte une lettre et l'autre un chiffre, la seule contrainte étant que si une carte porte un A d'un côté, elle doit avoir un 4 de l'autre. Cette règle une fois expliquée aux sujets, on les place devant quatre cartes de ce jeu, posées à plat sur une table. Les symboles qui apparaissent sont A, D, 4, 7. On demande alors de déterminer par la pensée, et de dire, quelles sont les cartes qu'il est nécessaire de retourner pour s'assurer que le jeu obéit bien à la contrainte stipulée (si A d'un côté, 4 de l'autre). Wason pense déterminer ainsi le degré de maîtrise d'une aptitude logique fondamentale, comprendre l'implication logique $P \supset Q$ ². Or il constate que loin d'obtenir à une très solide majorité la réponse correcte (à savoir « Il faut retourner les cartes portant le A et le 7, et celles-là seulement »), il ne trouve qu'un taux faible — de l'ordre de 10 % — de bonnes réponses. L'expérience est exemplaire pour deux raisons : elle est robustement reproductible, donnant des résultats très voisins quelle que soit — dans certaines limites, nous y revenons dans un instant — la formulation exacte du test, et quels que soient les sujets, leur formation, leur âge, leur appartenance sociale, leur culture, etc. ; et elle donne lieu à une répartition constante et caractéristique des erreurs (les deux réponses erronées « Carte A » et « Cartes A et 4 » recueillant environ trois quarts des suffrages).

De nombreuses autres expériences ont été menées : vérification d'énoncés découlant déductivement de la description d'une situation ; construction ou vérification de tables de vérité ; jugement sur la validité d'arguments déductifs ; invitation à appliquer *modus ponens* ou *modus tollens*, ou encore le principe de transitivité, invitation à fournir la conclusion d'un syllogisme, etc. Il est remarquable cependant que, même si nul ne conteste l'importance qu'elles prennent dans leur ensemble, aucune d'elles n'ait attiré, en psychologie mais tout particulièrement en dehors d'elle, ne serait-ce qu'une fraction de l'attention réservée à la tâche de sélection. Les raisons de cette inégalité de traitement — abstraction faite d'effets sociologiques — sont intéressantes. La première est qu'aucun autre type d'expérience n'a paru longtemps autant que celle de Wason mesurer directement et, il faut y insister, d'une

² Il est vrai que cette expérience prend place dans une étude de la mise à l'épreuve d'hypothèses (*hypothesis testing*), ce qui n'est pas exactement la même chose que la possession d'un concept logique. Mais mettre à l'épreuve cette possession, au niveau d'une maîtrise complète, revient à vérifier si le sujet est capable de s'assurer que le concept est instancié.

manière aussi robuste, une capacité logique absolument centrale. La deuxième est que la tâche de sélection donne lieu à une distribution caractéristique de réponses qui intrigue et appelle explication ; les tâches exécutées avec un taux de réussite très élevé intéressent peu les chercheurs, dans la mesure où elles ne semblent que confirmer une intuition commune (à savoir que les hommes sont parfois capables de raisonner logiquement), et celles qui ne sont accomplies correctement que par une très faible proportion des sujets sont manifestement inadaptées (puisqu'elles entrent en conflit avec l'intuition précédente). La troisième est que rapidement ont été découvertes des variantes de la tâche de Wason pour lesquelles les performances étaient très différentes, et sensiblement meilleures, ce qui a permis d'appliquer la méthode des différences pour faire, à partir de l'analyse des facteurs modulant la performance, des hypothèses sur les capacités cognitives sous-tendant le raisonnement « réel ». Ce qui est instructif pour le non-spécialiste est que l'expérience de Wason soit apparue longtemps comme la seule à remplir ces conditions — est-il donc si difficile que cela de tester l'aptitude à la déduction simple ?

Il se trouve que — ironie de l'histoire — de nombreux auteurs contestent aujourd'hui que la tâche de sélection ait un rapport direct avec une capacité *logique* élémentaire quelle qu'elle soit, et certains doutent même qu'elle teste l'aptitude au *raisonnement* ; ils insistent *a contrario* sur certaines expériences moins prisées, et soupçonnent un biais quasi idéologique³ dans le privilège accordé à la tâche de sélection. On assiste également à une remise en question des facteurs auxquels a été attribuée une influence cruciale sur les performances et qui ont nourri la réflexion sur la nature même du raisonnement. Mais ce retournement, du reste fort récent et qui est loin d'être unanime, ne nous dispense nullement d'examiner le paradigme de Wason et tout particulièrement les facteurs qui modifient la performance. En effet, l'état présent de la question est très largement déterminé par le débat suscité par ce paradigme ; que la ou les présupposés qui lui ont donné vie connaissent un début de défaveur n'y change rien. D'ailleurs, bon nombre d'autres expériences sont encore interprétées comme allant dans le sens de ces conjectures — en particulier, elles semblent confirmer que, des quatre règles d'inférence simples auxquelles donne lieu l'implication, à savoir les deux règles valides, MP et MT et les deux règles invalides, AC (affirmation du conséquent) et NA (négation de l'antécédent), seule MP est traitée convenablement par une forte majorité des sujets. Bref, la tâche de Wason, même si elle est soupçonnée de particularités qui restreignent sa portée, demeure caractéristique de l'esprit des recherches actuelles.

Outre les médiocres performances obtenues sur la tâche originale, les résultats qui ont le plus marqué les spécialistes concernent l'effet bénéfique exercé sur les performances par certaines modifications de cette tâche. Ainsi, en lui donnant un certain type de contenu concret, ou en la situant dans un certain contexte, on obtient un spectaculaire renversement des proportions de bonnes et de mauvaises réponses. S'il s'agit par exemple de vérifier (le plus économiquement possible) que des règles telles que les suivantes sont respectées :

« Pour avoir le droit de boire de la bière, il faut avoir dix-huit ans ou plus »,

ou encore :

« Pour pouvoir être affranchi au tarif réduit, un pli ne doit pas être scellé »,

les sujets s'en sortent à plus de 60 %. De même, si l'on donne au conséquent une forme négative :

« Une carte ayant un A d'un côté ne porte pas de 4 de l'autre »,

on améliore les résultats. Ainsi ont été postulés différents facteurs facilitants : le contenu, tout d'abord (le fait que les lettres et les chiffres soient remplacés par des propositions chargées de

³ Déçus par ce que les premiers résultats expérimentaux (en matière de raisonnement tant déductif qu'inductif – v. §4 *infra*) semblait révéler sur les aptitudes véritables d'*Homo sapiens* en matière de raisonnement, et partant de rationalité, les psychologues auraient cultivé un style pessimiste et amer et donné la préférence aux expériences et interprétations qui les confortaient dans leur sombre vision. Voir par exemple Lopes 1991.

sens), puis plus particulièrement le caractère concret de la règle, la familiarité avec la situation ou avec les objets manipulés (« S'il pleut, je prends la voiture »), le contexte déontique (obligation/permission). On a aussi proposé l'existence d'une tendance, ou « biais », qui pousse à choisir la réponse qui reflète les données de l'énoncé [*matching bias*] : un problème de la forme « Si P, alors Q » susciterait ainsi la (mauvaise) réponse « Carte P et carte Q », alors que les problèmes de la forme « Si P, alors non-Q » susciteraient la *même* réponse, bonne cette fois, du moins si la négative apparaît de manière implicite. (La notion de biais joue un rôle central dans l'étude du raisonnement inductif et nous y reviendrons au § 4.)

En fait, ces effets ne sont pas uniformément observés : ils peuvent se compenser mutuellement, ils peuvent disparaître dans certaines conditions, etc. Nous ne rentrerons pas dans ces considérations pour trois raisons : elles nous entraîneraient trop loin dans une voie dont il y a maintenant lieu de croire qu'elle est assez particulière ; elles ne sont pas nécessaires pour comprendre ce qui a motivé les théories du raisonnement qui ont été proposées à la lumière de l'ensemble des expériences ; enfin, nous sommes tenté de penser que l'explication générale de la tâche de sélection vient d'être fournie par Dan Sperber, Francesco Cara et Vittorio Girotto dans un texte non encore publié⁴ ; or cette explication, nous le verrons, fait appel à des idées assez différentes de celles qui viennent d'être évoquées.

1. 2 Trois théories du premier ordre

A une théorie cognitive du raisonnement déductif, il revient d'expliquer (1) les compétences déductives de l'esprit humain, (2) les nombreuses erreurs qu'il commet, (3) la distribution de ces erreurs, qui n'est nullement uniforme ni, par exemple, le simple reflet de la complexité apparente de la tâche, et enfin (4) l'influence exercée par le contenu et par le contexte.

Par « théorie du premier ordre », nous désignerons ici une théorie cognitive du raisonnement déductif qui prétend rendre compte de cette capacité indépendamment de toute autre. Le terme, il faut l'avouer, est doublement trompeur, puisqu'il n'est pas habituel de l'appliquer, comme nous le ferons, à des théories du raisonnement, et que ces théories n'ont rien à voir avec ce qu'un logicien entend par là. Nous n'avons pas trouvé mieux pour distinguer ces théories de celles que nous appellerons « du second ordre » parce qu'elles font intervenir des considérations sur les conditions dans lesquelles s'exercent les capacités dont les théories du premier ordre veulent rendre compte, et sur le « matériau » sur lequel ces capacités opèrent. La distinction est du reste provisoire, elle est seulement utile pour se repérer dans le champ passablement chaotique (« fragmenté », comme le dit Jonathan Evans, l'un des spécialistes les plus lucides du raisonnement déductif) des recherches sur le raisonnement. A terme ne survivra, peut-on espérer, qu'une seule théorie qui intégrera les deux niveaux ou les deux sortes de mécanismes.

On distingue en général, pour les opposer, trois théories du premier ordre. (En fait, comme y a insisté Evans, elles ne sont pas rigoureusement comparables. Nous y reviendrons au § 3.)

a) La logique mentale

Cette théorie attribue à l'esprit humain la possession de règles logiques de nature syntaxique, accompagnées le cas échéant de règles pour l'application de ces premières règles. Dans les exposés historiques du domaine, on fait remonter généralement à Piaget, voire à Boole la première idée de ce genre. Aujourd'hui, elle est défendue particulièrement, et testée expérimentalement, par Lance Rips et par Martin Braine et ses collaborateurs ; John Macnamara la développe (sous le nom de « psychologique ») dans un contexte épistémologique très large où il tente de redéfinir les rapports entre logique et psychologie. Actuellement, seule la partie propositionnelle de la théorie est pleinement développée.

Braine, auquel nous nous référons de préférence, a proposé un système de règles de type déduction naturelle, accompagné de règles de contrôle. Selon lui, ce qui explique notre capacité à effectuer des déductions est l'application de ce système logique interne à des représentations mentales des propositions à traiter. La logique mentale qu'il postule est non classique et incomplète : elle se veut « réaliste », c'est un « organe mental » de même nature

⁴ L'article est paru depuis la rédaction de cet essai.

que la grammaire universelle postulée par Chomsky. Donc même s'il s'agit d'une entité théorique qu'aucun microscope (ou « cérébroscope ») ne permettra jamais de voir, elle doit être conforme à ce que le jugement des psychologues tire de l'expérimentation et de la réflexion théorique d'ensemble. L'incomplétude n'est pas nécessairement un défaut, puisque c'est d'une capacité réelle, et non idéale, qu'il s'agit de rendre compte. Bref, la logique mentale est bien *une* logique, mais n'a pas exactement la même fonction que *la* logique des logiciens. A ce propos, deux importantes remarques s'imposent ici.

La première est que les règles de la logique mentale sont certes syntaxiques, donc ne prennent pas en compte le sens des propositions, mais seulement leur morphologie. Cependant, selon Braine, elles ne fonctionnent normalement que lorsque les représentations internes sont munies d'une interprétation. La déduction sur des propositions ininterprétées n'est pas de la compétence de ce système. Ce système, c'est la seconde remarque, doit selon Braine être compris comme une ressource *primaire* dont sont équipés tous les esprits normalement constitués, indépendamment de leur niveau d'instruction et notamment de leur compétence logique ou mathématique. Nous reviendrons au § 3 sur certaines questions soulevées dans les deux dernières phrases, qui, loin d'aller de soi, résultent d'un effort de clarification qui s'est étendu sur une quinzaine d'années et n'est pas encore, loin s'en faut, compris et accepté par tous.

b) Les schémas pragmatiques

Motivée par l'observation des effets facilitateurs des contextes déontiques sur la tâche de sélection, cette théorie rapporte nos capacités déductives à la présence dans le système cognitif de « schémas pragmatiques » qui se déclenchent, à la manière des systèmes de production de l'intelligence artificielle, dans certaines conditions et produisent une conclusion déterminée. Ce sont des *schémas* en un sens reçu depuis longtemps en psychologie cognitive : généralisations acquises par apprentissage, et dont la forme permet l'application à un nombre indéfini de cas nouveaux par instantiation de variables libres. Et ils sont *pragmatiques* parce qu'ils prescrivent la démarche appropriée dans les contextes créés par l'action.

Proposée dans le cadre de l'interprétation de la tâche de Wason par Patricia Cheng et Kenneth Holyoak, la théorie se ramène en pratique à la formulation des schémas dits de *permission* et d'*obligation*. On pourrait imaginer de l'étendre à d'autres situations au moyen d'autres schémas, mais pour diverses raisons, dont les faiblesses épistémologiques de la théorie sous sa forme présente, cette piste n'a pas été poursuivie, du moins directement, par beaucoup de chercheurs⁵. Disons donc rapidement ce qu'est par exemple le schéma de permission. Il s'applique à toute situation obéissant à une règle conditionnelle de la forme « Pour qu'une certaine action puisse être accomplie, une certaine précondition doit être satisfaite » et se compose lui-même de quatre règles :

- (R1) Si l'action doit être accomplie, la précondition doit être satisfaite.
- (R2) Si l'action ne doit pas être accomplie, il n'est pas nécessaire que la précondition soit satisfaite.
- (R3) Si la précondition est satisfaite, l'action peut être accomplie.
- (R4) Si la précondition n'est pas satisfaite, l'action ne peut pas être accomplie.

Nous ne nous attarderons pas sur les obscurités de cette proposition. On voit bien que des clarifications sont nécessaires, ne serait-ce que pour éviter une régression à l'infini : les règles du schéma sont elles-mêmes semble-t-il des conditionnelles de la même nature que celle dont le schéma est censé dire comment elle est traitée. Contentons-nous de rapporter

⁵ On nous permettra à ce propos une remarque d'ordre sociologique : les psychologues sont une communauté plus nombreuse d'un ordre de grandeur au moins que celles des logiciens ou des philosophes, et ils produisent des publications à un rythme peut-être cent fois supérieur. Il est difficile dans ces conditions de se priver complètement, dans l'évaluation des différents programmes de recherche, de critères quantitatifs.

que l'on peut en donner une interprétation cohérente ; l'idée fondamentale étant que les règles du schéma ne sont pas des propositions, mais comme on l'a dit des règles d'action comparables aux règles de production au sens de Newell : elles sont immédiatement applicables ou non, selon les circonstances, et il n'y a rien à en déduire ou à en tirer lorsqu'elles ne le sont pas. Ce qui n'est pas le cas de la règle de départ, dont le schéma a précisément pour fonction de dicter les conséquences à tirer dans les différents cas de figure.

Insistons en revanche un instant sur ce qui sépare la théorie des schémas pragmatiques de celle de la logique mentale. D'abord elle postule non seulement que les mécanismes déductifs s'appliquent à des propositions interprétées (on a vu à l'instant que la logique mentale, en tout cas selon Braine, est dans le même cas), mais surtout qu'ils sont sensibles à l'interprétation ; en ce sens, la théorie de Cheng et Holyoak se veut non formelle, voire anti-formelle. En deuxième lieu, elle soutient que ces mécanismes sont activés par la reconnaissance de situations ayant une pertinence dans le contexte humain, plus particulièrement de situations à caractère social (permissions et obligations sont typiquement de nature collective). Enfin, elle suggère que ces mécanismes sont spécialisés : ils dépendent fortement du domaine auquel l'individu a affaire. Or ces trois idées exercent actuellement une très forte influence sur différentes écoles de pensée. La non-formalité inspire non seulement la théorie des modèles mentaux dont il va être question dans un instant, mais les travaux en sémantique et en philosophie du langage associés à la *situation theory* de Jon Barwise et John Perry, ainsi que tout ce qui se réclame, dans les sciences cognitives et l'intelligence artificielle, de la « cognition située ». La sensibilité au contexte, et l'idée qu'un contexte est créé par une situation « écologiquement valide », importante et familière pour l'organisme ou l'individu, autant de thèmes chers aux tenants du contextualisme, aux divers sens du terme, en particulier aux défenseurs de la « cognition sociale » ; nous en verrons un exemple au § 3, celui de la théorie du contrat social de Lena Cosmides qui prolonge en un sens celle des schémas pragmatiques. Quant à la dépendance par rapport au domaine, sous l'étiquette de *domain specificity*, elle inspire un courant de recherche d'apparition récente mais qui s'enfle à vue d'œil (v. Hirschfeld & Gelman 1994).

c) *Les modèles mentaux*

Le psychologue anglais Philip Johnson-Laird, élève de Wason et d'abord rallié, comme à l'époque la plupart des spécialistes du raisonnement, à l'idée de logique mentale, a proposé au début des années 1980, pour expliquer le raisonnement syllogistique « réel », une nouvelle théorie, dite des « modèles mentaux », qui est désormais très largement majoritaire parmi les psychologues. Abondamment remaniée, raffinée, généralisée au fil des années, elle est donnée aujourd'hui par ses défenseurs comme la seule théorie qui soit à la fois conceptuellement complète et cohérente et capable de rendre compte de la plupart des résultats expérimentaux relatifs au domaine pris dans sa plus grande extension (qui comprend non seulement les raisonnements syllogistiques, pour lesquels la théorie a été d'abord proposée, mais tous les raisonnements relevant de la logique déductive, le raisonnement spatial, les raisonnements métallogiques impliqués dans les énigmes de type menteur, et même peut-être le raisonnement inductif au sens de notre § 4 *infra*). Nous exprimerons au § 3 quelques raisons de douter de ses qualités conceptuelles, mais comme dans le cas des schémas pragmatiques le plus important est d'abord d'en exposer les grandes lignes et surtout d'en saisir les motivations profondes.

