

LES PROCESSUS DE DECISION DES EXPERTS ET LA THEORIE DU COMPORTEMENT DU DECIDEUR

D. J. HILTON
Professeur au Groupe ESSEC

La démarche traditionnelle de la psychologie expérimentale consiste à contrôler les conditions de l'expérience afin d'exclure tout facteur autre que ceux spécifiquement étudiés.

Mais beaucoup d'études sur les processus de décision ne portent que sur des problèmes artificiels et excluent l'impact des connaissances déjà présentes chez le décideur.

Cette démarche gêne l'étude des processus par lesquels l'expert impose ses structures à un problème, lors de la recherche de ses données et de la génération de ses hypothèses.

Dans ce papier, une nouvelle approche est proposée ; elle a pour objectif d'étudier les processus cognitifs de l'expert dans le contexte rigoureux du contrôle expérimental.

1 - Introduction

Cet article présente tout d'abord les principales conclusions des recherches effectuées au cours des 25 dernières années dans le domaine de la théorie du comportement du décideur et plus spécifiquement sur les processus de décision de l'expert. La méthodologie expérimentale utilisée dans ces recherches présente certains inconvénients. La méthodologie proposée en alternative est illustrée par une analyse d'un cas de diagnostic médical.

2 - La démarche classique en théorie du comportement du décideur

L'individu de façon générale, analyse les problèmes qui se présentent à lui sur la base de connaissances formelles antérieures. Ainsi, dans le diagnostic d'un cancer chacun peut estimer la probabilité que le patient souffre d'un cancer sur la base de connaissances du type de celles présentées dans le tableau 1.

Tableau 1 - Données de base d'un diagnostic

Question : si un homme, âgé de 25 ans a passé le test qui diagnostique le cancer et que le test soit positif, quelle est la probabilité qu'il ait le cancer ?

On sait par ailleurs que la fiabilité du test est connue :

- quand le cancer est présent, le test est positif sans risque d'erreur dans 98 % des cas soit une erreur possible de 2 %,
- quand le cancer est absent, le test est négatif dans 95 %, soit un risque d'erreur de 5 %.

et que la distribution a priori du cancer parmi les hommes de 20 à 30 ans est de 1 pour 10.000

Spontanément, l'individu interrogé focalisera sur le résultat brut du test et aura tendance à répondre que la probabilité que le patient souffre effectivement d'un cancer est supérieure à 70 %. En fait, la vraie réponse se rapproche de 2 %, comme le montrent les calculs effectués dans le tableau 2 (page suivante).

Ce problème, très connu des statisticiens, est malheureusement trop peu connu des experts sur le terrain et de nombreuses erreurs de diagnostic sont ainsi commises car la distribution a priori de la maladie ("base-rate fallacy") n'est pas prise en compte. Cette "négligence" de praticiens se retrouve même dans des systèmes experts aussi réputés que MYCIN par exemple...

**Tableau 2 -
Probabilité effective de présence du cancer**

DISTRIBUTION A PRIORI		TEST SENSIBLE A 98 %
CANCER	1	.98 "positif" .02 "négatif" 95 % spécifique
PAS DE CANCER	9.999	499.95 "positif" 9 499.05 "négatif"
→ 500.93 dits "positif" → 8488.07 dits "négatif"		
Conclusion :		
Probabilité que "positif" signifie effectivement "cancer" : $\frac{0.98}{500.93} \approx 0.02$		

