



Incertitude, causalité et décision : Le cas des risques sociaux et du risque nucléaire en particulier

Reza Lahidji

► To cite this version:

Reza Lahidji. Incertitude, causalité et décision : Le cas des risques sociaux et du risque nucléaire en particulier. Gestion et management. HEC, 2012. Français. NNT : 2012EHEC0002 . pastel-00718790

HAL Id: pastel-00718790

<https://pastel.hal.science/pastel-00718790>

Submitted on 18 Jul 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

ECOLE DOCTORALE
Sciences du Management



Gestion – Organisation
Décision – Information



ECOLE DES HAUTES ETUDES COMMERCIALES DE PARIS

Ecole Doctorale « Sciences du Management/GODI » - ED 533

Gestion Organisation Décision Information

« Incertitude, causalité et décision :
le cas des risques sociaux et du risque nucléaire en particulier »

THESE

présentée et soutenue publiquement le 29 Février 2012

en vue de l'obtention du

DOCTORAT EN SCIENCES DE GESTION

Par

Reza LAHIDJI

JURY

Président du Jury :

Monsieur Bertrand VILLENEUVE
Professeur des Universités
Université Paris Dauphine

Directeur de Recherche :

Monsieur Philippe MONGIN
Directeur de Recherche au CNRS
Professeur Affilié HDR Ecole des Hautes Etudes Commerciales

Rapporteurs :

Monsieur Denis BOUYSSOU
Directeur de Recherche au CNRS LAMSADE
Université Paris Dauphine

Monsieur Jean-Christophe VERGNAUD
Directeur de Recherche au CNRS
CES Maison des Sciences Economiques

Suffragants :

Monsieur Patrick MOMAL
Economiste, Institut de Radioprotection et de Sécurité Nucléaire

Monsieur Bertrand MUNIER
Professeur des Universités IAE de Paris 1 – Panthéon Sorbonne

Ecole des Hautes Etudes Commerciales

Le Groupe HEC Paris n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

à Nouchi, Karim, Janne, Roxane, Morgane et Navid

Remerciements

Je tiens à remercier, en premier lieu et tout particulièrement, mon directeur de thèse Philippe Mongin. Tout au long de ce travail, Philippe Mongin m'a obligé par sa patience autant que par son exigence. Pourtant, je voudrais surtout dire ma gratitude au guide, celui qui amène à entreprendre ce que l'on ne ferait pas de soi-même.

Je remercie aussi chaleureusement les chercheurs dont la fréquentation m'a enrichi et éclairé : les théoriciens de la décision Mohammed Abdellaoui, Michèle Cohen, Itzhak Gilboa, Brian Hill, Bertrand Munier et Jean-Christophe Vergnaud ; et les experts de la sûreté nucléaire Geneviève Baumont, Giovanni Bruna, Bernard Chaumont, Eric Chojnacki, François Corenwinder, Patricia Dupuy, Pascal Leverd, Frédéric Ménage, Patrick Momal, Ambroise Pascal, Ludivine Pascucci-Cahen, Frédérique Pichereau, Vincent Rebour, Emmanuel Remond, Yann Richet, Oona Scotti, Romain Suchet et Michael Tichauer.

Je remercie enfin les institutions qui m'ont soutenu et m'ont permis de mener à bien ma recherche, l'Institut de Radioprotection et de Sûreté Nucléaire et le Groupement de Recherches en Gestion à l'école des Hautes Etudes Commerciales.

Conventions de notation

La thèse comporte huit chapitres numérotés de 1 à 8, ainsi que des parties non numérotées : une introduction et une conclusion générales, une bibliographie, un glossaire des termes de l'ingénierie nucléaire et une table des matières.

Les chapitres sont subdivisés en *parties* (1.1), les parties en *sections* (1.1.1) et les sections en *sous-sections* (1.1.1.1).

Les termes inclus dans le glossaire sont indiqués par un astérisque lors de leur première occurrence dans le texte.

Nous utilisons les caractères italiques d'une part pour souligner notre propos et de l'autre pour désigner les termes en langue étrangère.

Les notations mathématiques usuelles ne font pas l'objet d'indication ou d'explication particulières. Les conventions spécifiques sont indiquées dans le corps du texte.

Introduction générale

Se protéger ensemble de l'incertain : le projet est sans doute consubstantiel à la société humaine. Mais l'Europe des Lumières l'a formulé dans des termes nouveaux, en affirmant que la protection des membres de la société était une responsabilité des gouvernants et qu'ils devaient s'en acquitter en faisant appel au savoir et à la raison. Cette exigence de prise en compte rationnelle des risques sociaux a trouvé dans l'analyse des rapports de causalité, puis dans le calcul des probabilités, les outils de sa mise en œuvre. Nous voulons retracer, dans cette introduction, quelques étapes du développement de ces outils, souvent considérés antagonistes mais dont l'histoire nous montre la complémentarité.

La naissance du risque social

Le principe de causalité, qui proclame la primauté de l'entendement humain sur un ordre supérieur inaccessible, fut l'un des ingrédients essentiels de la révolution scientifique et philosophique du XVIIe siècle, avant d'atteindre l'ordre politique au XVIIIe. Galilée rejetant le cosmos aristotélien, Descartes fondant sa physique sur la seule notion de *mécanisme*, Spinoza énonçant *causa sive ratio* (« cause, c'est-à-dire raison ») disaient tous que « la connaissance scientifique d'une réalité quelconque consiste à écarter toute notion de finalité au bénéfice de la causalité » (Lecourt, 1999, p.174). Puis Leibniz en faisait un Principe de la raison suffisante, fondement son déterminisme métaphysique¹ : toute chose a une raison qui permet de l'expliquer.

Seulement vingt-cinq ans après Leibniz, Hume contestait la légitimité de la causalité à être le fondement de la raison en montrant qu'elle n'avait de racine ailleurs que dans notre expérience :

"Thus we remember to have seen that species of objects we call flame, and to have felt that species of sensation we call heat. We likewise call to mind their constant conjunction in all past instances. Without any farther ceremony, we call the one cause and the other

1. « Ainsi il faut que la raison suffisante, qui n'ait plus besoin d'une autre raison, soit hors de cette suite des choses contingentes, et se trouve dans une substance, qui en soit la cause, ou qui soit un être nécessaire, portant la raison de son existence avec soi ; autrement on n'aurait pas encore une raison suffisante, où l'on puisse finir. Et cette dernière raison des choses est appelée Dieu » (Leibniz, 1714).

effect, and infer the existence of the one from that of the other" (Hume, 1739, vol. 1, part III, section VI).

Pourtant Hume n'appelait pas à renoncer à la notion de causalité. Il considérait au contraire qu'il s'agissait de la relation associative la plus forte, plaidait pour une conception *subjectiviste* de la causalité² et proposait même huit règles empiriques permettant de porter un jugement de causalité (op. cit., vol. 1, part III, section XV). Son scepticisme n'avait, à ce stade, pas d'implications pratiques ; celles-ci ne se manifesteraient que bien plus tard.

Le principe de causalité était indissociable de l'idée de responsabilité de l'homme dans la compréhension et la maîtrise de l'enchaînement de causes et d'effets qui faisait la marche du monde. Une expression particulièrement forte de cette idée était la notion de devoir de la collectivité de protéger les hommes contre les méfaits de la Nature et du Sort. À mesure que la croyance en une punition divine faiblissait, que les personnes, villes et régions frappées par un malheur étaient perçues comme des victimes d'une injustice, cette notion de responsabilité collective se renforçait. Le point de basculement, s'il y en eut un, fut le désastre de Lisbonne.