Selon Johnson-Laird, le raisonnement ne fait pas intervenir de règles « logiques » : il ne consiste pas à appliquer des règles d'inférence formelles comme celles que postule la théorie de la logique mentale. Raisonner consiste au contraire à construire un modèle mental de la réalité décrite par l'énoncé du problème considéré et à « lire » sur ce modèle des informations nouvelles et utiles concernant la réalité en question. La construction du modèle procède elle-même en plusieurs étapes. Un premier modèle est construit à partir de la première indication de l'énoncé ; ce modèle est en général encore implicite à beaucoup d'égards : il reste « muet » sur bien des points ; ce sont les indications suivantes de l'énoncé (ou de son interprétation par le sujet) qui permettent de l'explicitier graduellement. Comme le processus consiste à combiner des données, il peut conduire à des incohérences. S'ouvre alors, du moins dans certains cas, une phase de vérification de la cohérence. Si une contradiction se révèle, une deuxième tentative est faite pour obtenir un modèle cohérent. D'ailleurs, au cours de la phase

de construction, il a pu arriver que plusieurs possibilités apparaissent, donnant naissance à plusieurs modèles plus ou moins développés. Dans tous les cas, il vient un moment où le raisonneur a construit un ou plusieurs modèles dans lesquels il n'a pas détecté d'incohérence. La phase finale peut alors commencer. Elle consiste, s'il s'agit d'une épreuve de vérification de la validité d'une conclusion C, à « lire » sur le ou les modèles la réponse à la question « C est-elle vraie dans le modèle ? » ; et s'il s'agit d'une épreuve ouverte, dans laquelle on demande par exemple ce qui découle des prémisses, le sujet cherchera à deviner ce qui est à la fois non trivial, non explicitement contenu dans les prémisses et vrai dans le ou les modèles construits.

Les tenants de cette théorie (évoquée ici de manière abstraite et générale et qu'il faudrait naturellement préciser et illustrer) lui voient deux avantages décisifs : d'une part elle rend raison de leur indéfectible conviction que le raisonnement naturel est un processus essentiellement sémantique ; d'autre part elle rend compte avec fidélité, selon eux, des différences de difficulté : les taux comparés d'erreurs reflètent le nombre de modèles dont la construction est nécessaire pour résoudre le problème.

Pour l'observateur critique en revanche, plusieurs questions restent entières, même après l'étude des exposés détaillés de la théorie : (1) En quel sens peut-on considérer que les opérations postulées sont élémentaires ? (2) En quel sens ces opérations sont-elles non logiques, sémantiques et informelles plutôt que logiques, syntaxiques et formelles ? (3) Quelle est la plausibilité de la théorie ? (4) Pourquoi faudrait-il qu'une théorie conçue pour rendre compte de la capacité à résoudre une famille de problèmes très artificiels (la construction ou la vérification de conclusions de syllogismes), quels que soient ses mérites en la matière, donne naissance par généralisation à une théorie du raisonnement déductif ?

2. THEORIES DU DEUXIEME ORDRE

Un bon nombre d'interrogations d'ordre conceptuel ou théorique ont été soulevées au passage. Pour commencer à y répondre, nous donnerons d'abord la parole à certains spécialistes du domaine qui, conscients des difficultés, ont voulu y parer en élaborant ce que nous avons choisi d'appeler, bien imparfaitement, des « théories du deuxième ordre ». Ces théories sont d'ambition inégale et ne s'adressent pas toujours aux mêmes questions ; elles peuvent se compléter mutuellement ou se combiner à certaines variantes de certaines théories du premier ordre. Pour ne constituer qu'un échantillon, celles que nous allons évoquer sont représentatives, pensons-nous, des choix théoriques disponibles.

2. 1 Théories de la rationalité limitée ; compétence et performance

Deux voies s'ouvrent de prime abord devant celui qui rejette l'idée que les écarts observés entre la norme logique et les performances des sujets sont l'indice d'une déficience essentielle. La première est de rapporter ces écarts à une nécessité purement matérielle, celle qu'impose la limitation des ressources du système cognitif humain. Rien de déshonorant, plaide-t-il, à ne pouvoir satisfaire à des exigences illimitées, particulièrement lorsque les tâches considérées présentent une forte complexité algorithmique, ce qui est le cas, déjà, de la logique propositionnelle. Au contraire, les ruses déployées pour compenser ces limitations sont la marque de l'intelligence. Naturellement, il y a un prix à payer : il est précisément dans le risque d'erreur. Inversement, poursuit le tenant de la rationalité « limitée » ou « minimale », il y a de la part du théoricien une manière de stupidité à vouloir élaborer un modèle du raisonnement humain sans partir de cette donnée fondamentale qu'est la finitude : quel ingénieur ferait-il abstraction de la résistance de l'acier en concevant un avion, quel général oublierait-il dans son plan de bataille que ses hommes se déplacent à une vitesse finie, ont besoin de nourriture et de repos, etc. ? Herbert Simon, qui est, on le sait, à la fois économiste et co-fondateur de l'intelligence artificielle et de la psychologie cognitives, a le premier fait valoir l'argument à propos de l'agent économique rationnel de la théorie économique classique. Le philosophe Christopher Cherniak, entre autres, a développé cette argumentation à propos de la cognition en général ; maints psychologues sont aujourd'hui

tentés de l'appliquer spécifiquement au cas du raisonnement déductif⁶ — il leur semble évident que la logique classique ne peut tout simplement pas fournir un modèle même approché des capacités humaines en la matière, puisqu'elle exige de supposer des ressources calculatoires dont nous savons que l'esprit humain n'en dispose pas.

L'argument ainsi présenté repose, semble-t-il, sur une confusion que nous dissiperons un peu plus bas. Tournons-nous alors vers la seconde possibilité.

Elle consiste à traiter la capacité de raisonnement à la manière dont Chomsky traite la capacité de langage : il y a un écart inéliminable entre le comportement linguistique observable d'un individu et le comportement théorique qui résulterait du déploiement de la base cognitive que le linguiste est fondé à postuler pour rendre compte de ce comportement. Pour le dire plus rapidement, la compétence du possesseur d'une langue ne détermine pas sa performance. Le philosophe Jonathan Cohen a le premier soutenu que les expériences de la psychologie du raisonnement ne prouvent, dans la mesure où elles sont significatives, rien de plus que l'existence d'un écart comparable à celui que décrit Chomsky ; et il tente de justifier cet écart par des considérations théoriques analogues. Macnamara, et d'autres psychologues qui défendent la théorie de la logique mentale contre l'objection des erreurs systématiques, adoptent la même stratégie. L'objection qu'ils doivent prévenir, tout comme Chomsky a dû le faire dans le cas du langage, est que cette stratégie risque d'être « immunisante » au sens de Popper : sans précautions théoriques, elle risque de paraître interdire toute invalidation empirique de la théorie de la compétence.

2. 2 Evolution et contrat social

On a évoqué plus haut les relativement bons résultats obtenus dans certaines situations — il s'agissait, dans la tâche de sélection de Wason, de celles qui impliquent des dispositions telles que l'âge légal pour la consommation d'alcool ou le droit au tarif réduit pour les plis postaux non cachetés. Pour expliquer ces phénomènes, Lena Cosmides a récemment proposé une théorie sociodarwinienne reposant sur trois idées principales : (i) pour l'animal social qu'est l'homme, les situations créées par l'existence d'un contrat social sont particulièrement importantes ; (ii) l'aptitude à gérer ces situations résulte d'un processus de sélection naturelle ; (iii) une telle aptitude repose en partie sur la capacité de faire respecter un contrat, et cette capacité repose à son tour sur l'aptitude à détecter les tricheurs. Sachant qu'un « contrat » a pour Cosmides la forme « Si l'on jouit d'un bien, il faut en payer le prix », le « tricheur » est pour elle celui qui jouit du bien sans en acquitter le prix. L'« algorithme de détection des tricheurs » dont nous serions équipés par la grâce de l'évolution permet donc, selon Cosmides, de repérer le joueur non payeur, c'est-à-dire de résoudre correctement la tâche de sélection correspondant aux situations de « contrat social ».

Laisant les spécialistes discuter de la validité empirique de la théorie, de ses rapports avec la théorie des schémas pragmatiques ou d'autres théories, de l'extension exacte de son domaine d'application, attardons-nous sur son originalité métathéorique. Voilà en effet une théorie qui premièrement, contrairement aux théories du premier ordre dont il a été question, n'a rien à dire sur la manière dont la tâche est accomplie *en général*, mais prétend expliquer *pourquoi* elle est exécutée correctement dans une famille spécifiée de cas et *comment* elle l'est ; et qui deuxièmement situe cette double explication en dehors du domaine logico-calculatoire : le *pourquoi* relève de la sociobiologie, le *comment* de la reconnaissance des formes. Bien entendu, le soubassement computationnel n'est pas nié : c'est bien un *algorithme* qui accomplit — qui explique que s'accomplisse — la tâche de détection du tricheur, c'est, on peut le présumer, un algorithme qui a préalablement permis d'identifier la situation contractuelle ; et c'est la présence effective, « implémentée » dans le système cognitif, de ces algorithmes dont rend compte l'explication darwinienne. Mais la fonction de l'algorithme n'est que de garantir la matérialité du processus : il n'y aurait pas un mot à changer dans la théorie si l'on découvrait demain que le rôle actuellement attribué aux algorithmes est en réalité tenu par le Saint Esprit⁷.

⁶ Voir par exemple Oaksford & Chater (1992).

⁷ Je défends l'idée que les sciences cognitives jouissent d'une certaine autonomie à l'égard de l'hypothèse fonctionnaliste ou « computo-représentationnelle » (voir par exemple l'essai « Les sciences cognitives entre cerveau et machine », ce volume). On en trouve ici une illustration.

La théorie de Cosmides n'élimine pas seulement de l'explication d'une tâche de raisonnement la logique : elle élimine le raisonnement lui-même. Pour le dire de manière non paradoxale, elle est le premier exemple que nous rencontrons d'explication éliminative dans le domaine du raisonnement : ce qui *apparaît* comme un raisonnement élémentaire est *en réalité* autre chose. Voici du même coup posée la question de savoir si tout raisonnement élémentaire ne risque pas de subir un sort analogue.

2. 3 La théorie heuristico-analytique

Evans a le premier développé dans le présent contexte une idée essentielle : toute tâche de raisonnement se déroule en deux temps. Le sujet commence par sélectionner les aspects de la situation qui lui semblent mériter son attention ; cette première phase est dite *heuristique* parce qu'elle repose sur des mécanismes non logiques, préconscients, qui sont du même ordre que l'attention sélective ou qu'une forme de perception intellectuelle, et non de l'ordre de la délibération. Ayant ainsi formé sa représentation personnelle du problème, le sujet s'attache alors à le résoudre ; cette seconde phase est dite *analytique* parce qu'elle est de l'ordre du raisonnement logique : elle consiste à traiter, au niveau conscient et de manière cohérente, l'intégralité des éléments de la représentation obtenue à l'issue de la phase heuristique. Evans insiste sur le fait qu'il ne présuppose aucune théorie particulière de la phase analytique : sa conception biphasique lui paraît compatible avec chacune des théories du premier ordre actuellement disponibles, comme avec diverses autres théories du deuxième ordre (par exemple celles qui veulent prendre en compte la complexité algorithmique de la tâche).

Le lecteur au fait de l'évolution de l'intelligence artificielle ou de celle de la linguistique s'étonnera peut-être que l'idée d'Evans n'aille pas de soi, ou du moins qu'il ait fallu longtemps pour qu'elle soit prise au sérieux. L'IA n'a-t-elle pas depuis longtemps mis au premier rang de ses préoccupations la « représentation des connaissances » comme facteur décisif dans l'exécution des tâches par l'application d'algorithmes ? Et la linguistique ne sait-elle pas au moins depuis Grice que la saisie d'un message met en jeu des processus pragmatiques qui interagissent avec les processus traditionnels du lexique, de la syntaxe, de la sémantique ? Or ce qu'Evans nomme heuristique, et qui conditionne le déploiement de l'analytique, n'est-il pas dans une large mesure l'homologue de la représentation des connaissances en IA et de la pragmatique en linguistique ? Peut-être les psychologues du raisonnement n'ont-ils tout d'abord pas songé qu'un tel facteur puisse entrer en jeu dans des tâches *élémentaires* : le passage de la présentation du problème qui est faite au sujet à la représentation qu'il en fait et sur laquelle il fait porter ses compétences proprement raisonnantes était vu sans doute comme une manière de copie. Si c'est le cas, ce n'est pas le seul piège que nous aura tendu l'idée d'élémentarité dans le domaine du raisonnement.

Quoi qu'il en soit, la conception d'Evans modifie considérablement la problématique. En premier lieu, elle jette un pont entre le raisonnement déductif et l'autre principale branche de la psychologie du raisonnement, appelée généralement raisonnement en situation d'incertitude et liée, comme nous le verrons au § 4, au programme dit « heuristiques et biais » de Kahnemann et Tversky. L'heuristique d'Evans introduit — avec d'autant plus de force qu'elle est selon lui « préconsciente » — un élément d'alogicité, donc potentiellement d'irrationalité, dans le raisonnement déductif même (apparemment) élémentaire, et permet d'expliquer les erreurs systématiques de la même manière que dans le raisonnement incertain, c'est-à-dire comme des « biais » : par l'application inappropriée de stratégies de sélection des aspects pertinents du problème.

En deuxième lieu, la question de la nature de l'heuristique est posée : comment opère la sélection des aspects pertinents, et peut-on donner une définition de la pertinence qui ne soit ni triviale ni dogmatique, c'est-à-dire qui n'identifie la pertinence ni à ce en vertu de quoi une information est sélectionnée par les processus heuristiques, ni à ce en vertu de quoi la « bonne » réponse est obtenue ? Evans et d'autres ont proposé des réponses partielles à ces questions, mais on peut penser que les travaux de Sperber, Cara & Girotto rendent ces réponses largement caduques.

En troisième lieu, il revient à Evans le mérite supplémentaire d'avoir rendu possible de voir dans la tâche de sélection, longtemps considérée comme représentative du raisonnement déductif élémentaire, une exception ou un cas-limite : Evans conjecture en effet que cette tâche ne met en jeu que la phase heuristique. Il opère ainsi dans ce cas, comme Cosmides mais d'une autre manière, plus générale et peut-être mieux motivée, l'élimination du raisonnement d'une tâche de raisonnement.

2. 4 L'approche pertinente de la tâche de sélection

La phase heuristique d'Evans consiste, on l'a vu, en la formation par le sujet d'une représentation interne du problème. Dan Sperber, Francesco Cara et Vittorio Girotto ont

appliqué la théorie de la pertinence de Sperber et Wilson à la description de ce processus dans le cas de la tâche de sélection. Cette théorie, on le sait, vise à rendre compte de la communication, plus précisément de la compréhension contextuelle des messages communiqués (principalement, mais non exclusivement, au moyen du langage). Trois objectifs sont atteints dans le travail de Sperber, Cara & Girotto : (i) Expliquer à partir de principes simples, clairement formulés et de portée générale l'ensemble proliférant des résultats empiriques obtenus dans la tâche de sélection, et réanalyser de manière lumineuse un grand nombre des explications « de second ordre » avancées jusqu'à présent. (ii) Prédire tout un ensemble de faits nouveaux, incompatibles avec les théories concurrentes, et les tester expérimentalement avec succès. (iii) Analyser le paradigme de Wason et la masse de ses variantes sur les plans logique, linguistique et cognitif avec une rigueur exemplaire, ce qui plaide, avec les succès expérimentaux mentionnés à l'instant, en faveur d'une extension de cette approche à d'autres problèmes que la tâche de sélection dont les auteurs soulignent, en accord avec Evans, le caractère très particulier.

L'idée qui préside à l'explication pertinentielle (au sens de Sperber et Wilson) de cette tâche est la suivante. La résolution du problème, comme l'a compris Evans, passe par la compréhension de la situation à l'aide de laquelle le problème est communiqué au sujet. Or cette situation est hautement complexe : elle comporte des aspects linguistiques, écrits ou oraux, mais aussi iconiques — par exemple les cartes, réelles ou représentées ; elle a surtout des composantes métalinguistiques (la structure du message complet — introduction, présentation des cartes, formulation de la question) et métathéoriques (cadre général d'une expérience de psychologie cognitive, rapports entre l'expérimentateur, omnipotent et omniscient, et le sujet, présumé apte à l'erreur, horizons culturels et épistémiques activés, ou non, au cours de la présentation de la tâche, stratégies savantes sollicitées ou non, stratégies spontanées encouragées ou découragées, etc.). L'interprétation par le sujet de cette situation, en particulier de ce que l'on attend de lui, est donc rien moins que triviale. Elle met en œuvre le « principe de pertinence » de Sperber et Wilson, principe qui comporte deux volets : [α] le sujet fait porter son attention sur les aspects de la situation qui sont pertinents, c'est-à-dire susceptibles de maximiser ses gains informationnels pour un coût de traitement minimal, et [β] le sujet présume que l'information communiquée par l'expérimentateur est pertinente.

Nous ne rapporterons pas la manière dont cette idée est déployée dans l'analyse détaillée de la tâche de sélection, ni comment elle permet la reprise critique des explications antérieurement avancées. Disons seulement que selon Sperber, Cara & Girotto, les sujets infèrent de la règle, par ordre décroissant d'accessibilité, certaines de ses conséquences observables, et s'arrêtent lorsqu'ils jugent que les conséquences considérées suffisent à donner à la règle la pertinence qu'ils en attendent. Cela fait, ils choisissent les cartes susceptibles de mettre à l'épreuve ces conséquences. Comme l'accessibilité et la pertinence varient selon les versions, il n'est pas étonnant que les résultats puissent varier de manière spectaculaire de l'une à l'autre. Il faut en revanche insister sur la prédiction très générale à laquelle cette analyse conduit : toute expérience dans laquelle la conséquence « Il n'y a pas de cas de P- & non Q » est aussi accessible, et plus informative, que la conséquence « Il y a des cas de P- & Q » conduira les sujets à choisir les cartes P et non Q. Cette prédiction prend la forme d'une « recette » pour fabriquer une version de la tâche de sélection sur laquelle on obtient d'excellents résultats :

- « — Choisir un couple de traits P et Q tels que le trait complexe P- & non Q soit (ou puisse être rendu) plus facile à représenter que le trait complexe P- & Q.
- Créer un contexte dans lequel le fait de savoir qu'il existe des cas de P- & non Q aurait de plus grands effets cognitifs que le fait de savoir qu'il existe des cas de P- & Q.
- Présenter la règle « Si P, alors Q » sous une forme pragmatiquement adaptée⁸. »

La prédiction est corroborée par cinq séries d'expériences sur des versions inédites de la tâche de sélection. Elle permet également de réanalyser les explications de second ordre qui ont été avancées, soit pour les faire apparaître comme des cas particuliers, soit pour les rejeter : la théorie pertinentielle rendrait compte de la manière la plus générale de la phase

⁸ Traduit de l'article Sperber, Cara & Girotto (1995), fig. 6. L'expression « pragmatiquement adaptée » [*pragmatically felicitous*] renvoie à un ensemble de conditions portant sur les propriétés pragmatiques de l'énonciation linguistique en contexte de la règle.