La question qui se pose est de savoir comment les experts en arrivent ainsi à démontrer ce type de faiblesse dans leurs propres domaines d'expertise. Il s'agit d'un domaine sensible puisqu'évidemment, si nous sommes malades, nous nous tournerons vers notre médecin pour connaître sa réaction face aux résultats des tests auquel nous nous sommes soumis. Pour résumer une vingtaine d'années de travaux de recherche dans le domaine de la décision, disons que les experts savent bien les facteurs à prendre en compte dans une décision. Mais un décalage peut être noté entre les connaissances de l'expert et l'application de ses connaissances. Ce processus de perfectionnement de ses propres connaissances est connu comme étant le phénomène de "bootstrapping" (Dawes, 1979), étudié en psychologie cognitive. Les chercheurs ont sur ce point noté que les procédures que l'expert lui-même utilise risquent parfois de le conduire à mal intégrer les informations pertinentes dans ses décisions. A plusieurs reprises, on a remarqué qu'un modèle linéaire bâti sur les connaissances d'un expert montre une performance dans la prise de décision supérieure de l'expert lui-même. Ainsi, on peut perfectionner la performance de l'expert par la mise de ses connaissances des facteurs importants dans le cadre d'un modèle linéaire subjectif. Ce modèle linéaire subjectif peut intégrer les données d'un cas d'une façon pondérée ou non pondérée, mais en tout cas, le résultat est presque toujours le même — le modèle linéaire est plus performant que l'expert qui lui a approvisionné ses connaissances.

Russo et Shoemaker (1991) ont comparé la fiabilité des prédictions formulées de manière intuitive par les experts, avec un modèle "bootstrapped" bâti sur la base de leurs indications (un modèle linéaire subjectif) et avec un modèle linéaire objectif validé par des résultats de décision. La première constatation est que dans tous les cas étudiés, les ordinateurs sont dans tous les cas plus performants que les experts (tableau 3).

Tableau 3 - Les études "bootstrapping"

Genres de jugements donnés aux experts	Prédiction intuitive	Modèle "bootstrapped"	Modèle objectif
• Performance scolaire des étudiants du 3 ^e cycle	.19	.25	.54
• La prévision de temps à vivre des patients atteints de cancer	-.01	.13	.35
• Changements de prix de valeurs	.23	.29	.80
• Maladie mentale par tests de personnalité	.28	.31	.46
• Notes et attitudes en cours de psychologie	.48	.56	.62
• Faillites des sociétés par ratios financiers	.50	.53	.67
• Evaluation des étudiants des performances des professeurs	.35	.56	.91
• Performance des ven- deurs d'assurance-vie	.13	.14	.43
• Performance IQ par test Rorschach	.47	.51	.54
Moyen (à travers plusieurs expériences)	.33	.39	.64

La première explication de ces résultats tient au fait que l'ordinateur ne fait bien sûr jamais de faute d'attention ou de mémoire et qu'il applique les règles et les critères (fournis initialement par l'expert lui-même) plus efficacement que l'expert lui-même.