Le 1er novembre 1755, la capitale portugaise fut frappée par un tremblement de terre et un raz-de-marée dévastateurs qui firent près de 10 000 morts³. L'événement inspira à Voltaire son Poème sur le désastre de Lisbonne, qui était une charge contre la philosophie optimiste de Leibniz et un appel au sentiment de solidarité entre des hommes également exposés au « sort ennemi » (Voltaire, 1756). Dans la Lettre sur la Providence qui engageait sa célèbre polémique avec Voltaire, Rousseau insistait pour sa part sur les responsabilités humaines dans le désastre : « la nature n'avait point rassemblé là vingt mille maisons de six à sept étages » (Rousseau, 1756).

Si les historiens ont reconnu dans cet événement un moment fondateur de l'Europe des Lumières, un « séisme philosophique » (Baczko, 1997), ses répercussions politiques et scientifiques sont sans doute moins connues. Les mesures politiques prises après le désastre de Lisbonne conjuguèrent, dans des termes actuels, une évaluation d'urgence des dommages, l'organisation systématique de secours et le maintien de l'ordre, un retour d'expérience, des mesures de prévention des risques futurs et un effort coordonné de reconstruction, c'est-à-dire l'ensemble des composantes d'une politique moderne en matière de risques sociaux. Déjà sur la pente du déclin économique, isolée des grandes capitales d'Europe, c'est « dans l'air du temps » que Lisbonne trouva les sources de sa rénovation architecturale et politique (Franca, 1965).

Dans le reste de l'Europe, l'événement suscita une effervescence scientifique sans précédent. Il est vrai que la secousse fut ressentie à des milliers de kilomètres et que durant les années qui suivirent, une grande partie du continent subit des séismes à répétition. Cette solidarité de fait contribua certainement à faire de la catastrophe un moment historique. Des débats et des concours furent

2. Voir à ce propos la citation placée en exergue du chapitre 8.

3. Soit 4 % de la population de la cité (Franca, 1965).

organisés, dans lesquels les savants firent assaut de connaissances (et d'imagination) pour proposer de nouvelles explications théoriques à la survenue des séismes. Un mois après la catastrophe, le géographe Philippe Buache élaborait la première carte des risques sismiques en Europe. À sa suite, Gueneau de Montbeillard établit une liste chronologique considérée comme l'un des premiers catalogues sismologiques, mais qui portait sur le périmètre plus vaste « des éruptions de volcans, des tremblements de terre, de quelques faits météorologiques, des comètes, des maladies pestilentielles, des éclipses les plus remarquables jusqu'en 1760 » (cité par Quenet (2005, p.313)). En 1756, informé que ses sujets avaient ressenti une secousse tellurique, Frédéric le Grand interdit que l'on évoquât la survenue d'un tel fléau dans le Royaume de Prusse et menaça de prison ceux qui voudraient ainsi « induire le public en erreur ». Ceci n'empêcha pas les gazettes de rapporter qu'un séisme s'était produit, le même jour et à la même heure, dans les pays limitrophes (Quenet, 2005).

Dans les dernières décennies du XVIII^e siècle, l'idée selon laquelle les accidents et désastres sont des phénomènes qu'il est possible d'anticiper et contre lesquels le pouvoir politique a le devoir de prémunir les citoyens était fermement établie. En 1789, à la veille de la Révolution française, on pouvait ainsi lire dans le tome de l'Encyclopédie Méthodique consacré à la police et aux municipalités, à l'entrée *Accident* :

« C'est en général un événement imprévu dont il résulte quelque malheur ou quelque trouble public. L'attention à prévenir les accidents, et à réparer les malheurs qu'ils ont occasionnés, lorsqu'on n'a pas pu les prévenir, forme un des plus importants et des plus essentiels devoirs des magistrats et officiers de police. C'est, en effet, sur leur vigilance à cet égard que reposent la tranquillité et la sécurité publiques. Il étoit, en effet, impossible que des hommes occupés journellement de leurs affaires personnelles, du soin de leur famille ou de leur commerce, pussent prévoir les accidents et les éviter. Il falloit qu'une police attentive s'en chargeât et veillât perpétuellement sur le repos public et la sûreté des citoyens. C'est à elle à les mettre à l'abri des accidens sans nombre auxquels l'homme est exposé dans l'état de société ; elle ne peut y manquer sans se rendre coupable d'un délit positif, et dont on a droit de la punir. C'est ainsi qu'en 1499 le parlement de Paris cassa les officiers municipaux de cette ville, et les déclara incapables de remplir aucune charge publique, pour ne pas avoir prévenu l'accident de la chute du pont Notre-Dame, chute annoncée en quelque sorte par l'état de vétusté et de délabrement où étoit ce pont. Mais tous les accidens ne peuvent pas être prévus ; il en est qui dépendent des intempéries des saisons et de l'action des éléments ; d'autres des entreprises et des travaux qui ont lieu dans la société ; d'autres enfin de la malice et de la méchanceté des hommes. On conçoit qu'il n'est pas toujours au pouvoir de la police de prévoir ni d'empêcher ceux de la première et de la dernière espèce ; mais il l'est au moins de tenir en

état tout ce qui est nécessaire pour diminuer les malheurs qui en sont la suite, ou qui les accompagnent. C'est ainsi qu'on ne peut pas prévoir une inondation, une grande chute de neige, les incendies produits par la foudre, etc. mais dans une ville bien administrée, on doit avoir des moyens toujours prêts d'arrêter le cours ou la multitude des malheurs qui suivent ces désastres, suivant la nature des lieux et les facultés des habitans. On ne peut pas non plus connoître ce qui se passe dans le cœur des hommes pervers ; et les soins les plus exacts d'une police vigilante n'empêcheront que faiblement des brigands, des ennemis du repos public, de causer de très-grands accidens, s'ils en ont fixément la volonté. [...] Quant aux accidens qui naissent des travaux, établissemens, entreprises, événemens publics qui ont lieu dans la société ; c'est-là qu'excelle la prévoyance de la police. Comme tout peut être prévu en pareil cas, il dépend d'elle de prendre de si sages mesures que les accidens ne puissent point naître. [...] Mais pour qu'un accident soit réputé cas fortuit à l'égard de la police, il faut qu'il ne soit jamais arrivé auparavant dans une circonstance semblable ; car lorsqu'un événement accidentel se présente deux fois de suite, il doit cesser d'être imprévoyable aux yeux d'une administration chargée de le prévenir » (Auteurs divers, 1789).

Il est remarquable, d'un point de vue actuel, que l'article évoque l'essentiel des concepts et principes de prise en compte des risques sociaux mis en œuvre au cours de deux siècles suivants sans faire aucune mention de risque, de probabilité ou d'aléa. La première phrase caractérise certes l'accident par référence à l'incertitude et à la possibilité de perte, mais toute la suite du paragraphe ne traite de cette perte éventuelle que sous l'angle de ses causes et des responsabilités qui en découlent, évoquant tour à tour : le devoir public de prévention et de compensation des dommages ; son lien avec le contrat social ; l'avantage informationnel des instances publiques ; le principe de leur responsabilité ; la catégorisation des aléas (naturels, accidentels, intentionnels) ; des principes différenciés d'action selon la nature des aléas ; la mitigation des dommages ; la difficulté fondamentale à prévoir les actes intentionnels ; enfin le principe d'inférence et d'anticipation. L'article distingue en particulier les accidents réellement imprévisibles, liés à l'action de la Nature ou à la méchanceté des hommes, de ceux qu'un examen responsable des faits permet d'anticiper et de prévenir. Même pour les premiers, toutefois, elle affirme qu'il est nécessaire d'intervenir *a posteriori* afin d'empêcher qu'ils produisent pleinement leurs effets. C'est, en quelque sorte, la proclamation d'un principe de prévention deux siècles avant celui de précaution.