« heuristique », au sens d'Evans. En revanche, comme y insistent les auteurs, elle ne permet pas — du moins à elle seule — de départager les différentes théories du premier ordre, puisque la phase « analytique », comme chez Evans, est selon la présente théorie réduite à néant. Ce n'est pas que le raisonnement ait entièrement disparu dans l'exécution de la tâche de sélection : il entre en jeu dans la mise au jour de conséquences (logiques) testables de la règle, et dans le processus de compréhension lui-même. Mais ce n'est pas la tâche de sélection, selon Sperber, Cara & Girotto, qui permet de trancher entre les différentes théories du raisonnement. La principale raison qu'ils avancent est que dans de nombreuses versions de la tâche, l'énoncé de la règle est pragmatiquement déviant : les mécanismes inférentiels à l'œuvre dans la compréhension sont alors pris en défaut et conduisent à des résultats sans grande portée (de même qu'il n'y a pas grand chose à tirer, quant au fonctionnement d'un moteur à explosion, de la constatation que l'ajout de sucre dans le réservoir à essence d'une voiture provoque une panne à retardement).

3. QUESTIONS THEORIQUES

La situation dans laquelle nous nous trouvons au terme de ce parcours a quelque chose de singulier. Le principal effort des psychologues pour isoler une aptitude logique fondamentale, ou encore une capacité élémentaire de raisonnement déductif, semble avoir échoué. Les observateurs — philosophes, logiciens, informaticiens — pourraient donc estimer légitime de s'en tenir à leurs conceptions et préoccupations traditionnelles. Deux choses devraient les en dissuader : d'une part les psychologues poursuivent leur enquête empirique et les mettent au défi de juger, sur le plan théorique, du sens et de la cohérence de leurs conceptions du raisonnement ; d'autre part tout ce remue-ménage a obligé certains observateurs⁹ à reconsidérer et à reformuler des questions aussi anciennes que fondamentales : d'avoir labouré durement sans avoir — sans doute — exhumé le trésor espéré n'en a pas moins conduit les psychologues à modifier durablement le paysage conceptuel que nous contemplons tous. A ces deux raisons on ajoutera peut-être, à l'intention de ceux qu'elles laisseraient indifférents, la considération suivante : admettons que nous soyons justifiés dans notre tranquille assurance concernant la déduction ; sommes-nous aussi sereins pour ce qui est du ou des raisonnements non déductifs, du ou des logiques non classiques ? et sommes-nous absolument certains que nos inquiétudes à cet endroit n'aient aucun retentissement sur nos conceptions de la déduction ? A tout le moins devons-nous admettre que l'étude (expérimentale ou théorique) de l'induction (nous prendrons ce terme en un sens très lâche, pour désigner de manière concise toute démarche, réelle ou idéale, faisant intervenir des inférences non démonstratives) prend appui sur nos théories de la déduction ; et plus celles-ci seront explicites et exhaustives, embrassant tout particulièrement les résultats et hypothèses des psychologues, mieux nous saurons mener notre enquête sur l'induction.

Quoi qu'il en soit, dès lors que nous acceptons d'examiner les travaux des psychologues et le débat qu'ils suscitent, nous ne pouvons éviter de poser à nouveaux frais tout un ensemble de questions épistémologiques des plus difficiles. Nous devons nous contenter ici de les effleurer, remettant à une occasion future leur examen approfondi. Elles peuvent être regroupées sous trois intitulés : logique et raisonnement ; bases cognitives du raisonnement ; logique, langage et pensée.

3. 1 Logique et raisonnement

Comme on pouvait sans doute s'y attendre, la logique classique sort indemne de l'épreuve : aucun argument n'est avancé en faveur de son remplacement par un autre système. Ce qui est en revanche fortement mis en cause est son rôle dans la conduite du raisonnement. Pour le dire autrement, il n'existe pas d'autre candidat que la logique classique à la dignité de théorie fondamentale universelle du raisonnement, mais les mérites de l'unique candidat sont sérieusement contestés.

⁹ Je pense notamment aux philosophes Jonathan Cohen, Alvin Goldman, Gilbert Harman ou Stephen Stich.

En réalité, la déstabilisation de la logique entraîne fatalement celle du raisonnement : à partir du moment où le raisonnement n'est plus de la logique appliquée, diluée, contaminée par les imperfections du réel, qu'est-il donc ? Les réponses de Descartes et d'Arnault ne sont plus disponibles, pas plus que leur manière de faire coïncider, à la limite, description et prescription.

Néanmoins, dans la discussion présente, les participants ne mettent pas réellement en cause la notion de raisonnement : ils admettent qu'il existe, dans les cas simples de déduction qui sont ceux qui les intéressent, une notion claire et univoque de ce qu'est le bon raisonnement, la marche qu'il faut suivre ou en tout cas la conclusion correcte. Ils s'opposent en revanche sur le sens et les raisons de l'écart entre les réponses correctes et celles que fournissent les sujets d'expérience dans la résolution de problèmes ultra-simples de raisonnement déductif.

Harman, on l'a rappelé dans l'introduction, se place sur le terrain de l'analyse théorique. Le raisonnement, dit-il, n'est tout simplement pas l'application ou le reflet de la logique. Il va plus loin : pour lui la logique n'a pas de rapport privilégié avec le raisonnement — elle autorise des inférences inutiles et n'a pas les moyens d'imposer les inférences utiles ; pire, elle autorise des inférences que la raison peut préférer rejeter plutôt que d'accepter une conclusion inadmissible pour elle. Pour ces raisons, Harman n'est pas loin de voir dans la locution « raisonnement déductif » un oxymore¹⁰. En revanche, il considère avec sympathie l'hypothèse d'une « logique de base » qui serait la capacité psychologique en vertu de laquelle l'esprit manipule des pensées composées à l'aide des connecteurs logiques ; cette capacité reposerait sur la saisie des rapports d'implication immédiate et de contradiction immédiate. Il doute cependant que le genre de système de déduction naturelle qui peut être édifié sur cette base soit véritablement de nature logique au sens classique du terme : il n'est nullement évident, selon lui, que tout ce qui est immédiat en ce sens soit de nature logique au sens propre (il y a des immédiatetés spatiales, temporelles, etc. dont on ne sait si elles sont médiées par des axiomatiques spécifiques, ni même, estime-t-il, s'il est possible d'en décider).

Ce rappel des thèses de Harman permet de mesurer la distance qui sépare son point de vue de logicien-philosophe de celui d'un psychologue antilogiciste tel que Johnson-Laird. Les raisons que ce dernier avance pour contester à la logique un rôle, exclusif ou même partiel, dans le raisonnement, sont, on l'a vu, de nature empirique. Les réponses que donnent les êtres humains, assure-t-il, s'écartent très souvent des résultats dictés par la logique pour des raisons qui ne sont pas triviales (ce ne sont pas de simples erreurs d'« inattention »). Plus précisément, il rejette la solution théorique qui consiste à postuler d'abord un système logique de base, puis des mécanismes de performance qui viennent perturber les productions de ce système. Il estime avoir d'autant plus de raisons de rejeter une telle solution qu'il pense avoir identifié des processus non logiques qui permettent de rendre compte du raisonnement « réel », ce qui à ses yeux prive la logique du dernier argument qu'elle pouvait évoquer, celui d'être le seul candidat en la matière.

Prises côte à côte, ou face à face, les deux positions que nous venons de caractériser nous mettent dans l'embarras. D'un côté, une analyse théorique solide mais dont les conséquences sur le plan psychologique restent entièrement dans l'ombre. De l'autre, une proposition apparemment assez claire sur le plan de la psychologie expérimentale, mais dont le statut théorique est loin d'être assuré. Elle semble en effet entrer en conflit direct avec deux évidences.

La première est que les processus « non logiques » postulés par le psychologue antilogiciste mettent en œuvre des aptitudes logiques complexes, plus complexes que la logique élémentaire, et qui semblent reposer sur elle. C'est criant dans le cas de la théorie des modèles mentaux, dont la plus sommaire description regorge littéralement de termes empruntés à la logique et qui désignent des sous-processus de nature au moins partiellement logique. La seconde est que l'être humain a bel et bien recours à la logique pour raisonner déductivement (ou, si l'on préfère avec Harman éviter l'expression, d'argumenter déductivement, c'est-à-dire de résoudre des problèmes nécessitant la recherche ou la vérification de preuves déductives) : c'est à quoi les mathématiciens, entre autres, consacrent une bonne partie de leur temps.

¹⁰ « Obviously, there is deductive argument, but it is not similarly obvious that there is deductive reasoning. » Harman (1983), p. 6.

Ces deux conflits obligent à poser une série de questions, qui toutes visent à mieux délimiter ce sur quoi, sous le nom de « raisonnement », sont censées porter les explications en présence.

a) Première question : de quel sujet parlons-nous ?

La référence au mathématicien pourrait être considérée comme sophistique : n'est-il pas un expert en matière de logique, et ne pratique-t-il pas une forme de raisonnement tout à fait particulière ? Pour ces deux raisons, ses aptitudes ne se placent-elles pas nettement en dehors de la sphère examinée ?

Une première réponse consiste à contester les attendus de la question. D'abord il n'est juste de considérer le mathématicien comme un expert logique qu'au sens où on peut lui attribuer la logique infuse, c'est-à-dire une connaissance implicite de la logique. En effet, il ignore en général les règles d'inférence de la logique formelle ; néanmoins on peut plaider qu'il pourrait sans mal les retrouver, en tout cas qu'il en reconnaîtrait immédiatement la légitimité si on les lui présentait ; on peut dire aussi que la logique formelle contemporaine (pas celle d'Aristote) est avant toute chose une théorie du comportement déductif du mathématicien. Pour autant l'antilogiciste ne prétendra pas que le mathématicien n'emploie pas les ressources de la logique dans ses raisonnements (typiquement il procédera par *modus ponens*, *modus tollens*, *reduction ad absurdum*, démonstration par cas — élimination de la disjonction —, pose puis congédiement d'hypothèses, etc. : bref, il déploiera une série de procédés dont la déduction naturelle, au sens de la logique mathématique, a justement pour mission de fournir le modèle). Par conséquent l'antilogiciste doit exclure le mathématicien, non parce que celui-ci ne fait pas référence aux règles logiques, mais du seul fait que son comportement déductif est conforme à ces règles et que ces règles en rendent compte exhaustivement. Bien entendu, le mathématicien est également membre de l'espèce *homo sapiens*, donc relève des schémas explicatifs généraux relatifs aux capacités cognitives de l'espèce. Mais *en tant que* mathématicien, c'est-à-dire en tant qu'il se soumet, au prix d'un entraînement très particulier, à des règles de comportement déductif non naturelles, il est une exception qu'il est légitime pour le psychologue d'écarter.

Cependant, le mathématicien est loin d'être le seul auquel il arrive d'appliquer des règles d'inférence logique. Il est caractéristique, au contraire, des comportements qualifiés de rationnels au sens le plus strict, d'être gouvernés par la logique, au sens non d'une référence explicite, mais d'un respect manifeste : le vigneron qui sait que pour qu'il devienne riche, ses vignes doivent être atteintes de pourriture noble et qui, constatant qu'elles ne le sont pas, en conclut qu'il restera pauvre, respecte *modus tollens* sans connaître le nom du principe qu'il respecte. La différence avec le mathématicien est qu'il n'a pas fait l'objet d'un entraînement susceptible de modifier ses aptitudes spontanées, et que, sous peine de frapper d'invalidité *a priori* la discussion dans son ensemble, on ne peut supposer que la logique rend compte « par construction » de son comportement. Le vigneron ne peut donc être écarté de la même manière que le mathématicien dans l'exercice de ses fonctions.

Nous devons donc envisager de restreindre la discussion à certaines aptitudes d'un sujet que nous appellerons *logiquement naïf*. Le sujet logiquement naïf (SLN) manifeste sa qualité par son incapacité à aborder un problème de nature déductive comme le ferait l'étudiant au cours de logique. Il doit cette qualité soit à une ignorance, implicite et pas seulement explicite, des règles de la logique, soit à une impossibilité provisoire de mettre en œuvre ses éventuelles compétences en la matière. Encore faut-il comprendre son « ignorance implicite » de la logique en sorte de ne pas préjuger de la question de savoir si son aptitude est ou non « fondée » sur la logique.

Il est essentiel d'observer que, pour logiquement *naïf* qu'il soit, notre sujet possède au moins une *certaine* compréhension des particules de la langue correspondant aux constantes logiques (nous serons amené à y revenir), et une *certaine* maîtrise de la notion de conséquence logique — faute desquelles ledit sujet ne serait pas à même de *comprendre* le problème posé. Il possède ainsi une aptitude qu'on pourrait être tenté de caractériser à l'aide du mot « logique », convenablement qualifié. Mais pour éviter de laisser croire que l'argument en faveur de cette aptitude rend oiseuses les questions que se posent les psychologues, ou y voir au contraire une pétition de principe, appelons-la « aptitude déductive primaire » pour l'opposer aux aptitudes secondaires qui résultent vraisemblablement de processus

d'apprentissage culturels et individuels longs et complexes et sont fondées sur les capacités déductives primaires ainsi que sur d'autres ressources cognitives telle l'utilisation d'artefacts (écriture, etc.)¹¹.

b) Deuxième question : quelle aptitude s'agit-il d'expliquer ?

L'objectif des psychologues qui s'inscrivent dans la tradition de Wason est de caractériser la capacité au raisonnement déductif du SLN, et leur méthode repose sur l'idée que les raisonnements étudiés en psychologie cognitive dans la tradition du *human problem solving* de Newell et Simon sont trop complexes pour permettre d'accéder aux bases de cette capacité. Ce qu'il faut d'abord comprendre, c'est le raisonnement élémentaire. Mais qu'entend-on par là ? On a vu combien la tâche de sélection en est éloignée, contrairement aux espoirs de Wason. On pourrait examiner d'autres paradigmes — raisonnement conditionnel, complétion de tables de vérité, raisonnement disjonctif, syllogismes, etc. Dispensons-nous-en et allons droit au but : les seuls candidats plausibles ne sont-ils pas les règles de la déduction naturelle ? Les tâches élémentaires sur lesquelles tester les sujets sont donc de la forme : « Vous savez que P implique Q. Maintenant, vous apprenez que P. Qu'en déduisez-vous ? », ou encore « Vous savez que P et Q. Pensez-vous que P ? »

Or de deux choses l'une : ou bien les sujets réussissent massivement dans ces tests, ou bien ils échouent avec une fréquence non négligeable. Dans le premier cas, le psychologue antilogiciste n'a plus de grain à moudre ; dans le second, la première hypothèse à considérer est que les particules du langage naturel qui dénotent des constantes logiques ne prennent pas pour les sujets, dans le contexte de la tâche, le sens de ces constantes. Il se trouve que c'est, du moins pour la plupart des règles¹², la première hypothèse qui est la bonne, ce qui nous dispense d'envisager ici la seconde, mais nous y reviendrons. Répondons à une objection que le second de nos exemples pourrait susciter : il ne serait pas étonnant, dira-t-on, que plus d'un sujet refuse de répondre « oui » à la question posée. En effet : elle est stupide. Pour mettre à l'épreuve la capacité de nos sujets à effectuer une élimination de la conjonction, il faudrait imaginer une formulation pragmatiquement acceptable. On pourrait par exemple s'assurer que, sachant que P et Q, le sujet se comporte infailliblement, *ceteris paribus*, comme s'il savait que P (par exemple qu'il croque ce qu'il a de comestible dans la main chaque fois qu'il a envie de croquer un bonbon, et alors qu'il sait qu'il a dans la main un bonbon et un cure-dent).

Deux difficultés sont ainsi apparues à la faveur de la dernière partie de la discussion. La première est que la recherche du raisonnement élémentaire, dès lors qu'elle est guidée par la conception traditionnelle, donc essentiellement logique, du raisonnement, semble fatalement conduire aux règles de la déduction naturelle, et que ces dernières ont peu de chances de donner lieu à des données empiriques permettant d'asseoir une conception antilogiciste du raisonnement. Mais quel autre guide que la logique suivre pour découvrir le raisonnement élémentaire ? On est tenté à ce point de poser une question à la Cohen¹³ : la non-logicité du raisonnement élémentaire peut-elle être établie par l'expérience ?

La seconde difficulté vient de ce que la validité de toute expérience sur le raisonnement élémentaire repose sur la combinaison de deux exigences : la saisie consciente par le sujet du problème posé (et non d'un autre problème), et la mise en œuvre spontanée d'une stratégie de résolution qui ne peut être consciente en tant que telle, dès lors que l'élémentarité de la tâche interdit que cette stratégie consiste en l'application délibérée d'une suite d'étapes. Or ni l'une ni l'autre de ces exigences n'est facile à satisfaire, et leur conjonction encore moins. A l'appui de ces remarques, rappelons, pour la première, les défauts que l'analyse de Sperber, Cara & Girotto met en évidence dans tout le corpus expérimental post-wasonien, et pour la seconde l'obscurité dont s'enveloppe la théorie des modèles mentaux : il semble implausible que le processus de construction puis d'étude des modèles soit inconscient ; mais il n'est pas davantage plausible qu'il soit conscient (Johnson-Laird n'étudie pas de protocoles de sujets, et

¹¹ On trouve dans Braine (1990) une distinction analogue entre « primary » et « secondary deductive skills ».

¹² Comme on l'a dit à propos de la logique mentale proposée par Braine, la possibilité que la logique intuitive ou spontanée (à quelque niveau qu'on veuille la situer) soit plus faible que la logique classique ne pose pas de difficulté conceptuelle ; elle soulève au contraire des questions empiriques intéressantes.

¹³ Cf. le titre de Cohen (1981).

n'apporte aucune autre donnée susceptible de soutenir l'hypothèse de processus conscients ; quant à l'introspection, elle plaiderait fortement, sauf peut-être chez un partisan convaincu de la théorie, contre l'interprétation consciente).

c) Troisième question : à quel niveau prétendons-nous rendre compte du comportement du SLN ?

Cette question peut se comprendre d'abord à la lumière du problème qui vient d'être soulevé, celui de la « saisie » de la tâche par le sujet et du caractère conscient de la stratégie de résolution. Mais elle est incontournable dans toute entreprise s'inscrivant dans le projet des sciences cognitives : dès lors qu'il est admis que la richesse des explications computationnelles ou informationnelles réside dans la possibilité de découpler au moins deux niveaux d'explication, et souvent en fait davantage, il est essentiel de dire à quel niveau se situe l'explication que l'on propose. Nous en arrivons ainsi à la deuxième partie de notre discussion.