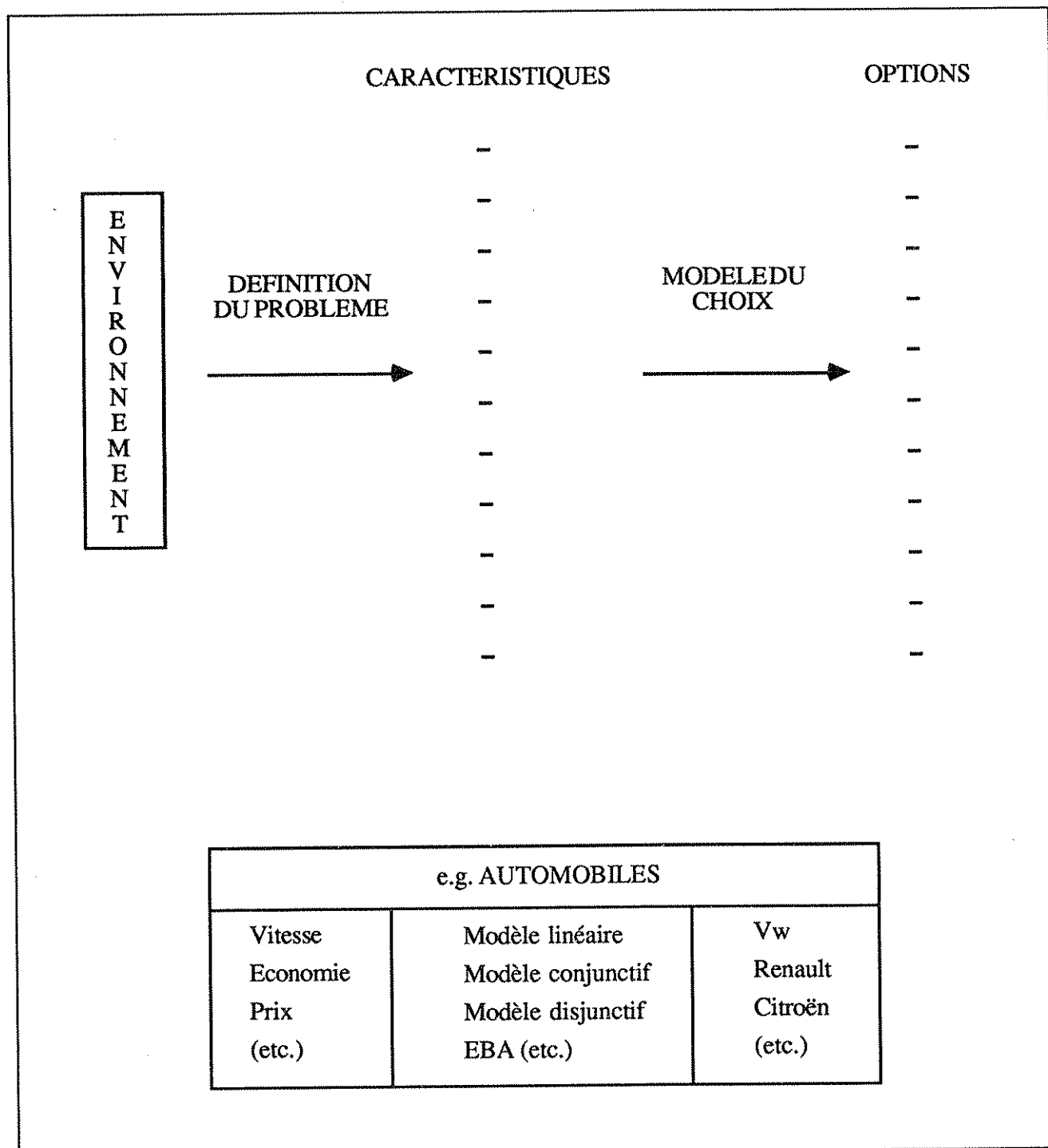
Mais le problème ne s'arrête pas là. Lorsqu'une décision est prise de manière répétée et donc que le résultat de chaque cas est connu, il est possible d'estimer les probabilités objectives d'une décision. Or le tableau 3 montre que ni la décision de l'expert, ni celle du modèle bâti sur les connaissances de l'expert (le "bootstrapped model") ne sont pas non plus optimales. Ici l'explication du décalage entre la performance des experts et la performance optimale peut résider dans le fait que les experts ont peut-être ignoré les facteurs importants, ou les ont mal estimés. Par exemple, ce phénomène a été noté dans une étude des

risques dans l'allocation du crédit aux Etats-Unis. Les faits prouvent que les risques sont plus importants avec des enseignants qu'avec des étudiants. Comme ce résultat ne correspond pas aux intuitions des experts, il existe un décalage entre les décisions prises par un juge et un logiciel qui utilise les valeurs objectives pour chaque facteur.

Finalement, même si le décideur connaît les facteurs

adéquats, il ne les prend pas toujours bien en compte lors du diagnostic. Cela tient au degré de structuration des problèmes. En effet, les processus présentés précédemment s'appliquent à des problèmes déjà structurés, c'est-à-dire les problèmes où toutes les caractéristiques et toutes les alternatives sont (ou peuvent être) présentes à l'esprit du décideur. La structure de la matrice décisionnelle se présente alors de la manière suivante (voir tableau 4).