En quelques décennies, les rapports entre l'individu et la société avaient profondément changé. La prise en compte de l'incertitude et du danger avaient joué un rôle important dans ce changement, et ses principes cardinaux étaient la responsabilité et la causalité. La probabilité en était totalement absente.

La vision probabiliste du risque

Pourtant, les probabilités et les statistiques avaient alors plus d'un siècle d'existence.

Les premiers fondements de la théorie des probabilités avaient été posés par Pascal et Fermat dans leurs célèbres échanges au cours de la décennie 1660. Moins de vingt ans plus tard, en 1678, Leibniz avait défini la probabilité comme étant un degré de possibilité dans son *De incerti aestimatione*, et déduit de son Principe de raison suffisante un postulat fondamental de la théorie naissante des probabilités, celui d'équipossibilité : puisque la raison pour laquelle un événement se produit plutôt qu'un autre est *pré-existante*, nous n'avons aucune raison de privilégier l'occurrence d'un événement par rapport à un autre tant que nous ne distinguons par une telle raison ; autrement dit, en présence de deux événements que rien ne distingue, nous devons considérer que leurs occurrences sont également possibles. Pour les successeurs de Leibniz, l'équipossibilité était ainsi celle des *causes*⁴. Lorsqu'en 1739, il remit en cause l'inférence inductive fondée sur les relations de cause à effet, Hume mit sur le même plan le raisonnement par les probabilités, puisqu'il s'agissait du même mode de raisonnement (Morris, 2009).

Les éléments initiaux formulés par Pascal et Leibniz ne prirent véritablement corps que quelques décennies plus tard, avec les travaux de Montmort, des Bernoulli et de Moivre (Hacking, 1975). En 1713, l'*Ars conjectandi* de Jakob Bernoulli formula une première version de la loi des grands nombres et, s'appuyant sur l'équipossibilité leibnizienne, proposa la première interprétation personnaliste (ou subjective) des probabilités. La même année, dans une lettre à Pierre de Montmort, son neveu Nicolas Bernoulli inventa le problème qui fut reconnu plus tard sous le nom de paradoxe de Saint Pétersbourg. En 1718, Abraham de Moivre établit le résultat d'approximation d'un échantillon de lois binomiales par la loi normale. En 1738, Daniel Bernoulli, cousin de Nicolas, résolut le paradoxe de Saint Pétersbourg en introduisant la notion d'utilité. En un quart de siècle, une théorie des probabilités avait pris forme.

Qu'il s'agisse du juste partage de gains ou de la foi en Dieu, les usages initiaux de la probabilité gravitaient uniquement autour de la question de la Raison et du comportement de l'Homme raisonnable. Dans cette optique, les probabilités subjectives apparaissaient comme un prolongement naturel des probabilités fréquentielles, puisque l'Homme raisonnable devait naturellement intégrer à son jugement toutes les régularités qui s'offraient à lui. La dualité de la probabilité, qui est pour Ian Hacking une sorte de trait congénital que lui ont légué les notions dont elle a été dérivée à l'origine, est donc restée invisible durant l'« âge classique » de la probabilité (Daston, 1988).

Les débuts de la statistique furent concomitants à ceux des probabilités, et indépendants de ceux-ci. À partir de la peste de 1603, un registre des décès fut tenu dans la ville de Londres comme

4. Le principe d'équipossibilité fut plus tard baptisé principe de raison insuffisante par von Kries, puis d'indifférence par Keynes.

moyen de détection et de suivi des épidémies. En 1662, le mercier John Graunt établit, à partir de soixante années de données, les premières tables de mortalité, et publia ses résultats dans ses *Natural and political observations made upon the bills of mortality*, considérés comme la première véritable étude statistique. Dès 1669, le mathématicien et médecin hollandais Huygens avança, à partir d'un raisonnement sur les annuités, une interprétation probabiliste des résultats de Graunt. Pourtant, il se passa plus d'un siècle, pendant lequel les études statistiques cherchèrent avant tout à montrer l'intérêt de la collecte de statistiques et les mathématiciens se tournèrent vers l'astronomie pour éprouver leurs outils probabilistes, avant que la rencontre de ces deux courants ne donne naissance à une science de la statistique sociale (Hald, 1990).

Les probabilités ne s'imposèrent pas davantage dans le domaine du droit et de l'assurance, bien que le risque fut une notion connue des juristes et des assureurs au XVII^e siècle. Ainsi que le note Pradier (1998, p. 66), l'usage du vocable *a*, dans ce domaine, de loin précédé une quelconque tentative de calcul effectif. Le risque permettait alors, dans le prolongement de la pensée scholastique, de justifier l'intérêt et le juste gain en les plaçant à égale distance de l'usure et du pari, activités immémoriales, très répandues, mais également répréhensibles du point de vue de la morale. C'était bien l'absence ou l'insuffisance d'incertitude, et non les conditions de rémunération, qui définissaient alors l'usure. L'incertitude était celle d'une opération d'inférence des causes aux effets ; portait à l'usure une situation où les causes semblaient suffisamment prévisibles pour qu'il fasse peu de doute qu'un certain effet se produirait. De même, lorsque les causes étaient mal connues ou imprévisibles, le degré d'incertitude était jugé excessif, ce qui caractérisait un pari. Les contrats aléatoires tentaient de donner un cadre juridique à cette idée (Le Goff, 1962).

Il fallu attendre le tournant du XIX^e siècle pour que la théorie des probabilités commence à être appliquée aux problèmes d'assurance sociale comme à l'analyse des statistiques sociales. Ce délai de près d'un siècle dans la diffusion d'un concept qui, d'un point de vue contemporain, semblait ne demander qu'à être mis en œuvre, intrigue. Il serait tentant de n'y voir que l'effet de l'ancrage des pratiques dans des traditions pittoresques, jusqu'à ce que celles-ci finissent par être balayées par l'avènement de la rationalité mathématique⁵.

Les historiens contemporains, même s'ils reconnaissent un certain rôle à l'inertie des mentalités, nous mettent cependant en garde contre cette interprétation exclusive. Les primes d'assurance appliquées au dix-huitième siècle au commerce transatlantique depuis le port de Nantes étaient strictement proportionnelles aux taux de sinistre constatés *a posteriori* (Ducoin, 1993). En dehors des cas d'escroquerie, les faillites d'assureurs semblent avoir été extrêmement rares, et les travaux disponibles au sujet des contrats aléatoires complexes (contrats *ad tempus*, prêts « à la Grosse ») indiquent que ceux-ci n'offraient pas d'opportunité d'arbitrage aisé (Pradier, 1998, pp. 72-4).

5. Boiteux (1968), cité par Pradier (op. cit., p. 66), caricature ainsi la méthode de l'assureur classique : "he looks at the sky and thinks of a rate".

Les assureurs savaient donc calculer leurs primes et s'ils résistaient à l'usage des statistiques, c'est qu'il leur fallait évaluer les risques au cas par cas. De ce fait, leurs meilleurs outils étaient une longue expérience et un réseau dense d'informateurs :

"It was not a world of constant surprises, but it was one where specific, up-to-the-minute, and above all personal knowledge counted" (Daston, 1988, p. 112).