3. 2 Bases cognitives du raisonnement

A ce point, le lecteur éprouve peut-être un sentiment de grande confusion. Il pouvait lui sembler au départ qu'il était placé devant une question bien posée, celle de départager le psychologue antilogiciste de son adversaire, partisan de la logique mentale. Abstraction faite des difficultés que nous avons cru détecter plus haut, nous pouvons formuler ainsi ce qui les sépare : selon le premier, le sujet résout le problème « à l'aide » de X, où X est (i) effectivement mis en œuvre, (ii) non consciemment invoqué, et (iii) tout différent de la logique ; le second affirme (i), il peut accepter (ii), mais il remplace (iii) par un (iii') de la forme « X est un certain système de déduction naturelle, éventuellement affaibli, et augmenté de règles procédurales ». Et si la question peut maintenant paraître mal posée, cela tient, pour simplifier et comme on vient de voir, à deux raisons principales : l'une est que l'on n'a aucune vision claire de ce qu'il faudrait prendre comme « problème » à faire résoudre, l'autre que le statut de X est obscur.

L'enjeu est alors le suivant : on voudrait d'une part se faire une idée plus claire de la forme que pourrait ou devrait prendre une théorie cognitive du raisonnement ; on voudrait d'autre part expliquer les phénomènes mis au jour par les psychologues, et comprendre ce dont les théories qu'ils avancent sont des théories.

Le second objectif est un exemple de « recherche inverse » dans lequel, après avoir montré que la réponse proposée ne répond pas à la question posée, on cherche à déterminer la question à laquelle elle répond. L'élucidation par Sperber et coll. de la tâche de sélection en apporte une illustration. Il faudra étendre leur démarche à d'autres types d'expérience, ce qui dépasse largement et notre propos et nos compétences.

Quant au premier objectif, s'il reste hors de portée, il n'est pas interdit de tenter d'y contribuer en examinant la question du statut, en particulier du niveau, du fameux système X. Nous nous contenterons ici, à titre d'ébauche de ce programme ambitieux, de suggérer une distinction tripartite et d'en déduire une caractérisation très générale des difficultés rencontrées dans le domaine que nous avons examiné.

Personne ne peut raisonnablement nier que l'être humain mette en œuvre, dans toute démarche de raisonnement, une capacité fondamentale qui partage avec la logique au sens habituel certaines propriétés essentielles, à commencer par la généralité, dont, remarquons-le, la formalité n'est qu'une espèce particulière. On a vu, par exemple, que pour déclencher et mettre en œuvre la stratégie conjecturée dans la théorie des modèles mentaux, il fallait disposer de ressources logiques en ce sens. Pour prévenir certaines objections, précisons que nous refusons la qualité de raisonnement aux processus quasi perceptifs ou réflexes — ceux (par exemple certains partisans du connexionnisme) qui pensent que tous les processus cognitifs sont de cet ordre, ou s'y ramènent au niveau de description approprié, maintiendront leur désaccord ; mais les psychologues dont il a été question ici ne sont pas du nombre ; au contraire, ils sont tous acquis à la doctrine cognitiviste ; Johnson-Laird en particulier est l'auteur d'une introduction aux sciences cognitives¹⁴ qui est un véritable manifeste du programme en question (il tient en particulier que la cognition s'explique

¹⁴ *The Computer and the Mind*, Johnson-Laird (1988).

comme un ensemble de processus *formels*, ce qui peut, mais ne doit pas, étonner de la part d'un aussi ardent défenseur de l'idée que le raisonnement n'est *pas* formel, mais, comme il dit, *sémantique*).

Nous admettons donc en premier lieu, pour les besoins de la discussion, l'existence d'une capacité cognitive, ou psychologique, qui, étant à la fois à la base de toute autre capacité postulée et de nature au moins quasi logique, mérite d'être qualifiée d'*Urlogique* [UL]. C'est là que se situe en particulier ce qui permet au sujet logiquement naïf de saisir les notions de conséquence et de contradiction logique.

En deuxième lieu, nos psychologues étudient le comportement des sujets humains face à des problèmes de nature logique. Ce comportement est gouverné par une capacité cognitive que nous pouvons nommer par souci de brièveté *résolution logique* [RL], puisqu'elle est ce en vertu de quoi des problèmes logiques sont (plus ou moins bien) résolus, sans pour cela présupposer qu'elle possède elle-même un caractère logique, moins encore exclusivement logique.

Enfin, nous devons postuler une troisième capacité qui est celle dont est armé, comme nous l'avons vu, le sujet logiquement naïf et lui confère ce que nous avons appelé, nous inspirant de Braine, son aptitude déductive primaire. Cette capacité mérite donc l'appellation de capacité déductive de base [CDB]. Remarquons, sans nous y attarder, que CDB pourrait aussi être vue comme une « logique naïve », ou populaire, ou encore spontanée — ce qu'on nommerait en anglais *folk logic*. Elle concerne en effet l'attitude spontanée du sujet naïf devant les pensées comprenant des chevilles correspondant en un certain sens aux constantes logiques.

A l'aide de ces trois notions, voyons si nous pouvons proposer des formulations simples des positions en présence. Commençons par ce que les différentes écoles de psychologie cognitive ont en commun. Elles admettent toutes que RL est logiquement défectueuse ; elles présupposent toutes nécessairement UL, avons-nous soutenu, sans généralement le dire, ni pour certaines être disposées à l'admettre¹⁵ ; enfin, elles voient dans CDB l'enjeu empirique essentiel.

Venons-en aux différences. Les théories du premier ordre portent sur CDB et en donnent des descriptions différentes et, dans certains cas et sur certains aspects, incompatibles. Les théories du second ordre sont avancées dans le but de rendre compte des conditions dans lesquelles CDB est mobilisé par RL ; certaines sont de forme négative ou limitative — ainsi les considérations de complexité¹⁶, de facteurs de performance ; d'autres introduisent un processus X qui entre en interaction avec CDB pour produire RL — ainsi l'heuristique d'Evans ou l'interprétation pertinentielle de la tâche chez Sperber et coll. (il arrive pour certaines tâches, on l'a vu, que X fasse à lui seul tout le travail).

Sur le plan conceptuel, les psychologues polémiquent et s'accusent de confusions¹⁷ ; les philosophes et les logiciens¹⁸ hésitent à entrer directement dans le débat, décidément trop confus à leurs yeux. Macnamara parmi les premiers, Harman parmi les seconds, font des contributions positives. Harman a réfléchi sur UL et fourni des raisons de douter que CDB ait la forme d'un système logique. Macnamara, partisan de la logique mentale (qu'il appelle psycho-logique), s'est interrogé sur son statut : si l'on postule en quelque sorte franchement une réalisation de la logique dans l'appareil cognitif, la distinction entre CDB et UL ne s'impose pas avec la même évidence que si l'on rejette un tel postulat ; mais la notion même de réalisation oblige à s'interroger sur le soubassement cognitif de la psycho-logique. Ce

¹⁵ Le titre d'un article de Johnson-Laird est significatif : « Reasoning without logic » (1986).

¹⁶ Nous disions que ces considérations semblent reposer sur une confusion. Il est exact qu'un système computationnel donné se heurte à une barrière de complexité. Mais rien ne dit que les tâches trop complexes entrent dans son domaine d'application naturelle. Dans le cas présent, on ne voit pas pourquoi CDB devrait résoudre des problèmes difficiles (ces problèmes relèvent des *secondary skills* au sens de Braine, et l'homme semble plutôt *bien* équipé pour les résoudre !) La situation est toute différente dans le cas de la vision, par exemple : nous savons que des tâches de haute complexité incombent au système visuel primaire ; c'est bien pourquoi les descriptions qu'on en propose font appel au parallélisme, à des processeurs hautement spécialisés, etc., aux antipodes de ce qui est envisagé pour CDB.

¹⁷ Voir en particulier Rips (1986) et le débat dans Johnson-Laird & Byrne (1993).

¹⁸ Le commentaire de W. Hodges, le logicien, dans Johnson-Laird & Byrne (1993), est caractéristique à cet égard.

soubassement, Macnamara le situe dans un langage de la pensée comme le conçoit Fodor : c'est lui qui jouerait le rôle de UL, CDB étant la psycho-logique.

On commence ainsi peut-être à entrevoir la possibilité de restructurer le débat et de surmonter au moins en partie ce que Evans appelle la « fragmentation » du domaine. L'objectif commun serait une théorie intégrée des capacités UL, CDB et RL. La difficulté théorique est néanmoins considérable : il ne s'agit pas seulement de déterminer la structure de chacune de ces capacités ; il faut aussi, et même, logiquement parlant, avant tout déterminer leur *statut*. La difficulté empirique n'est pas moins redoutable : nous semblons confrontés à un cas grave de problème de Duhem-Quine — quelles parts reviennent respectivement aux capacités UL, CDB et autres dans RL ? Comment concevoir des expériences qui contournent l'obstacle ?

Rien ne dit, d'ailleurs, que RL puisse faire l'objet d'une description unifiée : d'une part les facteurs de performance peuvent s'y opposer pour des raisons analogues à celles que Chomsky invoque pour le langage ; d'autre part il n'est pas exclu que des mécanismes très différents rendent compte de différentes familles de comportements déductifs.

3. 3 Logique, langage et pensée¹⁹

Une question demeure, qui se rattache à deux des interrogations précédentes, sans se confondre avec elles. On s'est demandé, d'une part, comment il est possible de s'assurer que l'interprétation faite par le sujet de la tâche à accomplir est identique à celle que l'expérimentateur a en tête ; on s'est interrogé, d'autre part, sur le statut d'une *Urlogique*. L'interprétation des constantes logiques, plus exactement des mots ou locutions utilisés dans l'énoncé du problème, est au cœur de celle de l'énoncé, puisque c'est l'aptitude logique que l'on cherche à évaluer, non l'aptitude à saisir le sens des autres éléments du texte ; cette interprétation est aussi au cœur de la question de l'*Urlogique*, puisque celle-ci doit contenir au moins les prémices, les précurseurs (en un sens quasi biochimique) d'une compréhension des parties du discours dénotant les constantes logiques (nous les appellerons, pour faire bref, les *mots logiques*) : la question de l'aptitude logique ne se pose qu'à propos d'organismes ayant une manière d'accès aux constantes logiques.

L'essentiel des expériences (en particulier celle de Wason) portant sur la logique propositionnelle, nous nous restreindrons aux mots logiques correspondant (en quel sens exactement, c'est tout le problème) aux connecteurs. Aucune difficulté n'est soulevée (dans le contexte présent) par la négation ou la conjonction. C'est la disjonction et surtout l'implication qui font problème. (Cette partition n'est pas fortuite, mais nous ne nous y attarderons pas.) Les difficultés de la disjonction sont peu analysées ; en fait ce sont pour l'essentiel celles de l'implication, avec en plus l'ambiguïté entre les sens exclusif (dont on dit volontiers qu'il est celui du « ou » naturel) et inclusif. L'implication est en tout cas au centre de toute discussion du « sens » des connecteurs, et c'est à tenter de dissiper certaines des confusions auxquelles elle donne lieu que nous nous attacherons.

Nous partirons d'un constat : la plupart des psychologues (et ils ne sont pas les seuls) semblent tenir pour acquis que l'implication de la logique (l'implication dite matérielle, i.m.) n'a pas le sens du « si » naturel. Sur ce point, partisans et adversaires de la logique mentale s'accordent. C'est ainsi qu'ils s'interrogent sur les raisons que peuvent avoir les sujets d'interpréter le « si » naturel comme la bi-implication logique (par exemple dans « Si tu es sage, je te donnerai un bonbon »). Et ils se fondent sur les fameux « paradoxes de l'implication matérielle »²⁰ pour arguer que l'i.m. ne saurait avoir le sens du « si » : on lit couramment, sous la plume d'auteurs habituellement lucides, des affirmations, présentées comme des évidences, telles que « dans le raisonnement ordinaire, du faux rien ne s'ensuit »,

¹⁹ Cette section présente pour la première fois quelques considérations qui vont à l'encontre d'opinions très largement partagées ; elles sont en ce sens originales et sujettes à caution, et j'aurais préféré m'en tenir à la discussion des positions internes à la problématique générale, comme je fais pour l'essentiel dans le reste de cet article. Mais cela m'a été impossible. Une présentation plus approfondie de mon argumentation devrait paraître sous la forme d'un article que je prépare.

²⁰ Ou d'autres, fabriqués pour la circonstance. Ainsi pour Jonathan Cohen (in Johnson-Laird & Byrne 1993, p. 341), la logique, et non le sens commun, autorise l'inférence de (a) *Si l'auto de Jean est une 2CV, Jean est pauvre, et si c'est une Rolls, il est riche* à (b) *Ou bien : si l'auto de Jean est une Rolls, il est pauvre, ou bien : si c'est une 2CV, il est riche*. Cet exemple est un concentré de perfidie.

par opposition à (ce que dicterait) l'i.m. Ou encore, on demande à des populations de sujets de dire s'il est vrai que p entraîne q lorsque p est faux, et on s'interroge sur les réponses, qui vont d'un refus de se prononcer à la réponse « non », en passant par divers intermédiaires. Tout cela est *totalemment* dénué de sens, malgré l'apparence de caution apportée par la tradition de logiciens qui de Lewis à Belnap ont prétendu que l'i.m. présentait de graves défauts qu'il fallait pallier. La position que nous prenons le risque de présenter ici sans la défendre s'articule autour des points suivants.

1. En tant que connecteur (donc dans son rôle logique au sens strict, notamment en tant qu'elle donne lieu à des tautologies jugées par certains « paradoxales »), l'i.m. ne possède pas de *sens* au sens où les mots du langage naturel ont un sens : c'est une erreur de catégorie caractérisée que de lui en attribuer un, que ce soit à partir de son « rôle inférentiel » ou de sa table de vérité.

2. La relation de conséquence prouvable (ou de conséquence tautologique, il n'est pas indispensable ici de les distinguer, et nous les confondrons, pour les besoins de la discussion, sous le nom de conséquence logique) appartient, elle, au métalangage, et possède bel et bien un sens strict auquel on peut envisager de comparer celui que lui attribue le sens commun. Cela suppose évidemment que le sens commun ait à son égard une opinion définie, et sur ce point tout dogmatisme serait déplacé. L'attitude qu'il convient d'adopter dans cette affaire est d'ailleurs un problème en soi, qui a été longuement discuté, entre autres, par Cohen et par Stich. Contentons-nous ici de présenter une thèse :

- (i) « A a pour conséquence logique B » signifie pour le sens commun que A est inférable de B *en toute circonstance*, ce que le philosophe peut vouloir interpréter comme « quel que soit le monde considéré » ;
- (ii) le logicien comprend que cette condition *équivalait* à « en vertu des formes logiques de A et de B » et peut réaménager, compte tenu de cette équivalence, la distribution des sens techniques de l'une ou l'autre des manières classiques de le faire.

Cette thèse repose sur la notion d'*inféribilité* dont il est évidemment crucial de donner une définition non circulaire. Avant de le faire, précisons que l'important dans cette thèse ne réside pas dans la formulation de (ii), laissée dans le vague à dessein ; mais dans l'affirmation que le rapport entre la notion naturelle et la notion technique de conséquence logique est très exactement de l'ordre de celui qui s'établit entre n'importe quelle notion « informelle » (lâche et quotidienne, *loose, everyday*) et la notion savante (stricte et technique) qu'on lui a associée au terme d'un cheminement scientifique et philosophique. Ce rapport n'est ni de différence ni d'identité, mais de représentation ou d'approximation.

3. « Inférer B de A » signifie ici (ce n'est en aucune façon la description supposée d'un usage courant, ni le rappel d'un sens technique, mais une stipulation) quelque chose comme « passer de A à B », ou « importer B sous la condition que A », ou « acheter B au prix de A », ou encore « déclencher/activer B en présence de A » — bref, inférer est une *action*. D'autre part, cette action est primitivement épistémique et non linguistique : c'est d'un commerce de pensées ou de croyances qu'il s'agit fondamentalement.

4. Le « si » du langage naturel a un sens et un seul, qui est celui d'une invitation à inférer, d'une licence d'inférer, ou encore d'une obligation d'inférer. Penser « Si A, B » (ou « B si A »), c'est se proposer d'inférer B de A ; affirmer « Si A, B », c'est inviter l'auditeur à inférer B de A. Comme on vient de le dire — et c'est crucial —, A et B sont des pensées, et ce n'est que de manière dérivée qu'ils peuvent être soit des énoncés soit des faits.

Invitation, piège, licence... : « si » est accompagné d'une pragmatique précise et obligatoire (au sens où les indexicaux le sont, par opposition aux autres parties du discours qui charrient en général une pragmatique floue et optionnelle). « Si » renvoie contextuellement à l'*existence* d'une *famille de situations* qui constitue le domaine de validité de l'invitation (ou de la licence, le domaine de fonctionnement du piège). « Si A, B » *exprime* une invitation à inférer B de A et *indique* une famille \mathcal{K} de situations (le philosophe peut vouloir parler de mondes possibles) dans lesquelles cette invitation est valable. La pragmatique de

« si » impose à la famille \mathcal{K} d'être non « A-triviale » c'est-à-dire de comporter au moins une situation dans laquelle A est vrai et une situation dans laquelle A est faux — en particulier, \mathcal{K} n'est jamais réduite à 0 ou à 1 élément. En d'autres termes, l'emploi de « si » présuppose une *variation* non triviale.