Tableau 4 - Schéma de la définition et prise de décision



Mais dans de nombreuses situations réelles, les caractéristiques et les alternatives pertinentes n'existent pas de manière formelle dans l'esprit du décideur : il doit lui-même les générer. Tel est le cas des "problèmes mal structurés" (Pople, 1981). Face à ce type de décisions, le décideur n'a d'autre alternative que d'imposer ses propres structures aux problèmes qui se présentent à lui.

Puisque l'une des analogies les plus répandues est celle des décisions dans le domaine de la gestion et dans le domaine médical, observons comment travaille un médecin de famille. Si le patient se présente à lui avec une plainte telle que : "mon estomac me fait mal", cela peut signifier l'indigestion, l'alcoolisme, le cancer, ou beaucoup d'autres causes. Pour le médecin, chaque alternative émise comme hypothèse entraîne des caractéristiques pertinentes qui seront recherchées. Par exemple, l'hypothèse de l'alcoolisme l'amènera à poser des questions au patient sur ses habitudes, en particulier sur sa consommation d'alcool ; celle du cancer entraînera le médecin à vérifier si le patient a perdu du poids récemment, etc. Donc la tâche cognitive pour le décideur consiste à se souvenir des alternatives pertinentes et ensuite des caractéristiques correspondant à ces alternatives. Mais cette démarche présente un risque fondamental, celui de ne pas passer en revue toutes les alternatives et leurs caractéristiques, c'est-à-dire à ne pas construire toute la matrice décisionnelle. Ce risque est intrinsèquement celui de la mémoire et de son processus de reconnaissance : alors que la psychologie expérimentale s'attache à améliorer le processus de reconnaissance, la réalité nous apprend que la priorité en est une de mémoire.

Ainsi, en psychologie expérimentale, respecte-t-on traditionnellement les étapes suivantes :

- 1 - Toutes les caractéristiques et toutes les alternatives sont présentées à l'expert, afin de s'assurer que tous les sujets agissent dans le même environnement cognitif.
- 2 - On fait varier les caractéristiques et les alternatives présentées aux sujets, afin d'estimer l'effet de chaque facteur sur la décision émise.
- 3 - La tâche est présentée de manière structurée et simplifiée aux décideurs, dans le but d'exclure les influences extérieures qui pourraient contaminer les manipulations expérimentales. La clarté des résultats peut ainsi être obtenue.

Ces précautions permettent de s'assurer que ce sont les manipulations effectuées au cours de l'expérience qui ont produit les effets observés et en aucun cas, d'autres facteurs inconnus. La validité interne de l'expérience est ainsi garantie. Mais la question est de savoir dans quelle mesure les résultats ainsi obtenus sont représentatifs des tâches décisionnelles dans le monde réel, c'est-à-dire qu'elle est leur validité externe.

Et la confrontation de la méthode préconisée en psychologie expérimentale avec le monde réel nous montre que les mêmes facteurs qui favorisent la validité interne militent contre la validité externe. Ainsi dans la réalité quotidienne deux phénomènes se produisent. D'une part, la tâche principale de l'expert est de se construire une matrice décisionnelle par la génération des hypothèses et l'acquisition des informations au fur et à mesure de ses interrogations et des réponses qu'il obtient. D'autre part aucune caractéristique ou option n'a de l'influence sur la décision émise si le décideur n'en est pas conscient. Ainsi faudrait-il étudier les processus par lesquels chaque caractéristique ou alternative est prise en compte. Ce sont donc ici les influences extérieures à la tâche expérimentale qui nous intéressent, influences qui peuvent être l'effet de l'expertise du décideur ou ses valeurs.

Elstein, Shulman et Sprafka (1978) ont étudié la manière dont le décideur expert résout les problèmes mal structurés. En s'appuyant sur les théories de Newell et Simon (1972), ils présentèrent à des médecins experts des dossiers médicaux et enregistrèrent ensuite leurs commentaires oraux sur leur processus de décision dans les "Think-aloud protocols", (c'est-à-dire leurs verbatims ou commentaires oraux sur leur processus de décision). Leurs protocoles oraux ont ensuite été analysés.