Lorraine Daston apporte une explication convaincante à l'absence d'une réelle mutualisation des risques durant l'âge d'or des probabilités : le manque de régularité. Dans un univers d'instabilité où les causes de décès subit, de ruine ou de naufrage étaient nombreuses et pouvaient s'étendre soudainement à une région, une fraction de la population ou une période temporelle, il importait que l'assureur apprécîât chaque cas en se fondant sur une évaluation *ad hoc* de ces facteurs. Un assureur qui aurait proportionné ses annuités à l'âge seul de l'assuré ou ses primes nautiques à la seule longueur du trajet se serait rapidement échoué sur les écueils de la corrélation des risques due aux guerres ou aux épidémies, et sur ceux de la sélection adverse au sein d'une population fortement hétérogène. Il lui fallait différencier les risques, c'est-à-dire, avant de se prononcer sur les chances d'un certain événement, de s'informer sur un ensemble aussi vaste que possible de ses *causes*.

Pour que s'impose le principe proportionnalité entre le risque et un petit ensemble de causes générales, autrement dit le choix d'ignorer les facteurs causaux répartis de façon hétérogène dans une population au motif qu'ils finiraient par se compenser, il a fallu que le poids de ces derniers commence à diminuer par rapport à l'influence des premières. C'est donc la transformation des sociétés dans le courant du XVIIIe siècle, avec le recul de certains fléaux⁶, la pacification des mœurs et l'émergence d'une classe moyenne, particulièrement en Angleterre, en résumé la montée de l'« homme moyen », qui aurait levé les derniers obstacles à la mutualisation des risques. L'expansion de l'assurance individuelle qui s'est ensuivie était en symbiose avec la mise en évidence de nombreuses régularités sociales par les appareils statistiques publics.

Une coopération souterraine

Dès le XVIIIe siècle, des données concernant de multiples aspects de la vie dans les sociétés européennes étaient collectées dans le cadre de relevés administratifs. À partir des années 1830, elles commencèrent à être publiées et étudiées en tant que « statistiques morales (Lazarsfeld, 1961). Le mathématicien belge Adolphe Quételet joua un rôle important dans le développement de cette pratique, d'une part en participant à la constitution des premiers systèmes statistiques nationaux rassemblant les recensements de population et les données sur l'état civil, la santé et la criminalité ; et de l'autre en montrant, par l'application des outils des probabilités, l'existence de régularités

6. Certes, pas encore les épidémies et les famines, mais les grands incendies, la forte mortalité infantile et la piraterie.

sociales : il se produisait chaque année, dans une société, à peu près le même nombre de crimes, de mariages, de suicides ou de naissances (Desrosières, 1993). Les travaux de Quételet déclenchèrent une onde de choc philosophique dont les effets étaient encore perceptibles dans la littérature du XXe siècle (Bouveresse, 1993). Outre la richesse des informations spécifiques qu'ils délivraient pour la première fois au sujet de faits sociaux, ils posaient la question de la nature des régularités émergeant de comportements individuels. Pour les partisans du déterminisme social, il s'agissait de lois ayant le même statut que les lois de la physique, indépendantes des individus et s'imposant à eux. Pour leurs opposants, en particulier les philosophes et statisticiens allemands de la première moitié du XIXe siècle, des régularités ne pouvaient tenir lieu de lois tant qu'elles ne correspondaient pas à l'action de causes identifiées (Hacking, 1990).

La controverse au sujet des lois sociales coïncida avec les débuts du positivisme sous la férule d'Auguste Comte. Comte ne discutait pas la validité de la notion de causalité, mais se prononçait en faveur d'un programme de recherche scientifique qui se satisfasse de ce qui était observable, et renonce à l'idée de recherche de causes *premières*, qui relevaient de la métaphysique. Il était, par ailleurs, très méfiant à l'égard de la statistique érigée en science sociale, et c'est malgré lui que ses préceptes furent repris pour soutenir une approche de la probabilité comme grandeur physique : la fréquence relative observée au cours d'une longue série d'observations.

Des « lois » sociales sans cause observable, mais dont la stabilité était montrée avec éclat par les fréquences observées : le déterminisme social marque ainsi le début du divorce entre les concepts de causalité et de probabilité.

En pratique, cependant, c'est bien l'usage conjoint des deux instruments qui produisait les effets les plus remarquables, notamment pour *rechercher* des causes, certes cette fois sociales et non individuelles, aux régularités observées et pour en tirer des conclusions opérationnelles dans un domaine comme l'hygiène publique. Lorsqu'en 1840, Louis-René Villermé produisit le *Tableau de l'état physique et moral des ouvriers*, il était déjà l'auteur de plusieurs études sur le rapport entre la mortalité et la pauvreté ou l'industrialisation. Il consacra la première partie du tableau à une description personnelle des conditions de vie des ouvriers du textile, dénonçant notamment le travail des enfants comme une « torture ». La Société industrielle de Mulhouse s'appuya sur son travail pour demander à la Chambre l'adoption d'une loi contre le travail des enfants (La Berge, 1992). La loi fut adoptée en 1841.

En analysant la localisation des victimes du choléra à Londres au cours des deux épidémies de 1848-49 et 1853-54, John Snow mit en évidence une liaison entre la prévalence de la maladie et la distribution d'eau, dont il inféra une nouvelle théorie sur la contagion. En 1854, sous son influence, les autorités municipales fermèrent la pompe d'eau de Broad street, et l'épidémie prit rapidement fin. L'étude complète de Snow sur le choléra fut publiée en 1855, trente ans avant que

Robert Koch ne parvienne à isoler l'agent responsable de la maladie⁷. Elle constitua l'acte fondateur de l'épidémiologie, qui se développa rapidement. La méthodologie était fondée sur l'analyse statistique de la distribution de la maladie dans la population à partir des tables de mortalité, et sur l'identification de spécificités biologiques, chimiques, physiques, sociales ou culturelles corrélées à sa prévalence. Ces *facteurs* permettaient d'expliquer les mécanismes de propagation de la maladie et, par induction, d'anticiper son impact futur.

La découverte de la corrélation par Francis Galton en 1888 donna l'occasion d'une nouvelle charge contre la causalité. La corrélation était la première mesure empirique, détachée de toute appréciation subjective, de la relation entre deux variables. Karl Pearson y voyait le substitut scientifique tant attendu de la notion de causalité :

"Galton turning over two different problems in his mind reached the conception of correlation : A is not the sole cause of B, but it contributes to the production of B ; there may be other, many or few, causes at work, some of which we do not know or may never know... This measure of partial causation was the germ of the broad category - that of correlation, which was to replace not only in the minds of many of us the old category of causation, but deeply to influence our outlook on the universe" (Pearson, 1914-30).

Au cours des années 1910-1920, pourtant, deux autres pères de la statistique moderne, Ronald Fisher et Jerzy Neyman, et une médecin britannique, Janet Lane-Claypon, élaborèrent la méthode des essais contrôlés (*randomised experiments*) qui prenait acte de l'insuffisance de la corrélation et posait les bases d'une théorie de la causalité ne disant pas son nom : il s'agissait de tester l'effet d'un traitement en mesurant les conséquences de deux *interventions* distinctes (administration du traitement ou non) sur deux populations identiques, ce qui était la première formalisation des théories dites interventionnistes de la causalité⁸.