La famille \mathcal{K} peut être constituée de toutes les situations matériellement possibles dans notre monde physique, et A avoir pour contenu un événement qui peut ou non se produire ; dans ce cas, B sera un effet ayant A pour cause, et l'on pourra être tenté de parler d'un « si causal » ; mais ce n'est pas au sens de « si » que se rattache la notion de causalité, c'est à la famille \mathcal{K} pragmatiquement associée à cet emploi de « si ». De même, \mathcal{K} peut être constituée de toutes les situations sans restriction, et A avoir pour contenu une pensée composée à l'aide de chevilles logiques ; B pourra alors être une conséquence logique de A, et le « si » ne sera « logique » qu'en vertu de \mathcal{K} . Et ainsi de suite : « si » peut annoncer une connexion indicielle (« s'il y a de la fumée, il y a du feu »), une connexion conventionnelle (« si c'est février, il n'y a que 28 ou 29 jours dans le mois »), une connexion mathématique (« si n est pair, le carré de n est pair »), une connexion performative (« si vous signez, vous devenez propriétaire du château » ou « si tu bouges, tu es mort »), déontique « si tu as péché, tu dois te confesser », etc. Il peut même arriver que \mathcal{K} comporte des situations impossibles : dans « si $2 + 2 = 4$, alors $(2 + 2)^2 = 4^2$ », \mathcal{K} comprend la situation dans laquelle $2 + 2$ est différent de 4. Bref, « si » annonce l'existence d'une connexion sans encoder la moindre information sur la nature de la connexion, ce qui explique que toute connexion puisse être en cause, et que les locuteurs du français puissent avoir une parfaite maîtrise du « si » sans avoir nécessairement d'idée bien précise sur ce qu'est l'inférence logique, la cause, la convention, etc. Même les usages tels que « si tu as soif, il y a de la bière dans le réfrigérateur » ou (message laissé sur un répondeur téléphonique) « si je suis bien chez M. X, qu'il me rappelle à tel numéro » qui paraissent très éloignés du « si » prototypique relèvent du même mécanisme. Dans le premier exemple, \mathcal{K} est la famille des situations dans lesquelles tu as soif ou tu n'as pas soif, et la pensée « tu as soif », présente dans le premier genre de situation, doit (dans l'intention du locuteur) conduire l'auditeur à activer la pensée « il y a de la bière dans le réfrigérateur ». De même dans le second exemple : \mathcal{K} est l'ensemble des situations dans lesquelles l'abonné atteint est M. X et celles dans lesquelles il ne l'est pas, et la licence accordée permet à l'auditeur de passer de la pensée « c'est bien chez X » à la pensée « l'auteur du message souhaite que X le rappelle à tel numéro » (les paraphrases sont grossières, mais peu importe). Le traitement de ces exemples et d'autres du même genre, abondamment cités dans la littérature, repose sur l'idée que le « si » porte sur des pensées et non directement sur le contenu des pensées (dans les exemples prototypiques, la pensée et son contenu sont indiscernables pour tout autre que le philosophe, d'où l'apparente ambiguïté entre connexion des faits et connexions des pensées).

Toute la difficulté de « si » pour le théoricien provient de la nature très particulière du renvoi pragmatique à la famille \mathcal{K} , de son mode d'existence et de détermination, et de sa nature dans la variété indéfinie des emplois légitimes de « si ». Quant aux prétendus paradoxes et aux erreurs de sujets mis en situation expérimentale, ils proviennent d'emplois illégitimes de « si », ou de montages pervers dans lesquels par exemple deux familles différentes de situations sont pragmatiquement mobilisées autour d'un seul « si », ou deux manières de quantifier sur les situations entrent en compétition²¹. C'est du moins ce que nous conjecturons, et qu'il faudra montrer de manière détaillée et systématique en un autre lieu.

5. S'il est vrai que l'i.m. n'a pas de « sens » au sens habituel, on ne peut se soustraire à l'obligation d'examiner ce qu'on pourrait peut-être appeler son « sens projeté » (projeté sur le

²¹ A titre d'illustration de ce dernier phénomène, je propose à la sagacité du lecteur l'explication du « paradoxe » obtenu en remplaçant dans la tautologie suivante :

$$[p \rightarrow (p \wedge q)] \vee [q \rightarrow (p \wedge q)],$$

la proposition p par « je suis riche » et la proposition q par « tu es riche ». (*Indication* : la quantification [implicite, métalinguistique] à laquelle se prête la tautologie porte sur la formule entière, alors que la paraphrase en langue naturelle force une quantification sur chacun des membres de la disjonction ; le paradoxe se réduit à l'erreur classique de distribuer la quantification universelle sur une disjonction, et n'a rien à voir avec l'acceptabilité de la table de vérité de l'i.m.)

langage naturel) et qui, selon la plupart des auteurs, serait différent du sens du « si » naturel. Ce sens projeté, attribuons-le par convention à un nouveau mot, « sim ». C'est ce que certains des auteurs en question appellent le « si » de la logique. « Sim A, B » (ou « B sim A ») est assertable, disent-ils, même lorsque A est faux, ou lorsque B est vrai « indépendamment » de A, etc., si bien que l'on peut dire « sim Paris est la capitale de la Prusse, ma tante est mon oncle », mais aussi « sim Paris est la capitale de la France, ma tante est mon oncle » et « sim Paris est la capitale de la Prusse, ma tante est ma tante », toutes choses inassertables avec « si » au lieu de « sim », sauf, en un sens ironique, la première (sur le patron de laquelle est construit le fameux « je veux bien être pendu sim Paris est la capitale de la Prusse »).

Qu'un tel mot n'existe pas dans la langue est évident ; ce qu'il faut expliquer, c'est d'une part pourquoi il n'existe pas, et d'autre part pourquoi la logique peut et doit attribuer à l'i.m. la table de vérité que l'on sait. Ce sont des questions difficiles, et nous devons nous contenter ici d'une indication. Si « sim » n'existe pas, c'est qu'il placerait l'auditeur d'une phrase le contenant dans une situation impossible — soit de rechercher une famille de variations tout en étant empêché de considérer toute autre situation que celle que la prémisse spécifie, soit de rechercher une connexion tout en étant obligé d'exclure qu'elle existe : il faudrait qu'à la fois on puisse savoir et ne pas savoir ce qu'il en est de A, ou savoir et ne pas savoir que A et B sont liés. Ce serait soit un « si » muni d'une pragmatique incohérente, soit un « si » arbitrairement privé de toute pragmatique, ce qui constituerait une incohérence au niveau supérieur.

Quant à la table de vérité de l'i.m., elle se comprend à la lumière de la distinction entre *fait* et *raison* : l'i.m. exprime *dans une situation entièrement fixée mais non nécessairement connue* le fait disjonctif « non-A ou B » ; la validité de cette disjonction *dans une variété logiquement définissable de situations* renvoie à une *raison* (nécessairement de nature logique) de cette invariance. Le « si » naturel indique de même *l'existence* d'une raison (non nécessairement logique) qui explique et garantit l'invariance relativement à une variété non nécessairement logiquement définissable de situations ; lorsque c'est un fait *singulier connu* qu'il s'agit d'exprimer, alors ce fait est soit non-A soit B, et il serait contraire aux règles de la pertinence d'appauvrir l'information « non-A » (respectivement « B ») en « non-A ou B ».

Quoi qu'il en soit, puisque « sim » n'existe pas, que peut faire le sujet d'expérience que l'on cherche à contraindre à interpréter « si » comme « sim » ? On le trompe, on se joue de lui : il n'est pas placé dans une situation normale de résolution de problème. Il doit faire face à une situation perverse, qu'il saisit mal, et ne peut que faire au mieux.

Que conclure de toute cette discussion, de manière plus générale, en ce qui concerne les expériences sur le raisonnement déductif ? C'est loin d'être clair. Trois difficultés semblent toutefois se dégager nettement. La première porte sur la distinction entre possession ou maîtrise des concepts correspondant aux mots logiques et aptitude déductive de base : quelle est au juste cette distinction, et comment en tenir compte dans l'investigation expérimentale du raisonnement élémentaire ? La seconde concerne la distinction entre conséquence logique et conséquence tout court : il est peu probable que la première de ces notions soit claire dans l'entendement naïf, *a fortiori* ne doit-on pas compter sur une perception claire de la différence. On ne peut donc mettre les capacités déductives à l'épreuve chez un sujet logiquement naïf en lui demandant de raisonner dans des situations dont la compréhension exige une maîtrise de cette différence (ce qui est le cas de la tâche de Wason). La troisième concerne le sens des mots logiques de la langue naturelle : s'il est très contestable que ces mots dénotent des constantes logiques non standard, il est non moins improbable qu'ils puissent indiquer par eux-mêmes la nature logique ou non logique des connexions qu'ils dénotent. On peut douter, par conséquent, qu'une formulation en langage naturel d'un problème de logique soit possible sans recours à une mise en scène explicitement logique, mise en scène qui ruine la possibilité de mettre à l'épreuve les capacités d'un sujet logiquement naïf !

4. UNE CATEGORIE DE RAISONNEMENTS INDUCTIFS : LE JUGEMENT EN SITUATION D'INCERTITUDE

On oppose traditionnellement, chez les logiciens, le raisonnement déductif au raisonnement inductif. Harman, on l'a vu, fournit des raisons de douter, à supposer qu'il y ait

là une distinction à faire, que ce soit une bonne manière de la formuler. Parlons donc plutôt de deux sortes d'arguments intervenant dans le raisonnement : ceux dans lesquels la conclusion suit avec certitude des prémisses (inférences *démonstratives*), et ceux dans lesquels elle ne suit qu'avec un degré d'incertitude (inférences *non démonstratives*).

Sous la dénomination de « jugement en situation d'incertitude »²², s'est développé, avec un retard de quelques années sur le paradigme de Wason, un programme de recherche en psychologie cognitive expérimentale visant à évaluer et à expliquer certaines capacités humaines en matière d'arguments de la seconde espèce²³. Ce programme (également appelé « *heuristics and biases* » du nom des deux principales notions mobilisées) a connu un développement quantitatif, mais surtout un retentissement, bien supérieurs à son homologue « déductif ». Nous reviendrons dans un instant sur l'importance prise par ce courant de recherche en l'espace d'une vingtaine d'années, depuis les premières publications d'Amos Tversky et Daniel Kahneman, ses chefs de file longtemps incontestés. Regardons d'abord quelques-uns de ses résultats les plus caractéristiques.

4. 1 Les expériences

Il y a deux façons d'évaluer une probabilité empirique : directement à partir de fréquences observées, ou indirectement à partir de données comparatives ou absolues, qualitatives ou quantitatives, sur des probabilités, et de relations entre les événements considérés. Toute détermination indirecte repose sur l'un ou l'autre des principes de la théorie des probabilités — on peut du moins le supposer.

a) *Observation directe de fréquences*

D'un ensemble assez complet d'expériences que nous laisserons en dehors de la discussion, il apparaît que de manière générale l'homme repère assez bien les fréquences d'événements identiques au sein d'une série d'épreuves (par exemple il évalue et classe sans erreurs importantes les proportions de boules de différentes couleurs tirées d'une urne sous ses yeux). On signale cependant une tendance à surestimer les fréquences très faibles. D'autre part, on a mis en évidence des perturbations plus graves lorsque les fréquences portent sur des événements véhiculant des informations significatives — une manifestation de plus de l'« effet de contenu ». Cette catégorie d'erreurs constitue l'un des premiers résultats de Tversky et Kahneman :

Expérience I. On fait défiler devant le sujet deux listes de personnalités.

La liste A comporte 19 hommes très célèbres, 20 femmes connues mais moins célèbres.

La liste B comporte 19 femmes très célèbres, 20 hommes connus mais moins célèbres.

On demande au sujet de comparer, pour chaque liste, le nombre d'hommes et de femmes.

Résultat : 80 % des sujets estiment qu'il y a plus d'hommes que de femmes dans A, et *vice-versa* dans B. (Kahneman & Tversky 1973)

Nous verrons un peu plus loin comment les auteurs de l'étude se proposent d'expliquer le phénomène. Mais leur explication s'applique à d'autres erreurs ; il sera plus économique de passer les exemples en revue avant de présenter les explications disponibles.

²² Comme nous allons le voir, il s'agit essentiellement, mais pas exclusivement, d'expériences sur des raisonnements probabilistes ou statistiques. C'est du moins ainsi qu'elles sont généralement présentées.

²³ Il ne sera question ici que d'un certain type d'argument non démonstratif ; les autres recherches dans le domaine portent principalement sur l'acquisition inductive de concepts, ou l'attribution de propriétés à partir d'informations finies, ou encore la saisie d'une règle à partir d'exemples. On peut observer à ce propos que la manière dont les psychologues découpent le domaine du raisonnement en sous-domaines est assez surprenante pour le logicien ; ainsi Garnham et Oakhill, dans leur manuel (1994), distinguent-ils entre « Raisonnement déductif » (chap. 5), « Raisonnement syllogistique » (chap. 6), « Induction » (chap. 7), « Mise à l'épreuve [*testing*] d'hypothèses » (chap. 8 – dans lequel est exposée notamment la tâche de sélection de Wason), « Raisonnement statistique » (chap. 9 – dans lequel il est question de ce qui nous occupe dans la présente section), et enfin « Prise de décision » (chap. 10 – de la réalité psychologique des principes de la théorie de la décision).

Les erreurs les plus significatives concernent, on ne s'en étonnera pas, les cas de détermination indirecte de fréquence. Toutes sortes de violations des principes probabilistes ont été mises en évidence. Nous présentons les principales, classées selon la difficulté apparente des principes probabilistes impliqués.

b) Sophisme de la conjonction

Peu de principes probabilistes sont plus simples et apparemment naturels que le principe de la conjonction, selon lequel la probabilité d'une conjonction d'événements ne peut excéder les probabilités de chacun des événements conjoints. Or ce principe semble régulièrement violé :

Expérience II. Linda a 31 ans, elle est célibataire, elle a son franc-parler, elle est très intelligente. Elle a terminé des études de philosophie. Etudiante, elle était extrêmement sensible aux questions de discrimination²⁴ et de justice sociale ; elle a également pris part à des manifestations anti-nucléaires.

Veillez noter les propositions suivantes selon leur probabilité, en donnant 1 à la plus probable et 8 à la moins probable :

- a) Linda est institutrice.
- b) Linda travaille dans une librairie et suit des cours de yoga.
- c) Linda milite dans le mouvement féministe.
- d) Linda est assistante sociale en milieu psychiatrique.
- e) Linda est membre de la Ligue des Electrices [*League of Women Voters*]
- f) Linda est employée de banque.
- g) Linda est représentante en assurances.
- h) Linda est employée de banque et milite dans le mouvement féministe.

Résultat : 89 % des sujets placent (h) avant (f). (Tversky & Kahneman 1983)

Expérience II bis. Même expérience, sauf qu'on ne propose que (f) et (h).

Résultat : 85 % des sujets placent (h) avant (f).

Expérience III. Une enquête de santé publique a été effectuée en Colombie britannique²⁵ sur un échantillon d'hommes adultes de tous âges et de toutes professions.

Veillez donner votre estimation des valeurs suivantes :

- (a) Le pourcentage des hommes étudiés qui ont subi un infarctus ou plus.
- (b) Le pourcentage des hommes étudiés qui ont 55 ans ou plus et ont subi un infarctus ou plus.

Résultat : 65 % des sujets fournissent un pourcentage supérieur pour (b) que pour (a). (Tversky & Kahneman 1983)

Expérience IV. Un dé non pipé porte 4 faces vertes et 2 rouges. On va lancer le dé 20 fois et on enregistrera la suite des verts (V) et des rouges (R) qui sortent. On vous demande de choisir l'une des suites qui figurent parmi les trois ci-dessous ; vous recevrez 25 \$ [de 1983 — environ 300 F de 1994] si une suite de jets consécutifs est identique à la suite que vous avez choisie. Cochez la suite sur laquelle vous souhaitez parier :

- A — R V R R R
- B — V R V R R R
- C — V R R R R R

Résultat : Une majorité de sujets (260 étudiants de Stanford et de l'University of British Columbia) cochent la suite B. (Tversky & Kahneman 1983)

²⁴ Est-il besoin de préciser que ces expériences ont été menées en Amérique du Nord ?

²⁵ Où Kahneman a enseigné de longues années.

c) Sophisme du taux de base (non-prise en compte des chances a priori)

Dans de nombreux problèmes, la solution dictée par la théorie des probabilités est fonction de la composition de la population de référence (donnée par ce qu'on appelle les taux de base ou chances *a priori*). On sait depuis assez longtemps²⁶ que les sujets, placés devant ce genre de problème, se montrent généralement insensibles au rôle des taux de base. Kahneman et Tversky ont proposé une série d'expériences devenues célèbres pour mettre en évidence la robustesse de cette erreur.

Expérience V. On compte, dans une certaine ville, 85 % de taxis verts et 15 % de bleus. Un accident a lieu dans lequel un taxi est impliqué. Un témoin affirme que le taxi était bleu. Une série de tests montre que le témoin est fiable à 80 % : il identifie correctement la couleur d'un taxi, qu'il soit vert ou bleu, dans 80 % des cas. Quelle est la probabilité que le taxi impliqué soit effectivement bleu ?

Résultat : La grande majorité des sujets non formés en probabilités répond : environ 80 %. [L'application du théorème de Bayes à partir du taux de base de 15 % de taxis bleus fournit 41 %.] (Kahneman & Tversky 1972a)

Expérience VI. Un test existe pour détecter une maladie rare. La maladie affecte 1 personne sur 1 000 dans une certaine population, et le test a un taux de faux positifs de 5 %. On teste un patient et il est positif. Quelles sont les chances, en l'absence de toute autre information, que le patient ait effectivement la maladie en question ?

Résultat : Sur 60 étudiants et enseignants de la Harvard Medical School, près de la moitié estiment les chances à 95 % ; la moyenne des réponses est 56 %, la réponse 2 % [correcte selon la formule de Bayes] est donnée par 18 % des sujets. (Casscells, Schoenberger & Grayboys 1978 — repris par Tversky & Kahneman 1982a)

d) « Loi des petits nombres » — signification et taille des échantillons

La loi des grands nombres, dont l'évidence est reconnue depuis l'aube du calcul des probabilités et la démonstration mathématique un exercice standard, ne donne pas seulement aux philosophes du fil à retordre²⁷, comme semblent l'attester les expériences suivantes.

Expérience VII. Il y a deux maternités dans une certaine ville. Dans la première, environ 45 bébés naissent chaque jour ; dans la seconde, 15. Comme vous le savez, à peu près 50 % des nouveaux-nés sont des garçons. Le pourcentage exact varie néanmoins d'un jour sur l'autre. Il peut être tantôt supérieur, tantôt inférieur. Au cours d'une année donnée, chaque maternité a compté le nombre de journées où ce pourcentage était supérieur à 60 %. Selon vous, quelle est la maternité qui a compté le plus grand nombre de telles journées ?

Résultat : Une majorité des sujets estime qu'il n'y a pas de différence entre les deux maternités. Les autres sujets se partagent à peu près à égalité entre les deux autres réponses possibles. (Kahneman & Tversky 1972)

Expérience VIII. On répartit 20 billes au hasard entre 5 enfants, André, Béatrice, Charles, Diane et Eléonore. On recommence un grand nombre de fois. Des deux répartitions suivantes, laquelle est la plus fréquente ?

Répartition I A = 4, B = 4, C = 5, D = 4, E = 3

Répartition II A = 4, B = 4, C = 4, D = 4, E = 4

Résultat : Une majorité des sujets estime que I est plus fréquente que II. (Kahneman & Tversky 1972)

²⁶ Il semble que ce soit dans le contexte du diagnostic médical que l'effet dont il va être question a été mis pour la première fois en évidence par des psychologues (Meehl & Rosen 1955).

²⁷ Karl Popper insiste sur l'impérieuse nécessité de rendre intelligible cette loi, que ses démonstrations dans un cadre axiomatique ne protègent pas du risque d'apparaître mystérieuse, voire miraculeuse. Voir Popper (1983).