Même si cette démarche a l'avantage d'étudier le médecin dans une situation assez naturelle, elle présente un inconvénient important, celui de la généralisation des résultats obtenus. Ainsi, en laissant le champ libre à chaque médecin de suivre sa propre démarche, les résultats obtenus ne permettront pas de comparer facilement les comportements des médecins entre eux. La monographie d'Elstein, Shulman and Sprafka (1978) s'est donc appuyée sur une analyse de contenu de trois dossiers dont le codage a été effectué par une seule personne. Il faudrait pour s'assurer de cette généralisation développer un système d'analyse du contenu assez sophistiqué, ce qui serait très coûteux en matière de ressources. Par exemple, la généralisation des résultats à l'ensemble des médecins et à l'ensemble des dossiers et un contrôle des erreurs de codage nécessiterait un certain nombre de médecins (disons un minimum de 20), de dossiers (disons un minimum de 3) et de codeurs (disons un minimum de 2). Cela ferait un minimum de 120 analyses (20 par 3 par 2) de protocoles au lieu de 3 dans l'étude d'Elstein, Shulman and Sprafka.

Dans une étude que nous avons réalisée (Weber, Böckenholt, Hilton et Wallace, 1991) le problème de la généralisation a été contourné de la manière suivante :

L'un de nous, Brian Wallace, médecin de famille exerçant au Pays de Galles, a agi comme expert. Il a dé-

veloppé trois dossiers expérimentaux à partir de trois cas réels. Les dossiers étaient montés en fonction des objectifs précis de la recherche, à savoir les facteurs qui influencent la génération des hypothèses et l'acquisition des informations. Chacun des trois dossiers suivait une certaine structure (voir tableau 5). Nous nous sommes intéressés aux problè-

mes de diagnostics pour lesquels il y avait deux possibilités d'expliquer la condition présentée par un malade. Donc nous avons varié la quantité des informations soutenant l'hypothèse de la présence de la maladie commune et bénigne et celle de la présence de la maladie rare et grave.

Tableau 5 - Structure de l'expérience sur les processus de décision des médecins
(Weber, Böckenholt, Hilton et Wallace, 1991)

		QUANTITE DES INFORMATIONS INDIQUANT MALADIE A (commune et bénigne)		
		Beaucoup	Assez	Peu
QUANTITE DES INFORMATIONS INDIQUANT MALADIE B (rare et grave)	Beaucoup			
	Assez			
	Peu			

Telles conditions risquent d'être mal diagnostiquées parce que le décideur risque de n'émettre qu'une hypothèse et d'ignorer l'autre possible. Ainsi, nous avons mesuré la conviction du décideur dans ses hypothèses, parce que nous supposons qu'il chercherait plus d'informations s'il n'avait pas confiance dans sa série actuelle d'hypothèses.

La principale conclusion fut que chaque résultat apparaissait dans chacun des trois dossiers et donnait l'impression que les résultats pouvaient refléter des processus généraux plutôt que des processus spécifiques à chaque dossier. Ainsi :

- il apparut que les médecins ont pris en compte des base-rates (distribution a priori des maladies) dans leur diagnostic. Cela veut dire que, si le patient qui présente une perte de poids est vieux et de sexe masculin, il est plus probable que le médecin diagnostiquera un cancer, alors que si le patient est jeune et de sexe féminin, il diagnostiquera l'anorexie ;
- il parut aussi évident que les médecins étaient convaincus de leurs hypothèses. Chaque dossier a

montré deux diagnostics différents : l'un impliquant une maladie commune et bénigne (comme la gastrite), l'autre une maladie rare et grave (comme l'ischémie myocardique). Comme prévu, les signes indiquant la présence de chaque maladie augmentaient la certitude que la maladie en question existait.