Le travail de Lane-Claypon sur les problèmes de maternité au cours de ces années 1910-1920 fournit les instruments du renouveau de l'épidémiologie après l'éradication des grandes maladies infectieuses en Europe (Winkelstein, 2004). En 1912, Lane-Claypon mit en œuvre la première étude de cohorte pour mesurer l'influence de l'alimentation des nourrissons. En 1926, sur demande du Ministère de la Santé du Royaume-Uni, elle réalisa la première étude épidémiologique sur le cancer, en comparant pour la première fois un groupe de malades à un groupe de contrôle. Son travail pionnier fut repris dans les années 1950 par Richard Doll et Austin Bradford Hill, dans le but d'établir une relation entre usage du tabac et cancer du poumon d'une part, maladies cardio-

7. La véritable découverte du bacille du choléra revient en fait à l'anatomiste italien Filippo Pacini, qui l'isole en 1854 avant de décrire le mécanisme de déshydratation qu'il provoquait et de proposer des traitements efficaces, mais ses travaux furent rejetés par la plupart de ses pairs dans son pays et restèrent inconnus à l'étranger.

8. Nous étudierons ces théories au chapitre 5. Sur le lien entre essais contrôlés et causalité, on pourra se reporter à Pearl (2000, pp.259 et suivantes).

vasculaires de l'autre (Doll et Hill, 1950 ; 1954). Le retentissement de ces études, notamment de la seconde (connue sous le nom de *British doctors study*) ouvrit un nouveau champ à l'investigation épidémiologique. Dans les années suivantes, l'épidémiologie s'intéressa à un nombre croissant de facteurs collectifs ou individuels qui contribuaient à l'apparition d'une pathologie.

L'objectif premier de l'épidémiologie des maladies chroniques fut, comme dans le cas du choléra, de leur associer statistiquement des « facteurs de risque » avant que les mécanismes causaux qui les lient ne soient entièrement compris. Ceci explique le recours très répandu aux critères de causalité proposés par Hill (1965), qui actualisent les règles de l'inférence inductives proposés par Hume et, après lui, John Stuart Mill : force, généralité, spécificité, temporalité, régularité, plausibilité, cohérence, observabilité, analogie. On en était ainsi revenu à des critères empiriques (Rothman, 1976), ce constituait un recul par rapport à la méthode rigoureuse de l'essai contrôlé. L'épidémiologie moderne, toutefois, a corrigé cette erreur et contribué au renouveau des théories de la causalité (Rothman, Greenberg et Lash, 2008).

Contexte et objectifs de la recherche

Bien que séparés par un courant majoritaire de la statistique théorique et de l'épistémologie, la probabilité et la causalité ont donc constitué, depuis deux cent cinquante ans, les deux piliers de l'inférence inductive et de l'action publique en matière de risques sociaux. Or cette double fracture, entre probabilités et causalité en théorie d'une part, entre théorie et pratique de l'autre, pourrait être en voie de réduction.

Les années récentes ont vu un renouveau de l'intérêt pour les théories de la causalité, grâce à l'attrait de modèles dérivés des réseaux bayésiens (Pearl, 2000 ; Spirtes, Glymour and Scheines, 1993 ; Woodward, 2003). L'idée centrale de ces modèles est de caractériser la causalité par la réponse à l'intervention humaine : ce qui distingue une relation causale entre A et B d'une corrélation entre A' et B' , c'est qu'en *provoquant* A , nous pouvons observer la réalisation de B . En transcrivant cette idée dans le cadre de réseaux probabilistes, ces modèles réconcilient la probabilité et la causalité et répondent au défi de Pearson : A peut être *une* cause de B , ainsi que C , D ou E , et nous pouvons même mesurer la force de cette relation. Par ce fait même, ils portent la promesse de permettre l'étude d'une palette de phénomènes allant du déterministe au stochastique à l'intérieur d'un cadre formel unifié.

Apparus il y a une vingtaine d'années, ces modèles ont fait converger vers eux l'intérêt d'un grand nombre de disciplines, depuis l'intelligence artificielle jusqu'aux sciences sociales. Dans le champ de la théorie bayésienne de la décision, toutefois, l'événement n'a été guère remarqué que par des philosophes (Joyce, 2007 ; 2010). La forte influence de l'empirisme logique sur la tradition bayésienne n'est sans doute pas étrangère à cette indifférence.

Pour qui s'intéresse à la décision face à un risque social, pourtant, la nouvelle théorie de la causalité constitue un point de passage obligé. La raison n'en est pas seulement que, comme ce qui précède l'a illustré, la causalité a été un instrument essentiel de la construction des décisions en matière de risques sociaux. C'est aussi que le risque social constitue un sujet épineux pour la théorie de la décision. Malgré le renouvellement de son cadre et la diversification de ses modèles entraînés par le relâchement des postulats de Savage, celle-ci n'a pas encore, par exemple, trouvé de traduction convaincante au principe de précaution. Les modèles d'aversion pour l'ambiguïté (Gilboa et Schmeidler, 1989 ; Klibanoff, Marinacci et Mukerji, 2005) ne correspondent pas à l'idée d'une incertitude radicale empêchant l'évaluation du risque, qui est centrale dans le principe de précaution. La théorie de la décision ne permet pas davantage de rendre compte de raisonnements déterministes, qu'elle assimile au choix irrationnel d'une distribution de probabilités dégénérée, alors que ceux-ci sont fréquents dans de nombreux domaines de risques sociaux. Pour finir, elle ignore totalement la phase critique d'évaluation du risque, dans laquelle une représentation d'une situation incertaine est construite et les actions possibles sont identifiées.

On voit donc se dessiner, autour de la question du risque social, une potentialité théorique d'un côté et un manque de l'autre. Peut-on proposer une représentation convaincante des risques sociaux à partir de la nouvelle théorie de la causalité ? Peut-on rendre compte par ce biais des raisonnements causaux et probabilistes qui coexistent dans ce domaine ? Peut-on enfin examiner les problèmes de décision en situation de risque social à partir d'une telle représentation ?

Le projet de recherche délimité par ces questions suppose une démarche d'inférence inductive : n'ayant pas de théorie prête à l'usage, nous devons commencer par décrire et comprendre un champ pratique avant d'espérer forger notre outil théorique.

Le champ que nous avons retenu pour cela est celui de la sûreté nucléaire. Le secteur nucléaire présente des spécificités qui l'ont conduit à faire un effort exceptionnel de compréhension du risque consistant à décomposer celui-ci en un ensemble structuré de scénarios accidentels. Cet effort a pu être entrepris grâce à la nature particulière des systèmes que sont les installations nucléaires, notamment au fait que leurs interactions avec leur environnement sont très strictement contrôlés. Il a ainsi paru possible d'énumérer tous les événements potentiellement dommageables et de décrire les développements auxquels ils conduisaient. Cette démarche de représentation formelle et d'identification des sources du risque à l'intérieur d'un système fait du secteur nucléaire un sujet d'étude parfaitement adapté à nos besoins. L'un des enjeux du travail sera la pertinence d'une approche du risque social inspirée du cas nucléaire pour d'autres catégories de risque, qui dépendra notamment de la généralité que nous parviendrons à donner au concept de système.