Expérience IX. On considère une urne contenant des boules, dont deux tiers sont d'une couleur et un tiers d'une autre. Amédée tire 5 boules, dont 4 rouges et 1 blanche. Blandine tire 20 boules, dont 12 rouges et 8 blanches. Lequel des deux est en droit de se sentir le plus sûr de la répartition $2/3$ de rouges et $1/3$ de blanches, plutôt que l'inverse ?
Résultat : La majorité des sujets jugent qu'Amédée est plus justifié dans sa réponse que Blandine. [Selon le calcul des probabilités, Blandine est 2 fois plus justifiée qu'Amédée.] (Tversky & Kahneman 1974)

De ces expériences et d'autres se dégage l'hypothèse que les sujets ont du mal à saisir le rôle de la taille comparée des échantillons : ils respectent la « loi des petits nombres », affirment Tversky & Kahneman (1971), c'est-à-dire qu'ils tendent à surestimer le caractère significatif des petits échantillons. Le célèbre sophisme du joueur, dont l'expérience prouve abondamment que les sujets le commettent très souvent²⁸, relève de la même erreur.

e) Incompréhension de la variance

Une distribution de probabilité ne détermine pas seulement l'espérance, mais aussi les moments d'ordre supérieur, en particulier la probabilité d'un écart donné par rapport à l'espérance. Les sujets semblent mal armés pour évaluer une telle probabilité, ce qui n'étonnera sans doute pas grand monde. Le sens d'une expérience telle que la suivante est de suggérer des stratégies susceptibles d'être mobilisées par les sujets dans des cas où la compétence probabiliste leur fait défaut.

Expérience X. Un lycée comprend deux filières. Dans la filière A, il y a 65 % de garçons ; dans la filière B, il y en a 45 %. Il y a le même nombre de classes dans chaque filière. Vous entrez dans une classe au hasard et vous comptez les garçons : il y en a 55 %. Si vous deviez parier à quelle filière appartient la classe en question, choisiriez-vous A ou bien B ?

Résultat : La plupart des sujets parient A. [Or la loi binomiale a une variance plus forte pour $p = 0,45$ que pour $p = 0,65$, donc le bon pari est B, puisque l'écart est le même (0,1) par rapport à chacune des deux moyennes.] (Kahneman & Tversky 1972)

f) Evaluations non probabilistes

Toujours dans l'espoir de mettre en évidence les stratégies d'estimation mises en œuvre dans le jugement incertain, diverses autres expériences ont été menées. En voici deux parmi les plus caractéristiques :

Expérience XI. Un groupe de 10 personnes doivent constituer des commissions. Combien de commissions différentes de 2 membres peuvent-elles être constituées ? Et combien de commissions de 6 membres ?

Résultat : La plupart des sujets donnent une réponse supérieure à la première question qu'à la seconde. [Au pays du triangle de Pascal, chacun sait naturellement que l'ordre est inverse.]

Expérience XII. a) Vous allez disposer d'exactly 5 secondes pour répondre à la question suivante : évaluer le produit $1 \times 2 \times 3 \times 4 \times 5 \times 6 \times 7 \times 8$.

b) Vous allez disposer d'exactly 5 secondes pour répondre à la question suivante : évaluer le produit $8 \times 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1$.

Résultat : La médiane des réponses est de 512 pour (a) et de 2 250 pour (b). [La réponse exacte est 40 320.]

²⁸ Gigerenzer *et al.* (1989) signalent même l'erreur dans le manuel d'un professeur de probabilités de l'École polytechnique dans les années 1920 !

4. 2 Les explications de l'école « Heuristics and biases »

Ces expériences, et beaucoup d'autres menées dans le même esprit, prouvent selon Tversky et Kahneman que l'esprit humain est affecté, dans ses jugements en situation d'incertitude, par ce qu'ils appellent des « biais », comparables aux biais de certaines méthodes statistiques, ou de certains instruments, et responsables d'« illusions cognitives » qui seraient l'équivalent d'illusions d'optique, inévitables même en présence de données objectives qui devraient les dissiper.

Cette façon de voir oriente la recherche vers des explications de type fonctionnel : si le système cognitif responsable des jugements en situation d'incertitude est victime d'illusions systématiques, c'est qu'il fonctionne selon des principes non conformes à la norme rationnelle, mais qui répondent sans doute à des impératifs de *design* computationnel et qui fournissent néanmoins à l'organisme des réponses rapides et acceptables dans une bonne proportion des situations écologiques. Ces principes sont des « heuristiques » au sens que ce terme a pris en IA. Tversky et Kahneman ont cru détecter trois de ces heuristiques. La première, qui est aussi la plus clairement formulée et la mieux corroborée, est l'heuristique de *représentativité* ou de *similarité* ; Daniel Osherson (1990) propose une version simplifiée de ce principe :

Pour juger de la probabilité qu'un objet *O* appartient à une catégorie *C*, évaluer le degré de similarité de *O* à *C* : plus ce degré est élevé, plus la probabilité est grande.

C'est l'application de ce principe qui explique, selon Tversky et Kahneman (1974, etc.), les résultats observés dans les expériences II à X. En fait, il faut selon eux (1982) distinguer deux formes de l'heuristique de représentativité. Dans l'expérience II, une Linda employée de banque *et* féministe est plus représentative de la jeune étudiante qu'était la Linda d'autrefois (plus semblable à la catégorie des jeunes étudiantes en philosophie intelligentes et militantes) qu'une Linda *simple* employée de banque : le jugement opère *par le biais* d'une comparaison de similarités perçues, puis conclut par application du principe simplifié d'Osherson. Dans l'expérience IX au contraire, l'échantillon d'Amédée (4 + 1) est jugé plus représentatif de la répartition 2/3 — 1/3 que l'échantillon de Blandine (12 + 8) : ici c'est la distance perçue (le degré de similarité) qui informe un jugement *sur* la représentativité, et la représentativité fixe le choix en vertu d'un principe apparemment trivial selon lequel plus un échantillon est représentatif, plus ses traits pertinents ont de chance d'être proches de ceux de la population entière.

L'heuristique de *disponibilité* expliquerait les expériences I et XI : les jugements comparatifs se basent sur la disponibilité d'exemples de chacune des catégories (dans I, pour la liste A les exemples masculins sont plus faciles à évoquer que les exemples féminins, et inversement pour la liste B ; dans XI, davantage de commissions distinctes de 2 membres sont facilement imaginables que de commissions de 6 membres, qui se recoupent fortement et se distinguent donc plus difficilement).

Enfin, l'heuristique d'*ancrage plus ajustement* expliquerait l'expérience XII : les sujets feraient une toute première estimation en effectuant les premières opérations (c'est l'ancrage), puis la corrigeraient pour tenir compte des opérations restantes qu'ils n'ont pas le temps d'effectuer ; or cet ajustement est en général insuffisant, les sujets répugnant à trop modifier leur première estimation (peut-être en raison du phénomène général de persistance des croyances initiales, illustré par la résistance au *debriefing*²⁹). On comprendrait ainsi la considérable différence entre les résultats des tests (a) et (b).

4. 3 Critiques de « Heuristics and biases »

²⁹ Il s'agit de situations, réelles ou expérimentales, dans lesquelles des sujets sont dans un premier temps induits en erreur et doivent exécuter diverses tâches sur la base de ces données trompeuses ; dans un deuxième temps, on leur explique la tromperie, ce qui ôte tout fondement à leurs croyances de la première phase ; on constate néanmoins que ces croyances persistent souvent, fût-ce sous une forme atténuée. Dans le cas qui nous occupe, la première estimation serait à la fois d'emblée tenue pour fautive, et prise, faute de mieux, comme (approximativement) juste au cours de la première étape ; le « *debriefing* » par le sujet lui-même ne suffirait pas pour le convaincre de modifier radicalement sa réponse. (Ce rapprochement est une spéculation de l'auteur du présent article.)

L'énorme importance accordée aux thèses de Kahneman et Tversky explique le nombre considérable de travaux critiques qu'elles ont suscités. Souvent, il s'agit d'expériences tendant à montrer que certains effets disparaissent lorsqu'on fait varier les paramètres ou d'autres conditions d'expérience, ce qui permettrait de légèrement « remonter la note » de notre système de jugement spontané sous incertitude. Mais quelques auteurs sont bien plus sévères : ils contestent le bien-fondé de toute l'entreprise. Nous évoquerons ici les deux critiques les plus radicaux, Gerd Gigerenzer et Jonathan Cohen.

Le premier (1991) déploie son attaque sur deux fronts. Fort de ses connaissances scientifiques et historiques de la théorie des probabilités et de la statistique, il conteste d'abord que dans son interprétation classique cette théorie ait quoi que ce soit à dire concernant la plupart des problèmes chers à Tversky et Kahneman. Selon lui en effet, ces problèmes portent sur l'évaluation de la probabilité d'un événement unique, alors que l'interprétation classique, c'est-à-dire fréquentiste³⁰, des probabilités porte sur le comportement asymptotique de suites d'épreuves identiques. C'est donc à la norme invoquée qu'il s'en prend en premier lieu³¹. En second lieu, il montre qu'en reformulant les problèmes de manière à les situer clairement dans le champ d'application de la théorie fréquentiste, on peut obtenir des améliorations spectaculaires des performances — au point que les fameuses « illusions cognitives » se dissipent, non tant chez le sujet que chez le psychologue qui avait cru les déceler. Voici, à titre d'exemple, comment Cosmides & Tooby (1990), cités par Gigerenzer, reformulent le problème de l'expérience VI :

Expérience VI_{fréq.} Un Américain sur 1 000 est atteint de la maladie X. Un test permettant de détecter la maladie chez un patient a été mis au point. Chaque fois que le test est administré à quelqu'un qui est atteint de X, le résultat est positif. Mais il arrive qu'il soit positif alors que la personne est en parfaite santé ; pour être précis, sur 1 000 personnes en bonne santé, 50 réagissent positivement au test.

Imaginez que nous ayons réuni un échantillon aléatoire de 1 000 Américains. Ils ont été choisis par tirage au sort. Ce tirage a été exécuté par des gens n'ayant aucune information sur l'état de santé des personnes concernées. Combien de personnes réagissant positivement au test sont effectivement atteintes de la maladie : ... sur ... ?

Résultat : 76 % des sujets fournissent la réponse 2 % (1 sur 50 ou 51).

Une version « iconique » du problème, dans laquelle on demande aux sujets de représenter les populations considérées sur une grille, fournie, de 10 sur 10, donne des résultats encore meilleurs (92 %).

De la même manière, la version fréquentiste de l'expérience II bis fait virtuellement « disparaître » le sophisme de la conjonction :

Expérience II bis_{fréq.} [Même description de Linda du temps de ses études.]

Il y a 100 personnes qui correspondent à cette description. Combien d'entre elles sont :

(a) des employées de banque ?

(b) des employées de banque qui militent dans le mouvement féministe ?

Résultat : À peine plus de 20 % des sujets donnent une réponse supérieure pour (b) que pour (a).

Pour Gigerenzer, ces expériences, et d'autres du même genre, montrent que les sujets, contrairement aux psychologues qu'il critique, sont sensibles à la différence entre événement singulier et fréquences dans les séries longues.

Jonathan Cohen ne prétend pas, quant à lui, opposer aux expériences de « *heuristics and biases* » d'autres expériences. Philosophe, il se place sur le terrain de l'analyse théorique, et sa

³⁰ Gigerenzer ne rejette pas l'interprétation subjectiviste, même si elle est minoritaire ; pour lui, le subjectiviste conséquent doit également rejeter les conclusions de Kahneman et Tversky, quoique pour d'autres raisons que le fréquentiste.

³¹ Il détecte aussi des erreurs de formulation ; par exemple, dans le problème VI dont il va être question, il manque le taux de faux négatifs, et surtout une spécification de la nature de l'échantillonnage : est-il représentatif de la population générale, ou d'une population « auto-sélectionnée » par les symptômes exhibés ?

première interrogation porte sur ce que des expériences de psychologie *en général* sont susceptibles de révéler des facultés humaines de raisonnement (Cohen 1981). Son étude critique est au service d'un projet plus vaste, qui est de défendre la rationalité du sens commun — d'un sujet logiquement et probabilistiquement naïf ; cet objectif est lui-même subordonné pour Cohen à la volonté de justifier la philosophie analytique, qui n'est rien d'autre, selon lui, que l'étude *raisonnée* des raisons : la *raison* qui examine les raisons (dans tous les domaines et à tous les niveaux), cette raison-là est forcément « naïve » — qu'elle se révèle foncièrement et grossièrement défectueuse, et c'est toute l'entreprise de la philosophie analytique qui se trouverait privée de raison d'être (Cohen 1986).

La question de la rationalité — et du lien entre rationalité et raisonnement — n'entre pas dans le propos du présent article ; *a fortiori* n'aurons-nous rien à dire ici de la nature et du fondement de la philosophie analytique. En revanche, l'analyse que fait Cohen des travaux de Tversky et Kahneman, plus particulièrement des expériences II et VI dont il vient d'être question, retiendra un instant notre attention.

Le volet critique de cette analyse relève de la stratégie, également appliquée par Gigerenzer, de l'arroseur arrosé : Cohen — philosophe des méthodes inductives et stochastiques³² — met en doute la capacité des psychologues à faire ce dont ils jugent leurs sujets incapables, à savoir résoudre correctement les problèmes probabilistes posés. Plus précisément, il conteste que les solutions bayésiennes présentées par Tversky et Kahneman comme seules solutions correctes soient d'une part uniques, d'autre part optimales dans le contexte des expériences incriminées. Sa cible principale est précisément le taux de base que les psychologues accusent leurs sujets de négliger : Cohen soutient d'une part qu'il peut y avoir de bonnes raisons de le faire, et en présente certaines (ce qui lui vaut, à propos de l'expérience VI, l'accusation de ne devoir qu'au fait de n'être pas directeur d'hôpital de n'être pas non plus un assassin³³ !); il rappelle d'autre part que dans d'autres expériences³⁴, les sujets se montrent parfaitement capables de prendre le taux de base en compte (il arrive même qu'ils en surestiment l'importance). Si donc les sujets négligent le taux de base dans certains cas, il se pourrait qu'ils le fassent délibérément, ce qu'ils sont *a priori* en droit de faire : c'est un choix qui relève de la sélection des données pertinentes, laquelle est par excellence affaire d'intuition (et c'est pourquoi d'ailleurs la question intéresse tant Cohen).

Pendant, en l'absence d'un candidat plausible à la qualité de raison, pour les sujets, de ne pas tenir compte du taux de base, l'argument contre les psychologues qui estiment qu'ils le font sans raison (donc irrationnellement) demeure incertain. Reconnaisant cette faiblesse³⁵, Cohen fait pour y remédier un pas important dans une direction constructive, très différente de celle que suit Gigerenzer mais jouant un rôle dialectique analogue. Il existe selon Cohen une distinction capitale entre deux sortes de probabilités. Soit C une classe de référence (une population au sens statistique) et P une qualité à laquelle est attachée une probabilité p : cette probabilité est dite *contrefactualisable* si elle s'applique aussi bien aux non-membres qu'aux membres de C . Soit par exemple C l'ensemble des ouvriers d'usine d'amiante âgés de 50 ans — à une période et dans une région considérées —, P le fait de mourir avant 70 ans, et supposons que la probabilité p pour un ouvrier de l'amiante de 50 ans de mourir avant 70 ans soit de 0,8. Cette probabilité ne dépend pas des individus particuliers qui se trouvent composer la classe P : si Malcolm, philosophe d'Oxford, se trouvait avoir travaillé depuis toujours dans une usine d'amiante et être âgé de 50 ans, la probabilité qu'il meure avant 70 ans serait égale à 0,8 ; ainsi, la probabilité p est contrefactualisable. Au contraire, supposons que la probabilité pour un ouvrier de l'amiante de 50 ans, à l'époque et dans la région considérées, d'avoir les yeux bleus soit de $q = 0,6$; cette probabilité-là est (selon toute vraisemblance) non contrefactualisable : elle ne s'applique *pas* à Malcolm. Pour le dire en peu

³² Voir Cohen (1970, 1977, 1989).

³³ Dans un commentaire de Cohen (1981).

³⁴ Les expériences qu'il mentionne (Cohen 1986, p. 163, note 21) sont antérieures, pour la plupart, à la période faste de *heuristics and biases* ; Gigerenzer souligne de même que Tversky et Kahneman mettent un terme à une période *optimiste* au cours de laquelle avaient été réunis des témoignages expérimentaux encourageants quant aux aptitudes probabilistes des sujets naïfs. Ce mouvement pendulaire illustre les « biais » qui affectent le domaine (Lopes 1991).

³⁵ C'est un progrès sensible sur Cohen 1981, accompli dans Cohen 1986.

de mots, est non contrefactualisable une probabilité associée à une propriété *accidentelle* de la classe de référence.

Il échoit au philosophe de fonder cette distinction, et Cohen s'y emploie (*op. cit.* § 18). Mais pour qu'elle soit pertinente sur le plan psychologique, il faut que le sujet naïf y soit sensible, et que cette sensibilité puisse expliquer la manière dont il résout les problèmes que lui soumet le psychologue. Cohen pense que c'est bien le cas. Selon lui, les sujets donnent spontanément la préférence aux explications causales ; or les probabilités contrefactualisables sont plus proches de la causalité que les probabilités non contre-factualisables. Les sujets attribuent à l'expérimentateur les mêmes préférences, et conjecturent par conséquent que ce qui l'intéresse, c'est leur capacité à construire des explications causales, ou à défaut, à évaluer des probabilités contrefactualisables, dans la mesure où les données du problème n'imposent pas clairement l'interprétation non contrefactualisable.

Or, dans le problème des taxis (expérience V) par exemple, le sujet n'a pas de raison de considérer la proportion de taxis bleus comme contrefactualisable, contrairement au taux de fiabilité du témoin. Refusant à juste titre de mélanger les deux sortes de probabilité, il interprète alors la question comme portant sur des probabilités contrefactualisables, avec absence d'information sur la probabilité contrefactualisable d'apparition d'un taxi bleu ; il adopte donc pour celle-ci la valeur 0,5, puis applique correctement le raisonnement probabiliste et obtient la réponse correcte dans l'interprétation considérée.

Cette « préférence causale » que Cohen attribue aux sujets lui permet d'expliquer d'autres « erreurs » : dans l'expérience VII, par exemple, il conjecture que les sujets pensent qu'on leur demande si la taille de l'hôpital a une influence causale sur la répartition en sexes des naissances ; de même, dans l'expérience II, c'est sur la cohérence causale d'un scénario que les sujets s'imagineraient être mis à l'épreuve. Et justement, lorsque aucune information interprétable causalement n'est disponible, les sujets semblent fort capables de respecter les principes probabilistes de la conjonction et de la disjonction³⁶.