Cependant, ce qui était imprévu et intéressant est que la confiance des médecins dans la série des hypothèses était confortée par les symptômes existants de la maladie commune et bénigne mais qu'elle était également réduite par les signes d'apparition de la maladie rare et grave. En fait, la confiance des médecins en leur série d'hypothèses ne respectait pas les règles normatives de la probabilité.

Une possibilité est ici que les médecins ne supportent pas facilement que deux maladies puissent apparaître simultanément chez le même patient. Si cette supposition est vraie, les médecins risqueraient d'ignorer la deuxième maladie. Ce risque pourrait être

augmenté par la possibilité que les informations soutenant l'apparition de la maladie "normale", c'est-à-dire la maladie commune et bénigne leur confirme que la série d'hypothèses est complète. Cela pourrait avoir des conséquences graves, par exemple d'ignorer une maladie qui est présente. Par exemple, aux Etats-Unis, on a constaté que pour plus d'1/3 des personnes qui décédaient à la suite d'un problème au foie lié à l'alcoolisme, cette cause n'avait pas été diagnostiquée lors de leur admission à l'hôpital. Ceci montre l'importance du risque d'ignorer la présence d'une maladie rare et sévère.

En résumé, l'expérience effectuée avec les médecins a révélé un bon et un mauvais côté du processus de diagnostic. Le bon côté est que nos médecins ont utilisé la distribution a priori des maladies en faisant leur diagnostic. Le mauvais côté est qu'il semble bien possible que dans la génération des hypothèses, les médecins négligeraient d'explorer la possibilité de l'apparition d'une maladie rare et grave soit parce qu'ils sont convaincus que leur série d'hypothèses est complète, soit parce qu'ils n'admettent pas facile-

ment la possibilité qu'une deuxième maladie puisse intervenir en plus de leur premier diagnostic.

J'espère que je vous ai fait goûter un peu la "Behavioural Decision Theory", théorie qui traite du comportement du décideur. Les travaux effectués au fil des années ont fait état de nombreuses sources de faiblesse dans les performances des experts et les années à venir permettront d'approfondir nos connaissances dans ce domaine. Nos connaissances pourraient nous amener à améliorer la qualité des décisions prises par deux moyens. L'un est de faire réfléchir les experts sur leur propre processus de décision par des interventions de conseil, ou pendant leur formation en leur enseignant la théorie du décideur. L'autre est de donner aux experts des aides à la décision comme les systèmes experts. En raison de l'échec des systèmes de décisions statistiques venant du fait qu'ils ne sont pas acceptés par de nombreux experts, il sera important de bien étudier la psychologie du décideur afin de chercher les méthodes pratiques qui pourraient les aider à améliorer leurs décisions.

BIBLIOGRAPHIE

- Dawes, R.M. (1979), "The robust beauty of improper linear models in decision making, *American Psychologist*, 34, 571-582.
- Elstein, A.S., Shulman, L.S. et Sprafka, S.A. (1978), *Medical problem-solving: An analysis of clinical reasoning*, Cambridge, MA : Harvard Press.
- Newell, A. et Simon, H.A. (1972), *Human problem-solving*, Englewood Cliffs, N.J. : Prentice-Hall.
- Pople, H.E. (1981), Heuristic methods for imposing structure on ill-structured problems, in P. Szolovits (Ed.), *Artificial intelligence in Medicine*, Boulder : Westview Press.
- Russo, J.E. et Schoemaker, P.J.H. (1991), *Confident decision making*, London : Judy Piatkus.
- Weber, E.U., Böckenholt, U., Hilton, D.J. et Wallace, B.B. (1991), Clinical and base-rate information as determinants of medical diagnoses and of physicians' confidence, Unpublished manuscript, University of Chicago.

REMERCIEMENTS

Je remercie beaucoup Albert Bemmaor, Michelle Bergadaà, Anne-Marie Dumas et Véronique d'Estaintot pour leur gentillesse et leur aide dans la rédaction de ce document.