Annnonce du plan

L'exercice de définition des termes peut être convenu ; il ne l'est pas ici. Outre le flou sémantique qui entoure le terme risque, c'est l'absence de notion bien établie de risque social qui nous y contraint. Il nous faut donc tenter de caractériser ce qu'est un risque social, en particulier en l'opposant au risque privé, et de procéder à une discussion terminologique qui nous permette de progresser dans notre travail sans être gênés par l'ambivalence de nos propos. C'est l'objet du premier chapitre, que nous entamons en nous donnant une première définition volontairement vague du risque social : une situation incertaine comportant la possibilité de pertes significatives du point de vue de la société. Armés de ce concept provisoire, nous faisons un tour d'horizon des phénomènes qui semblent en relever, au premier chef les catastrophes naturelles, les accidents industriels et technologiques et les problèmes de santé publique. L'analyse des terminologies du risque associé à chacune de ces familles de phénomènes montre que la polysémie du risque n'est pas simplement liée à des différences de tradition, et qu'elle est encore moins artificielle. Elle apparaît comme l'inévitable conséquence de la difficulté à délimiter précisément le risque, qui a conduit au développement d'une méthodologie de *représentation* du risque propre à chaque secteur. Nous en tirons trois conclusions : premièrement, la diversité terminologique est irréductible et nécessaire ; deuxièmement, il existe cependant, en-deçà des représentations du risque, des éléments communs qui nous permettent de définir un risque social ; troisièmement, le risque est indissociable de sa représentation.

La décision et l'action face au risque forment le deuxième volet de notre travail préalable de clarification. La difficulté principale, ici, n'est pas posée par le contenu des termes mais par l'identification du décideur. Peut-on parler de décision sociale comme on parle de décision privée ? Existe-t-il une instance qui puisse incarner la figure du décideur telle qu'elle est campée par la théorie de la décision, c'est-à-dire une entité consciente, considérant les conséquences possibles de ses actions et évaluant ces dernières au moyen de préférences cohérentes et stables ? Après avoir passé en revue certains développements législatifs, réglementaires et institutionnels récents et brièvement étudié le problème de la gestion des risques d'inondation en France, nous répondons par la négative à ces deux questions. Nous défendons néanmoins l'utilité du modèle du décideur individuel comme outil d'aide à l'analyse du risque et d'information de la société.

Nous abordons alors notre étude du secteur nucléaire. Le troisième chapitre est consacré aux méthodes d'analyse de la sûreté nucléaire, qu'il est d'usage de répartir en deux grandes familles, l'une déterministe et l'autre probabiliste. Le développement de ces méthodes a accompagné celui de l'industrie nucléaire civile, et elles ont été progressivement corrigées et complétées sur la base de l'expérience opérationnelle accumulée et des problèmes de sûreté observés. Nous présentons en détail les deux démarches, en décrivant et en commentant à la fois leurs principes, leurs modalités pratiques et les principales étapes de leur évolution. Ceci nous permet de porter une appréciation

nuancée sur la complémentarité supposée de ces démarches. La démarche déterministe permet de couvrir une part d'incertitude ignorée par la démarche probabiliste et cette dernière, en retour, peut mettre en évidence certaines failles dans la première. Cependant, toutes deux ont aussi des limites communes, qui sont celles de notre compréhension des phénomènes en jeu dans une centrale nucléaire. Ces limites ne font pas l'objet d'une réflexion méthodique aujourd'hui.

Dans le quatrième chapitre, nous étudions les politiques de sûreté nucléaire suivies aux États-Unis et en France sous l'angle historique. La genèse particulière de l'industrie nucléaire civile, dans le sillon de l'arme atomique, l'attitude extrêmement volontariste des autorités publiques durant la période de montée en charge industrielle (dans les années 1950 aux États-Unis et 1970 en France), enfin l'absence, au moins jusqu'au milieu des années 1970, d'autorités de contrôle indépendantes expliquent que le secteur ait bâti un mythe de la sûreté absolue dont il s'est retrouvé l'otage par la suite. Les principes généraux de la sûreté nucléaire ont été posés à une époque où l'on s'affirmait convaincu de l'inviolabilité de l'enceinte de confinement et où l'on choisissait de parler de « démonstration de la sûreté »⁹. Les méthodes probabilistes de la sûreté ont été adoptées en large part pour corriger ce défaut originel, notamment après l'accident de Three Mile Island. Depuis, la dualité déterministe/probabiliste semble avoir occupé tout l'espace de réflexion relative à la prise en compte de l'incertitude.

Nous passons, à ce stade, aux parties plus formelles de notre projet.

Le cinquième chapitre propose une synthèse et une discussion de la théorie des réseaux bayésiens causaux, dont nous développons une version adaptée à nos besoins. Nous y rappelons la définition et les principales propriétés de réseaux bayésiens, avant de présenter leurs deux interprétations causales, la première fondée sur les distributions de probabilités et la seconde sur les modèles fonctionnels. Ces interprétations font appel à la notion d'intervention, qui consiste à modifier localement le réseau afin d'y injecter une certaine valeur. La réponse du réseau constitue une mesure de l'effet causal de la variable qui a fait l'objet de l'intervention. Après avoir justifié notre préférence pour la forme fonctionnelle, nous proposons un élargissement de la notion d'intervention et montrons que le modèle théorique obtenu permet de reproduire l'ensemble des raisonnements déterministes et probabilistes étudiés dans le cas nucléaire. Le modèle constitue en outre un cadre conceptuel unique pour comparer ces raisonnements et analyser leurs conditions de validité. Nous l'utilisons alors pour proposer une heuristique des démarches déterministe et probabiliste d'inférence et nous refermons le chapitre sur une définition générale du risque dans le prolongement de nos considérations du premier chapitre.

Dans le sixième chapitre, nous revenons au cas nucléaire, mais cette fois équipé de notre outil théorique. Nous proposons une analyse de l'accident de Fukushima d'où ressort l'importance

9. La « démonstration de la sûreté » reste l'expression officielle, aux États-Unis comme en France, pour désigner le fait qu'une installation est conforme aux *exigences réglementaires de sûreté*.

des éléments d'incertitude ignorés par les évaluations déterministes et probabilistes de la sûreté, notamment ceux relatifs aux risques sismiques. Nous en recherchons les causes spécifiques dans le contexte institutionnel japonais, avant de nous tourner vers les enseignements plus généraux de l'accident. Le processus de retour d'expérience qui a commencé à être mis en place dans les pays européens semble marqué par la volonté d'améliorer la prise en compte de l'incertitude pour la catégorie particulière de phénomènes à l'origine de l'accident de Fukushima, c'est-à-dire les aléas naturels. Nous y voyons à la fois une confirmation de notre analyse générale relative aux limites des doctrines de sûreté nucléaire et le signe d'une évolution positive, mais peut-être trop limitée, de ces doctrines. Nous illustrons ce propos par une analyse de l'évaluation des risques d'inondations dans les centrales nucléaires françaises.

Enfin, nous développons notre modèle de décision causale en deux temps. Dans le septième chapitre, nous commençons par en exposer les principes et les avantages à partir d'un cas d'école. Nous discutons en détail les relations entre notre modèle et le cadre de Savage, et nous montrons l'apport d'une représentation causale à l'analyse des certains problèmes classiques de la théorie de la décision tels que la dépendance aux états ou l'aléa moral. Le huitième chapitre présente les développements et résultats théoriques et les positionne par rapport à certains développements récents de la théorie de la décision.

Première partie

Une problématique du risque social

Chapitre 1

Examen du concept de risque social

*Ici non plus, comme c'est si souvent le cas, l'unité du mot
ne garantit en rien l'unité de la chose.*

Nietzsche (1878)

1.1 Introduction

Qu'appelle-t-on risque social et comment caractériser les décisions publiques en situation de risque social ?