Ces principes apparaissent du reste à Cohen si solidaires de la notion même de probabilité (de même d'ailleurs que la formule de Bayes) qu'en tout état de cause, si on lui prouvait que les sujets sont incapables d'appliquer de tels principes, il lui semblerait préférable de dire qu'ils ne possèdent pas le concept de probabilité, ou qu'ils ne saisissent pas que c'est de ce concept qu'il est question dans les problèmes qu'on leur soumet. On retrouve là une des difficultés centrales que soulèvent les expériences sur le raisonnement déductif³⁷.

5. DEDUCTION ET INDUCTION, LES DEUX MAMELLES DU RAISONNEMENT ? QUELQUES ENSEIGNEMENTS TIRES DE LEUR TRAITEMENT PAR LA PSYCHOLOGIE

On a signalé plus haut le retentissement particulier qu'a connu le programme de *heuristics and biases*³⁸. Il s'explique d'au moins deux façons. Les examiner nous permettra de saisir très rapidement les différences profondes qui séparent les deux ordres de phénomènes, l'argument démonstratif et l'argument non démonstratif.

La première raison de la notoriété supérieure du programme de Tversky et Kahneman est qu'en matière de jugement en situation d'incertitude aucune norme aussi solide que la logique d'Aristote et de Frege ne vient baliser le terrain, que ce soit pour fixer un cadre de concepts et de termes, pour dicter les méthodes et déterminer les réponses, ou enfin pour orienter l'interprétation des expériences de psychologie. La logique s'impose doublement, en tout cas dans un premier temps : elle fournit l'interprétation privilégiée des mots logiques,

³⁶ Ici encore, ce sont des expériences datant de l'époque « optimiste » (1966) que Cohen invoque (*op. cit.* p. 186, note 43).

³⁷ La discussion est en fait plus compliquée que cela : le non-respect par le sujet des lois d'usage d'un concept n'entraîne pas le verdict de non-possession par le sujet du concept, mais la disjonction de ce verdict et de celui de non-possession du concept *par le théoricien*. S'ouvre ici le débat sur l'« équilibre réflexif » entre théorie et intuitions pré-théoriques, sur lequel Cohen fait fond plus particulièrement dans son article de 1981, et d'une manière que Stich (1990) conteste avec force.

³⁸ Gigerenzer (1991) mentionne de vastes secteurs d'activité influencés par ce programme : droit, économie, médecine, gestion, etc. Le grand public a été informé également : Lopes (1991) fait référence par exemple à l'article vedette d'un numéro « récent » de *Newsweek*. En revanche, l'article de *l'Economist* du 4 juillet 1992 reflète une relative « remontée » des recherches sur le raisonnement déductif, présentées sur le même plan que les travaux sur le jugement incertain.

elle est dotée de fondements solides ; elle occupe de la sorte une place pratiquement inexpugnable dans ce qu'on pourrait appeler l'« intuition instruite ». Aussi les résultats empiriques de la psychologie de la déduction se sont-ils heurtés à des convictions suffisamment fortes pour les empêcher d'être pris, sinon au sérieux, du moins pour argent comptant. Notons que la logique opère à deux niveaux : en tant que norme officielle du raisonnement déductif, puis en tant que « pilier » de notre régime cognitif « instruit ». La psychologie de la déduction se trouve ainsi prise en tenaille, si l'on peut dire, par deux corps d'armée sous un même commandement.

La psychologie du jugement en situation d'incertitude n'a face à elle qu'une seule troupe : la théorie mathématique des probabilités en tant qu'interprète officiel des concepts et calculs relatifs à l'incertain. Rien dans l'« intuition instruite » ne vient en quelque sorte redoubler les certitudes offertes par la théorie mathématique : devant un problème de chances, même fort simple, que faisons-nous ? Le langage naturel ne nous a pas préparés à nous en saisir, aucune voie de résolution n'y est en quelque sorte frayée ; les mots autant que les procédures font totalement défaut — même la notion d'ordre de grandeur est pratiquement inexprimable, en tout cas elle n'est pas manipulable dans la langue ordinaire. Nous n'avons donc pas d'autre choix que d'appliquer le calcul des probabilités en priant le ciel de ne pas tomber dans l'un des pièges redoutables que cette théorie nous tend, comme nous le savons d'expérience.

Cette relative fragilité de la norme, dans un premier temps, a formidablement contribué au crédit accordé aux thèses de *heuristics and biases*. Lorsque le psychologue met en évidence un écart considérable et systématique entre les résultats prescrits par l'application de la théorie des probabilités et les performances des sujets, on n'a à lui opposer rien qui redouble, dans l'intuition, l'évidence mathématique : comment le psychologue n'aurait-il pas raison, et que pourraient montrer ses expériences, sinon que les gens sont incapables ne serait-ce que d'approcher la vérité en matière de jugements quantitatifs d'incertitude ? La situation ne se retourne que dans un deuxième temps, lorsque l'on s'avise enfin de s'interroger sur le statut de la « vérité » brandie par le psychologue, et que fournit selon lui la théorie mathématique. Mais les premières thèses auront occupé le devant de la scène près de vingt ans, et le problème désormais est posé en grand.

La seconde raison de la notoriété des travaux psychologiques sur l'incertain tient à l'importance accordée par beaucoup à la fonction cognitive qu'ils sont censés mesurer. Autant, en ces temps de « nouvelle alliance » antilogiciste, on peut être tenté d'estimer que la déduction est bien peu de chose, autant le rôle que jouent les jugements de probabilité dans la conduite rationnelle apparaît capital : tout agent économique, tout praticien des arts et des techniques, tout enquêteur, tout magistrat ou juré, tout responsable politique, etc. exerce à longueur de journée son jugement sur les chances comparées d'événements incertains. Peu d'informations sur les aptitudes cognitives réelles peuvent donc rivaliser en importance avec celles que le programme *heuristics and biases* prétend nous apporter. Nous ne serions pas plus étonnés et inquiets, sans doute, si les psychologues et neurophysiologistes de la perception et du mouvement venaient nous apprendre que l'homme est incapable d'estimer correctement la vitesse et la trajectoire d'objets tels que les automobiles.

De manière plus générale, les résultats du programme *heuristics and biases* paraissent affecter notre conception de la rationalité humaine réelle davantage que les recherches sur la déduction parce que, comme l'annonce son objet officiel (le jugement en situation d'incertitude), elles touchent au jugement, qu'une certaine sagesse place depuis toujours³⁹ (bien antérieurement, donc, à la vague antilogiciste actuelle) très au-dessus de la déduction. Relayés et amplifiés par la psychologie sociale, les résultats négatifs de ce programme ont donné une nouvelle jeunesse, et une caution scientifique, au vieux pessimisme philosophique sur les limites de l'intelligence humaine.

Ces raisons contribuent sans doute à expliquer l'inégalité des traitements réservés aux deux grands courants de recherche sur le raisonnement. Mais elles s'alimentent elles-mêmes à une conception erronée des phénomènes étudiés.

Cette conception, nous l'appellerons, pour faire court, « symétrique » : déduction et induction, ou encore raisonnement déductif et raisonnement probabiliste (ou incertain) sont

³⁹ Georges Pompidou ne disait-il pas que la première qualité d'un homme d'Etat est le jugement ? On croit plus en Maigret qui y va au *jugé* qu'en ce logicien de Sherlock Holmes.

des facultés qui existent côte à côte, elles sont sur le même plan, comme labourage et pâturage sont les deux mamelles de la France. Ou encore, comme la logique (mathématique) et la théorie (mathématique) des probabilités sont deux branches des mathématiques.

Puisque dans les présentations par les psychologues tout nous y invite, et qu'*a contrario* rien ne nous en dissuade, faisons nôtre, provisoirement, cette conception, et posons, à propos du jugement incertain, les questions que nous avons posées à propos de la déduction (au § 3).

En premier lieu, au rapport entre logique et raisonnement (déductif) doit correspondre le rapport entre probabilités et raisonnement en situation d'incertitude. Se pose alors la question du sujet dont les capacités en matière d'analogie sont l'objet des recherches : ce doit être l'analogue, pour l'induction, du « sujet logiquement naïf » (SLN), donc un « sujet probabilistiquement naïf » (SPN). Or, on s'en souvient, nous avons été amené à attribuer au SLN une maîtrise pré-théorique de certaines notions logiques, à défaut desquelles il ne serait pas capable de saisir les problèmes logiques qu'on lui soumet. La question majeure à laquelle nous devons répondre est celle-ci : de quelles notions probabilistes pré-théoriques devons-nous créditer le SPN pour qu'il soit en mesure de saisir les questions que lui posent les psychologues ? Sans doute celle de « probabilité subjective » — mais qu'est-ce que cela ? Aucune réponse ne semble s'imposer. Mais poursuivons. En matière de déduction, nous avons tenté de préciser la nature de l'aptitude psychologique fondamentale qu'il s'agissait de caractériser. Il fallait en tout cas qu'elle soit élémentaire. Quelle est l'aptitude élémentaire que nous voulons cerner chez le SPN ?

Nous pourrions être tentés de répondre : la mesure de fréquences ou de corrélations à partir d'observations, mais nous avons vu que d'une part cette aptitude, à l'estime des psychologues, semble bien développée chez l'homme, et que d'autre part ce n'est pas elle qui est en question dans les expériences qui nous occupent. Nous devons donc nous tourner vers les *relations* entre fréquences (ou probabilités, ou chances, ou propensions, etc.), et à défaut d'une « théorie naïve des probabilités subjectives », si conjecturale qu'elle soit, nous n'avons d'autre choix qu'entre des fragments plus ou moins étendus de la théorie classique des probabilités⁴⁰. Or quel fragment s'impose-t-il ? Rien n'indique même qu'il faille le prendre dans l'axiomatique de Kolmogorov. D'abord cette axiomatique n'a aucune prétention à être « naturelle » au sens de la déduction naturelle en logique mathématique, et il n'existe pas (on peut même douter qu'il existera jamais) de théorie mathématique « naturelle » des probabilités ; de plus, il semble qu'une notion que l'on aimerait pouvoir attribuer au SPN est celle de probabilité conditionnelle — or elle ne figure pas en tant que telle dans l'axiomatique de Kolmogorov ; enfin, le problème de l'omniscience logique se pose : quelles conséquences logiques de l'axiomatique « native » ou naïve doivent-elles être incluses dans l'aptitude étudiée ? Naturellement, c'est là, en un sens, toute la question *empirique et scientifique* qui est posée, et nous n'avons pas à nous étonner de l'absence d'une réponse toute faite. Ce qui est gênant, en revanche, c'est la difficulté qu'il y a à imaginer si vaguement que ce soit *ce que pourrait être* une réponse plausible, sachant que nous voulons cerner une aptitude *élémentaire*. Une tension semble exister entre les ressources disponibles (la théorie mathématique) et l'objectif (caractériser une aptitude cognitive élémentaire)⁴¹.

Venons-en à la question des niveaux : nous l'avons posée, pour la déduction, par le biais du « système X » que semblaient rechercher les psychologues, système dont la nature et la fonction demandaient à être précisées.

La question de la fonction se pose avec une acuité toute particulière pour le jugement incertain. La situation dialectique est en effet très différente dans ce cas. En matière déductive, les compétences naturelles de l'homme ne sont pas nulles, même si elles sont prises en défaut avec régularité ; au contraire, en matière d'évaluation de résultats incertains, il n'y a pour ainsi dire, dans le programme de Tversky et Kahneman, que des défauts. Ce serait donc à l'absence d'un système Y, postulé ou espéré, qu'il permettrait de conclure, et à la présence d'un système Z (les heuristiques) qui n'accomplit pas ce que Y devait accomplir,

⁴⁰ Je ne crois pas que la considération des théories non classiques actuellement disponibles modifie radicalement la situation que nous examinons ici, mais ce n'est peut-être que l'effet de mon ignorance en la matière.

⁴¹ Un rapprochement avec la théorie de la vision pourrait être instructif : dans ce contexte-là, on n'hésite pas à créditer des composantes du système nerveux de la capacité d'exécuter des fonctions mathématiques fort complexes.

mais accomplit *autre chose*. Comme Kahneman et Tversky, Lopes (*loc. cit.*) et Gigerenzer (*loc. cit.*) le soulignent, restent fort vagues sur ce Z (sur la fonction des heuristiques), nous sommes en présence d'un problème de recherche inverse caractérisé (pour la déduction, c'est l'épistémologue soupçonneux qui le pressent ; ici c'est patent). On pourrait objecter qu'il est hâtif de conclure des résultats négatifs de *heuristics and biases* qu'il n'y a rien dans le système cognitif humain qui fasse ce que le système Y devrait faire : Gigerenzer ne montre-t-il pas que lorsqu'on évite d'induire le sujet en erreur, et qu'on déclenche chez lui une démarche fréquentiste, il produit des réponses correctes ? Sans doute, mais ce n'est pas le même genre de problème qui est posé. Il n'est pas nécessaire d'argumenter dans le détail sur les concepts engagés : il suffit de remarquer qu'il serait *absurde* de postuler une quasi-équivalence au niveau d'une « structure profonde » du problème de formulations délibérément différentes en « surface », alors même que cette différence a des effets dans les réponses. En outre, il semble plausible que les énoncés de Gigerenzer déclenchent une réaction de type combinatoire qui n'a aucun rapport avec l'évaluation que cherchent à mesurer Kahneman et Tversky. (Le mathématicien serait enclin à estimer que Gigerenzer oblige ses sujets à faire des mathématiques élémentaires, et que le fait qu'elles aient pour objet des dénombrements de cas est sans pertinence ; autrement dit, il se pourrait que le système cognitif G — s'il existe — mis à l'épreuve par Gigerenzer soit celui — s'il existe — qui sous-tend l'aptitude mathématique *en général* !) Enfin, Osherson, par exemple, conjecture qu'il existe non pas un seul système, mais plusieurs, dont l'un serait un bayésianisme mental qu'il importerait d'encourager. Mais comment séparer la contribution des différents systèmes, à moins qu'ils ne forment collectivement un super-système, ou à moins que l'on puisse caractériser indépendamment les aptitudes qu'ils sous-tendent respectivement ? Bref, de quelque côté qu'on se tourne, le problème inverse se pose avec force.

Hasardons-nous alors à demander à quel niveau se situerait le système — Y, Z ou G — qu'il s'agit de cerner grâce aux expériences des psychologues. Si nous nous en tenons à la conception symétrique, il devrait prendre place entre une faculté primitive de saisie de l'incertain, homologue de l'*Urlogique*, et une aptitude de résolution de problèmes d'incertitude. Or que doit contenir cette faculté primitive ? D'abord, on l'a vu, un (ou des) concept(s) précurseur(s) de la probabilité, de l'incertitude, des chances, des propensions, etc. Ensuite, rien de moins que l'*Urlogique* elle-même. Comment, en effet, se passer du concept de la conjonction (ou du moins d'un précurseur), pour penser les chances d'une conjonction (que Linda soit employée de banque *et* militante), du concept de la négation pour penser les chances d'un non-événement, et surtout du concept de conséquence pour penser les relations logiques qui s'établissent entre relations probabilistes ?

A ce point de la discussion, il serait futile de maintenir la fiction de la symétrie. La logique, ou son homologue psychologique — une capacité déductive de base — est un préalable à tout raisonnement, qu'il comporte ou non des arguments non démonstratifs. Reasonner sur l'incertitude, en tant que c'est *raisonner*, n'est pas différent à cet égard de raisonner sur la tâche de Wason, sur les configurations spatiales, sur les relations temporelles, etc. (On opposerait la même objection, si c'était le lieu de le faire, à la manière dont Johnson-Laird traite le « raisonnement spatial » comme s'il était une espèce de raisonnement *à côté* et sur le même plan que le raisonnement propositionnel ou syllogistique.) Bref, il y a de la logique dans tout raisonnement, parce qu'il n'existe pas de raisonnement, quel qu'en soit le contenu, qui n'implique pas de manipulation de propositions complexes dont les articulations sont les chevilles logiques au sens classique.

Inversement, la plupart des raisonnements, y compris ceux qui ne concernent ni de près ni de loin des probabilités, des ordres de grandeur, ou toute espèce d'évaluation d'incertitude, reposent sur une capacité d'évaluation de données incertaines : la pertinence et la fiabilité des prémisses, les chances d'aboutir à une conclusion acceptable et utile, la fiabilité des inférences elles-mêmes, tout cela appelle des évaluations non démonstratives.

Prenons alors un peu de hauteur. Les sciences cognitives fournissent d'une part un cadre très général pour mener l'enquête sur les capacités de l'esprit, et d'autre part une méthode qui, lorsqu'elle aboutit à la caractérisation d'une aptitude particulière, semble, en tout cas jusqu'à présent, devoir son succès au fait que cette aptitude possède une identité naturelle, qu'elle est naturellement isolée, ou encore *modulaire*, dans la terminologie de Fodor (1983). Ce n'est pas seulement qu'il se trouve que la liste de leurs succès est incluse dans la liste des aptitudes modulaires : en l'absence de modularité, il semble difficile aux sciences cognitives

telles que nous les connaissons de même *poser des questions* susceptibles de recevoir une réponse : la modularité apparaît comme une condition de scientificité, un peu comme un critère de démarcation, au sein de la cognition, entre question scientifique et question qui ne l'est pas, ou pas encore. Adopter ce point de vue, ce n'est pas dire que d'autres questions soient dépourvues de sens ; ce n'est pas non plus se rallier à la caractérisation détaillée que Fodor propose des capacités modulaires, ni à sa conception de la distinction (exhaustive) entre processus modulaires et processus centraux⁴². C'est seulement (dans un premier temps du moins) se doter d'une heuristique. Sur le sujet qui nous occupe, cela consiste à se demander, parmi les diverses capacités évoquées dans la discussion, lesquelles ont des chances de donner prise à la méthode expérimentale, et à chercher des candidats parmi celles qui présentent de la manière la plus nette certains traits associés à la modularité, tels que le caractère irrépressible, automatique et rapide, l'isolement informationnel, l'ouverture limitée aux processus centraux, la spécialisation (*domain specificity*) ou encore la forme caractéristique des pannes.

Dans ce que ce crible retient, on ne distingue d'abord que l'*Urlogique* et peut-être la capacité logique de base (CDB). Il laisse filer, naturellement, résolution logique (RL) et résolution de problèmes d'incertitude. Reste la question de ce qui pourrait jouer, du côté de l'incertain, le rôle de UL et de CDB.

On peut penser en effet que des notions telles que le couple probable/improbable et certaines relations d'implication partielle ou d'ordre partiel entre hypothèses ou propositions (*assumptions*) plus ou moins probables font partie d'un équipement de base au même titre que les précurseurs des constantes logiques et de la relation de conséquence logique. Tout cela constituerait un module *Urprobabilité* UP (sans rapport direct, bien entendu, avec une théorie des probabilités même imparfaite ou incomplète) analogue de UL. Un deuxième module aurait pour fonction de proposer aux processus centraux certains résultats concernant des propositions probables particulières et des relations entre ces propositions : ces propositions et relations seraient fournies au module en question par UP, lequel aurait reçu en entrée quelque chose comme les données d'un problème de jugement en situation d'incertitude. (Évitons de spéculer sur la manière dont UL intervient : est-ce en parallèle avec UP, ou avant ? Il importe seulement de ne pas oublier qu'il intervient de quelque façon.) Ce deuxième module, appelons-le *évaluateur automatique de l'incertain* (EAI) — pour éviter quelque chose comme *capacité probabiliste de base* qui risquerait de nous faire pencher vers une théorie naïve des probabilités au sens de la théorie mathématique des probabilités, ce qu'il est méthodologiquement indispensable d'éviter. Qu'allons-nous mettre dans EAI ? Peut-être bien les heuristiques de Tversky et Kahneman, ou quelque chose d'approchant⁴³. Dans ce cas, la théorie des probabilités serait complètement écartée d'un domaine où elle n'aurait peut-être jamais dû occuper une place importante. Après tout, l'expérience XII (où il s'agit d'estimer la factorielle de 8) est peut-être moins marginale qu'il n'apparaît dans un contexte dans lequel on en est venu à parler de jugement probabiliste ou statistique pour désigner le domaine dans son ensemble. Si cette voie devait être poursuivie, elle ferait droit, d'une certaine manière, à la fois aux intuitions de Tversky et Kahneman et à celles de Gigerenzer et de Cohen.

*

Celui qui attendait de la psychologie du raisonnement des bénéfices immédiats en matière de modélisation du raisonnement, au sens où on peut l'entendre en informatique et en intelligence artificielle, celui-là ne peut qu'être déçu. Il ne voit poindre ni de bonnes raisons de se détourner résolument de la logique, ni les prémices d'une autre logique.

Au contraire, la logique classique semble plus indéradicable que jamais : elle intervient avant même que ne se déclenchent des stratégies de résolution, et elle conditionne ce

⁴² Sur cette question, voir Sperber (1994) ; Benny Shanon était allé dans le même sens (introduire de la modularité dans les processus centraux) dès 1985 environ.

⁴³ Il faudrait également examiner les « modèles mentaux probabilistes » de Gigerenzer *et al.* (1991), ce que je n'ai pas eu le loisir de faire.

déclenchement. Primaire, la logique est essentiellement *triviale*⁴⁴ : les deux vont ensemble (et contribuent sans doute conjointement à expliquer la difficulté d'une *théorie* de la logique). Une première leçon semble donc se dégager, purement négative : ne pas chercher à remplacer la logique classique.

Indéracinable, la logique est radicalement insuffisante pour construire tout raisonnement quel qu'il soit. Non seulement, comme on le sait depuis longtemps, parce qu'elle est permissive et doit donc être complétée par un dispositif de pilotage, mais aussi, comme le montre Harman, parce qu'elle est indifférente au poids et à la pertinence des propositions, hypothèses ou conclusions virtuelles, entre lesquelles elle ne fait qu'établir un certain type de relations, relations qui peuvent entrer en conflit interne, (contradiction logique ou externe, tension avec des croyances ou contraintes d'arrière-plan). Une deuxième leçon, non moins négative que la première, serait qu'il est vain d'espérer modéliser le raisonnement, *même déductif*, à l'aide seulement de la logique classique et de stratégies de pilotage.

Enfin, puisque tout raisonnement fait appel à des arguments non démonstratifs, il est essentiel de comprendre ce qu'est un tel argument ; or, et c'est une troisième leçon négative, ce n'est pas la théorie des probabilités qui semble pouvoir nous y aider.

Mais il y a peut-être des bénéfices indirects à retirer de l'analyse des tentatives des psychologues. En matière de logique, d'abord, l'inamovibilité de la logique classique est compatible avec son enrichissement, et avec une multiplicité de tels enrichissements, qui seraient autant de modules placés en aval de la logique primaire ou *Urlogique*. Cette disposition autorise notamment un dédoublement de niveaux, dans lequel l'*Urlogique* serait reflétée dans un module en aval, son image réfléchie entrant dans ce module en interaction avec d'autres primitives, propres au module. La question de savoir quel statut donner à ces autres primitives — sont-elles de nature logique, sont-elles de l'ordre d'axiomes spécifiques, sont-elles de l'ordre du « régime inférentiel » (au sens de Barwise 1986 et 1987) qui conditionnerait la nature du raisonnement ? — tout cela est ouvert : en plaçant la logique classique dans une sorte de sanctuaire, on gagne une liberté à peu près totale dans la conception de systèmes cognitifs de raisonnement, naturels ou artificiels — c'est un autre bénéfice indirect du travail des psychologues. De plus, c'est le module aval qui va déterminer la « coloration » du système de raisonnement dans son ensemble ; il peut aller jusqu'à masquer la nature logique du module amont. On peut donc à la fois faire droit au rôle de la logique et militer contre l'idée d'un raisonnement « essentiellement » logique. Enfin, rien ne s'oppose à l'idée d'une pluralité des raisonnements.

S'il y a une idée vers laquelle les efforts des psychologues, des informaticiens, des logiciens, des philosophes et des pragmaticiens semblent converger, c'est que nous sommes encore très ignorants sur la nature du raisonnement. En particulier, nous prenons conscience du rôle clé de l'inférence non démonstrative, non seulement, en grand, comme une *sorte* ou un *domaine* de raisonnement, mais aussi, en petit, comme une *composante* intime de tout raisonnement. Or ce qu'est une inférence non démonstrative, nous n'en savons encore pas grand chose.

BIBLIOGRAPHIE

- Andler, D. (1994), « A quoi pensent les sciences cognitives ? », *Raison présente* 109, 29-50.
 Bar-Hillel, M. (1980), « The base-rate fallacy in probability judgments », *Acta Psychologica* 44, 211-33.
 Barwise, J. (1986), « Information and circumstance », *Notre-Dame J. of Formal Logic*, 27, p. 324-338.
 Barwise, J. (1987), « Unburdening the language of thought », *Mind and Language* 2, 82-96.
 Barwise, J. & Perry, J. (1983), *Situations and Attitudes*, MIT Press, Cambridge, Mass.
 Braine, M.D.S. (1978), « On the relation between the natural logic of reasoning and standard logic », *Psychological Review*, 85, 1-21.

⁴⁴ C'est ce qu'avait fort bien vu Popper, redécouvrant en Nouvelle-Zélande, dans l'immédiat après-guerre, la déduction naturelle (cf. Popper 1948).

- Braine, M.D.S. (1990), « The « Natural Logic » approach to reasoning », in W.F. Overton, ed., *Reasoning, Necessity, and Logic : Developmental Perspectives*, Lawrence Erlbaum, Hillsdale, NJ.
- Braine, M.D.S. (1992), « Approches empiriques du langage de la pensée », in D. Andler *et al.*, dir., *Epistémologie et cognition*, P. Mardaga, Liège.
- Braine, M.D.S., Reiser, B.J. & Rumain B. (1984), « Some empirical justifications for a theory of natural propositional logic », in G.H. Bower (ed.), *The Psychology of Learning and Motivation*, vol. 18, Academic Press, New York.
- Casscells, W., Schoenberger, A. & Grayboys, T. (1978), « Interpretation by physicians of clinical laboratory results », *New England Journal of Medicine* 299, 299-300.
- Cheng, P. & Holyoak, K.J. (1985), « Pragmatic reasoning schemas », *Cognitive Psychology*, 17, 391-416.
- Cherniak, Ch. (1986), *Minimal Rationality*, MIT Press, Cambridge, Mass.
- Cohen, L. J. (1970), *The Implications of Induction*, Methuen, Londres.
- Cohen, L. J. (1977), *The Probable and the Provable*, Clarendon Press, Oxford.
- Cohen, L. J. (1981), « Can human irrationality be experimentally demonstrated ? », with Open Peer Commentary, *Behavioral and Brain Sciences* 4, 317-70.
- Cohen, L. J. (1982), « Are people programmed to commit fallacies ? Further thoughts about the interpretation of experimental data on probability judgements », *Journal for the Theory of Social Behaviour* 12, 251-74.
- Cohen, L. J. (1986), *The Dialogue of Reason. An Analysis of Analytical Philosophy*, Clarendon Press, Oxford : §§ 16-19.
- Cohen, L.J. (1989), *An Introduction to the Philosophy of Induction and Probability*, Clarendon Press, Oxford.
- Cosmides, L. (1989), « The logic of social exchange : Has natural selection shaped how humans reason ? », *Cognition* 31, 187-276.
- Cosmides, L. & Tooby, J. (1990), « Is the mind a frequentist ? », communication présentée au Second Annual Meeting of the Human Behavior and Evolution Society, Los Angeles.
- Evans, J. St. B.T. (1989), *Bias in human reasoning : Causes and consequences*, Lawrence Erlbaum, Hillsdale, NJ.
- Evans, J. St. B.T. (1991), « Theories of human reasoning : The fragmented state of the art », *Theory and Psychology* 1, 83-105.
- Evans, J. St. B.T. (1994), « Bias and rationality », in Manktelow, K.I. & Over, D.E., eds., *Rationality*, Routledge, Londres.
- Evans, J. St. B.T. (1994a), « Bias in thinking and judgement », in Keane, M.T. & Gilhooly, K.J., eds., *Advances in the Psychology of Thinking* (vol. 1), Harvester-Wheatsheaf.
- Evans, J. St. B.T., Newstead, S.E., & Byrne, R. M. J. (1993), *Human Reasoning*, Lawrence Erlbaum, Hillsdale, NJ.
- Fischhoff, B. (1988) « Judgment and decision making », in Sternberg, R. & Smith, E., eds., *The Psychology of Human Thought*, Cambridge University Press, Cambridge., 153-187.
- Fodor, Jerry A. (1983). *The Modularity of Mind*, MIT Press, Cambridge, Mass. Trad. fr. *La Modularité de l'esprit*, Minuit, Paris, 1986.
- Gamut, L.T.F. (1991), *Introduction to Logic* (vol. 1 : Logic, Language, and Meaning, vol. 2 : Intensional Logic and Logical Grammar), Chicago University Press, Chicago.
- Gärdenfors, P. (1988), *Knowledge in Flux*, MIT Press, Cambridge, Mass.
- Gigerenzer, Gerd (1991), « How to make cognitive illusions disappear : Beyond "Heuristics and Biases" », in Stroebe, W. & Hewstone M., *European Review of Social Psychology*, vol. 2, John Wiley & Sons, Chichester, 83-115.
- Gigerenzer, Gerd, Hoffrage, Ulrich & Kleinbölting, Heinz (1991), « Probabilistic mental models : A Brunswickian theory of confidence », *Psychological Review* 98, 506-528.
- Gigerenzer, G. & Hug, K. (1992), « Domain-specific reasoning : Social contracts, cheating, and perspective change », *Cognition* 43, 127-71.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Krüger, L. (1989), *The Empire of Chance : How Probability Changed Science and Everyday Life*, Cambridge University Press, Cambridge.
- Gilhooly, K.J., Keane, M.T.G., Logie, R.H., Erdos, G. (1990), *Lines of Thinking : Reflections on the Psychology of Thought*, vol 1 : Representation, Reasoning, Analogy and Decision Making, John Wiley & Sons, Chichester.

- Giroto, V. (1991), « Reasoning on deontic rules : The pragmatic schemas approach », in Politzer (1991).
- Goldman, A. (1986), *Epistemology and Cognition*, Harvard University Press, Harvard.
- Grice, H.P. (1989), *Studies in the Way of Words*, Harvard University Press, Cambridge, Mass.
- Halpern, J., ed. (1986), *Theoretical Aspects of Reasoning About Knowledge*, Morgan Kaufmann, San Francisco.
- Harman, G. (1986), *Change in View*, MIT Press, Cambridge, Mass.
- Henle, M. (1962), « On the relation between logic and thinking », *Psychological Review* 69, 366-378.
- Hirschfeld, L.A. & Gelman, S.A., eds. (1994), *Mapping the Mind. Domain Specificity in Cognition and Culture*, Cambridge University Press, Cambridge.
- Holland, J.H., Holyoak, K.J., Nisbett, R.E. & Thagard, P.R. (1986), *Induction : Processes of Inference, Learning, and Discovery*, MIT Press, Cambridge, Mass.
- Holyoak, K.J. (1990), « Problem solving », in Osherson & Smith (1990).
- Johnson-Laird, P. N. (1983), *Mental models : Towards a Cognitive Science of Language, Inference, and Consciousness*. Cambridge University Press, Cambridge.
- Johnson-Laird, P. N. (1986), « Reasoning without logic », in *Reasoning and Discourse Processes*, Academic Press, Londres.
- Johnson-Laird, P. N. (1988), *The Computer and the Mind. An Introduction to Cognitive Science*, Harvard University Press, Cambridge, Mass.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1991), *Deduction*, Lawrence Erlbaum, Hillsdale, NJ.
- Johnson-Laird, P. N. & Byrne, R. M. J. (1993), « Précis of Deduction », with Open Peer Commentary, *Behavioral and Brain Sciences* 16, 323-80.
- Kahneman, D. & Tversky, A. (1972), « Subjective probability : A judgment of representativeness », *Cognitive Psychology* 3, 430-54.
- Kahneman, D. & Tversky, A. (1972a), « On the psychology of prediction », Oregon Research Institute Research Bulletin 12.
- Kahneman, D. & Tversky, A. (1973), « On the psychology of prediction », *Psychological Review* 80, 237-51.
- Kahneman, D., Slovic, P. & Tversky, A., eds. (1982), *Judgment under Uncertainty : Heuristics and Biases*, Cambridge University Press, Cambridge.
- Lopes, Lola L. (1991), « The rhetoric of irrationality », *Theory and Psychology* 1, 65-82.
- Macnamara, J. (1986), *A Border Dispute : The Place of Logic in Psychology*, MIT Press, Cambridge, Mass.
- Manktelow, K.I. & Over, D.E. (1990), *Inference and Understanding : A Philosophical and Psychological Perspective*, Routledge, Londres.
- Manktelow, K.I. & Over, D.E., eds. (1993), *Rationality*, Routledge, Londres.
- Meehl, P. & Rosen, A. (1955), « Antecedent probability and the efficiency of psychometric signs, patterns, and cutting scores », *Psychological Bulletin* 52, 194-215.
- Newell, A. (1973), « Production systems : Models of control structures », in Chase, W.G., ed., *Visual Information Processing*, Academic Press, New York.
- Newell, A., & Simon, H.A. (1972), *Human Problem Solving*, Prentice Hall, Englewood Cliffs, NJ.
- Nisbett, R. & Ross, L. (1980), *Human Inference : Strategies and Shortcomings of Social Judgment*, Prentice-Hall, Englewood Cliffs, NJ.
- Noveck, I., Lea, R.B., Davidson, G.M., & O'Brien, D.P. (1991), « Human reasoning is both logical and pragmatic », in Politzer (1991).
- Oaksford, M. & Chater, N. (1992), « Bounded rationality in taking risks and drawing inferences », *Theory and Psychology* 2, 225-230.
- Osherson, D. (1975), « Logic and models of logical thinking », dans R.J. Falmagne (ed), *Reasoning : Representation and Process*, Wiley, New York.
- Osherson, D. (1990), « Judgment », in Osherson & Smith (1990), 55-87.
- Osherson, D. & Smith, E., eds. (1990), *Thinking*, MIT Press, Cambridge, Mass.
- Peacocke, C. (1994), « Content, computation and externalism », *Mind and Language* 9.
- Politzer, G. dir. (1991), Numéro spécial « Pragmatique et psychologie du raisonnement », *Intellectica* 11 [revue de l'ARC, Association pour la recherche cognitive].
- Popper, Karl R. (1948), « The trivialization of logic ».

- Popper, Karl R. (1983), *Realism and the Aim of Science*, Hutchinson, Londres ; trad. fr. *Le réalisme et la science*, Hermann, Paris, 1990.
- Rips, L.J. (1986). « Mental Muddles », in M. Brand & M. Harnish, eds., *The Representation of Knowledge and Belief*, University of Arizona Press, Tucson, 258-86.
- Rips, L.J. (1988). « Deduction », in Sternberg, R. & Smith, E., eds., *The Psychology of Human Thought*, Cambridge University Press, Cambridge., 116-152.
- Sainsbury, M. (1991), *Logical Forms*, Blackwell, Oxford.
- Scholtz, R.W., ed. (1983), *Decision Making Under Uncertainty*, Elsevier, Amsterdam.
- Smith, E.E., Langston, Ch., & Nisbett, R. (1992), « The case for rules in reasoning », *Cognitive Science* 16, 1-40.
- Sperber, D. (1994), « The modularity of thought and the epidemiology of representations », in Hirschfeld & Gelman.
- Sperber, D., Cara, F., & Girotto, V. (1995), « Relevance Theory explains the selection task », *Cognition* 57, pp. 31-95.
- Sperber, D. & Wilson, D. (1986), *Relevance : Communication and Cognition*, Blackwell, Oxford ; tr. fr. *La pertinence : Communication et cognition*, Minuit, Paris, 1989.
- Stalnaker, R. (1984), *Inquiry*, MIT Press, Cambridge, Mass.
- Stich, S. (1990), *The Fragmentation of Reason*, MIT Press, Cambridge, Mass.
- Tversky, A. & Kahneman, D. (1971), « Belief in the law of small numbers », *Psychological Bulletin*, 76, 105-10.
- Tversky, A. & Kahneman, D. (1973), « Availability : A heuristic for judging frequency and probability », *Cognitive Psychology* 5, 207-32.
- Tversky, A. & Kahneman, D. (1974), « Judgment under Uncertainty : Heuristics and Biases », *Science* 125, 1124-31.
- Tversky, A. & Kahneman, D. (1982), « Judgments of and by representativeness », in Kahneman, Slovic & Tversky 1982, 84-98.
- Tversky, A. & Kahneman, D. (1982a), « Evidential impact of base rates », in Kahneman, Slovic & Tversky 1982, 153-160.
- Tversky, A. & Kahneman, D. (1983), « Extensional versus intuitive reasoning : The conjunction fallacy », *Psychological Review* 90, 292-315.
- Wason, P.C. (1966), « Reasoning », in B.M. Foss, ed., *New Horizons in Psychology*, Penguin, Hammondsworth.
- Wason, P.C. & Johnson-Laird, P.N. (1972), *Psychology of Reasoning : Structure and Content*, Batsford, Londres.