S'interrogeant sur la définition d'un fait social, Emile Durkheim observait que l'on emploie couramment la qualification de social « pour désigner à peu près tous les phénomènes qui se passent à l'intérieur de la société » et « qu'à ce compte, il n'y a, pour ainsi dire, pas d'événements humains qui ne puissent être appelés sociaux » (Durkheim, 1937, p. 3). Comme le substantif risque couvre un champ qui n'est guère plus restreint, le risque social apparaît aujourd'hui comme une notion extrêmement vaste et imprécise. En particulier, nombre d'actes publics peuvent être considérés, à un titre ou l'autre, comme des décisions en situation de risque social.

Il y a vingt-cinq ans, Ulrich Beck proposait même de caractériser la société occidentale contemporaine comme une « société du risque », où le danger, pour la collectivité, provient de multiples décisions publiques et privées, et non plus, comme par le passé, de quelques sources extérieures¹. Toutefois, Beck se gardait de définir précisément ce qu'il entendait par risque, notant que cette question tendait à être instrumentalisée par les parties en présence et que l'omniprésence du risque

1. "In contrast to all earlier epochs (including industrial society), the risk society is characterised essentially by a *lack* : the impossibility of an *external* attribution of hazards. In other words, risks depend on *decisions* [...]" (Beck, 1986, p.183, l'auteur souligne). Dans notre introduction générale, nous avons au contraire cherché à montrer que la notion de risque social est inséparable de la conviction selon laquelle « le risque dépend des décisions », et que l'une et l'autre ont émergé au milieu du dix-huitième siècle, c'est-à-dire peu de temps *avant* le début de la Révolution industrielle.

résultait de ce qu'*on* dénommât ainsi tout phénomène qui avait la modernisation pour cause et des dommages pour effet indirect². Cette non définition fait peser d'emblée sur la démonstration de Beck un soupçon de tautologie qui ne se dissipe pas à la lecture de l'ouvrage.

La question de la clarification des termes comme préalable à l'analyse, pour fastidieuse qu'elle soit, est donc particulièrement importante dans ce domaine. C'est l'objet de ce premier chapitre. Précisons : nous ne cherchons pas ici à développer une terminologie stricte à laquelle nous nous tiendrions dans le reste de ce travail. Au contraire, l'argument principal de ce chapitre est que les difficultés sémantiques sont soutendues par des problèmes méthodologiques. En l'avancant d'emblée, nous espérons gagner le droit à une certaine souplesse terminologique dans la suite.

Avant d'essayer de caractériser précisément le risque social, nous devons convenir de quelques limites à lui imposer. Cette première approche doit, *a minima*, comporter une référence à l'incertitude et à la possibilité de dommages, qui sont constitutives de l'idée de risque, ainsi qu'un critère de distinction entre risque social et risque privé. Considérons donc, à titre provisoire, comme risque social *toute situation incertaine comportant la possibilité de pertes significatives du point de vue de la société*. Nous verrons que cette caractérisation d'apparence extensive conduit malgré tout à exclure certaines questions couramment désignées comme des risques sociaux.

À partir de cette définition volontairement vague du risque social, la première partie de ce chapitre passe en revue les situations ainsi délimitées et analyse la terminologie qui y est employée. Ce tour d'horizon nous permet d'affiner notre approche dans la seconde partie, et d'élaborer une définition plus précise de notre objet d'étude.

1.2 Survol des domaines concernés

Trois catégories d'événements nous semblent représenter par excellence la possibilité de pertes sociales significatives, pour des raisons tant pratiques qu'historiques : les aléas naturels, les problèmes de santé publique et les accidents industriels et technologiques.

Mais ces catégories sont loin de constituer une typologie exhaustive des risques sociaux. Tout d'abord, elles sont assez vaguement définies ; leurs frontières sont perméables et doivent parfois être aménagées pour accueillir de nouveaux phénomènes, comme le montrent de façon éclatante les exemples du tabagisme et des « conduites à risque ». En outre, comme nous l'indiquerons dans la dernière partie de cette section, certains phénomènes peuvent relever de notre définition du risque social sans pour autant faire l'objet d'une action publique, ou parfois sans que celle-ci soit communément reliée à l'idée de risque. D'autres, sont fréquemment appelés risques sociaux, mais n'en sont pas selon nous.

2. "Actually, one can relate everything to everything else, as least experimentally, so long as the basic pattern is retained modernization as the cause, damage as the side effect" (op. cit., p.31).

Chacune des trois catégories considérées a été étudiée par une grande variété disciplines scientifiques ou techniques, et fait l'objet de lois, de règlements et de dispositions publiques ou privées. Il s'est ainsi constitué, pour elle et même souvent pour chacune de ses composantes, un domaine de connaissances et de compétences spécifique, avec sa terminologie propre. Bien que le terme risque, ainsi que d'autres notions qui lui sont associées, soient d'usage courant dans l'ensemble de ces domaines, il est notoirement difficile de parvenir à en proposer une définition opérationnelle qui soit acceptable par tous. Au-delà des approximations qu'autorise la malléabilité du concept, cette diversité semble être le fruit de difficultés méthodologiques que chaque domaine a cherché à contourner à sa manière.

1.2.1 Catastrophes naturelles

La possibilité de pertes sociales significatives trouve son expression la plus extrême dans les catastrophes naturelles, qui font, encore de nos jours, des dizaines de milliers de morts et des millions de sans-abri tous les ans dans le monde³. Ces chiffres peuvent parfois être largement dépassés à l'occasion d'un seul événement, particulièrement lorsqu'il survient dans une région pauvre, comme le tremblement de terre de Haïti en 2010 ou le tsunami du Sud-Est asiatique en 2004. Cependant, même dans les pays riches, des événements singuliers comme le grand tremblement de terre et le tsunami de Tohoku au Japon en 2011 ou le cyclone Katrina aux États-Unis en 2005 peuvent coûter la vie à des milliers de personnes, paralyser des régions entières et causer des dommages économiques considérables.

Les risques liés aux catastrophes naturelles sont souvent définis comme une fonction de la probabilité d'occurrence d'un aléa et de ses conséquences, sans autre précision quant à la forme de cette fonction. Ainsi, selon le glossaire de l'agence de l'Organisation des Nations Unies en charge de la prévention des risques de catastrophe, le risque est « la combinaison de la probabilité d'un événement et de ses conséquences négatives » (UNISDR, 2009). De même, la Directive européenne relative à l'évaluation et à la gestion des risques d'inondation définit ces risques comme « la combinaison de la probabilité d'une inondation et des conséquences négatives potentielles pour la santé

3. Les trois principales sources d'information relative aux catastrophes sont les bases de données du Centre de recherches sur l'épidémiologie des désastres (CRED) et des réassureurs Munich Re et Swiss Re. La base EM-DAT du CRED (<http://www.emdat.be>) rassemble depuis 1988 l'information relative aux événements ayant eu l'une au moins des conséquences suivantes : au moins 10 pertes humaines ; au moins 100 personnes affectées (déplacées ou ayant eu besoin d'une aide d'urgence) ; déclaration d'un état d'urgence ; appel des autorités nationales à l'aide internationale. La base inclut près de 18 000 catastrophes survenues depuis 1900. La base NatCat de la compagnie de réassurance Munich Re (<http://www.munichre.com/>) contient de l'information concernant 20 000 catastrophes survenues depuis l'an 79. Elle recense environ 800 événements nouveaux chaque année, selon des critères de sélection qui ne sont pas communiqués. La base Sigma de la compagnie de réassurance Swiss Re (<http://www.swissre.com/>) répertorie 7 000 événements survenus depuis 1970 et remplissant l'une au moins des conditions suivantes : au moins 20 pertes humaines ; au moins 50 blessés ; au moins 2 000 sans abri ; pertes *assurées* supérieures à un certain montant, variable selon le secteur. Sur la base de ces critères, Swiss Re identifie environ 300 nouvelles catastrophes chaque année. Outre les différences méthodologiques entre ces bases, les données disponibles sont sujettes à des biais de collecte importants (voir à ce propos la note 11).

humaine, l'environnement, le patrimoine culturel et l'activité économique associées »⁴. En France, la loi n'explicite pas la notion de risque naturel, bien qu'elle en fasse un usage extensif⁵. Pour le réseau d'information sur les risques majeurs mis en place par le Ministère français du développement durable : « Un événement potentiellement dangereux ou aléa n'est un risque majeur que s'il s'applique à une zone où des enjeux humains, économiques ou environnementaux sont en présence »⁶. Ici, le risque *est* donc un aléa menaçant des enjeux.

Dans les applications quantitatives, la forme de la fonction est parfois spécifiée ; il s'agit alors, sauf exception, d'un simple produit, de sorte que le risque devient une espérance de perte (Dilley et al., 2005; PNUD, 2004). Comme nous les verrons dans la suite⁷, ceci revient à supposer que la société n'a aucune aversion pour le risque, ce qui constitue une hypothèse qu'il conviendrait, pour le moins, d'indiquer explicitement.

1.2.1.1 Aléa

L'*aléa* et le *danger* sont les deux équivalents français du terme anglais *hazard*. Dans les usages qui nous intéressent ici, ils semblent avoir été débarrassés de leurs nuances sémantiques⁸ pour désigner simplement la source, la composante à *l'origine* du risque. Il existe une tendance, mais qui n'est pas unanime, à préférer l'usage d'aléa pour les risques naturels, et à réserver le terme danger aux applications sanitaires et industrielles.

Les aléas naturels sont classés en deux grandes catégories. Les aléas d'origine climatique regroupent les inondations, les tempêtes, les avalanches, la sécheresse, les feux, voire les vagues de chaleur ou de froid. Les dangers d'origine géophysique comprennent les tremblements de terre, les raz de marée (tsunamis), les éruptions volcaniques et les glissements de terrain⁹. La plupart de ces classes sont très hétérogènes. Les inondations ne sont pas de même nature selon qu'elles résultent de remontées de nappe ou de crues torrentielles. La classe des tempêtes inclut les cyclones, les tornades, ou encore les dépressions de latitude moyenne, qui sont des phénomènes très disparates en termes d'intensité et de fréquence. Les tremblements de terre peuvent survenir dans des zones de subduction (de chevauchement de plaques tectoniques), le long de failles de décrochement (à l'intérieur de ces plaques), voire suite à des éruptions volcaniques.

4. Directive 2007/60/CE du Parlement européen et du Conseil du 23 octobre 2007 relative à l'évaluation et à la gestion des risques d'inondation, Journal officiel de l'Union européenne, L 288/27, 6 novembre 2007.

5. Voir la sous-section 2.2.1.

6. <http://www.risquesmajeurs.fr/definition-generale-du-risque-majeur>

7. Voir la sous-section 2.4.2.

8. Alors que le danger désigne une situation *subie* où l'on craint des événements néfastes et graves, l'étymologie d'aléa se rapporte à une *entreprise risquée*, dont il peut résulter des gains comme des pertes et dont les *chances* sont incertaines, comme dans cet extrait du *Tour du monde en quatre-vingts jours* : « Pendant les premiers jours qui suivirent le départ du gentleman, d'importantes affaires s'étaient engagées sur "l'aléa" de son entreprise. On sait ce qu'est le monde des parieurs en Angleterre, monde plus intelligent, plus relevé que celui des joueurs. Parier est dans le tempérament anglais. » (Verne, 1873).

9. Pour être tout à fait complet, il faudrait ajouter à ces deux catégories traditionnelles celle des dangers cosmiques que sont les corps célestes ou les tempêtes électromagnétiques, qui offrent un exemple intéressant de risques sociaux ayant toujours existé mais qui commencent seulement à être considérés comme tels.

La compréhension de ces phénomènes et de leurs effets suppose qu'ils soient décomposés en une succession d'événements liés par des relations de cause à effet. Or le choix de l'élément de cette chaîne qui doit être qualifié d'aléa ne s'impose pas naturellement. Si l'on veut privilégier la proximité de l'aléa avec les conséquences du phénomène pour l'environnement humain, alors il peut rarement être considéré comme une variable exogène dans la représentation du risque. Dans le cas d'un séisme, par exemple, la variable qui détermine les dommages aux structures de la façon la plus fiable est l'accélération au sol ; toutefois, pour être en mesure d'estimer la probabilité de réalisation de cet « aléa », il est nécessaire de reconstituer, à partir d'un événement initial, la propagation des ondes sismiques dans le sous-sol. Si l'on veut, en revanche, que l'aléa soit un événement situé en amont des mécanismes causaux étudiés, par exemple de manière à échapper à tout contrôle¹⁰, alors il faut admettre l'existence d'étapes supplémentaires, et donc de sources additionnelles d'incertitude, avant la détermination de ses effets. Ainsi, l'endiguement d'une rivière ne permet plus de considérer sa crue comme indépendante de l'action de l'homme. Pour définir un aléa exogène, il faut s'orienter vers les causes d'une crue qui déborderait la digue, telles que les précipitations. La crue devient, pour sa part, un événement intermédiaire dans la représentation du risque ; entre l'aléa et lui, d'autres facteurs tels que le ruissellement des sols entrent en jeu.

1.2.1.2 Conséquences

Les *conséquences* sont l'ensemble des pertes dues à l'aléa : les vies et la santé des personnes, les dommages aux actifs corporels (tels que les logements, les biens mobiliers, les infrastructures, les équipements, le patrimoine culturel ou environnemental) et incorporels (tels que l'image touristique), les perturbations de l'activité économique, les défaillances d'entreprises et les pertes d'emplois, la dislocation des structures sociales, la souffrance psychologique, etc.

Le concept de conséquences s'avère difficile à manier. Les méthodologies développées pour aider à sa mise en œuvre distinguent des conséquences directes, qui résultent de la seule action de l'aléa, et indirectes, qui font intervenir des actions ou des mécanismes supplémentaires, notamment de marché (ECLAC, 2003). Parmi les premières, certaines, comme le nombre de personnes tuées, blessées et déplacées dans un intervalle de temps limité après l'événement et le coût de remplacement du capital physique détruit, peuvent généralement¹¹ être évalués avec une précision suffisante *a posteriori*. D'autres types de conséquences restent très incertaines même longtemps après un désastre. Les estimations de coûts des catastrophes se sont longtemps limitées à une partie des conséquences directes, sans d'ailleurs que son périmètre soit toujours clairement indiqué. Il en est probablement

10. L'aléa est en effet parfois défini comme un phénomène naturel « entièrement indépendant de l'action de l'homme » (Ramade, 1993).

11. L'adverbe n'est pas superflu. À propos de glissements de terrain qui ont ravagé le Venezuela en 1999, les rapports et bases de données internationales font, aujourd'hui encore, état de 30 000 à 50 000 victimes, alors que selon l'enquête minutieuse d'un anthropologue à partir des registres de décès, « le nombre total de morts n'excède pas 700 » (Altez, 2007, cité par SIRC, 2009).

résulté non seulement une sous-estimation systématique de l'impact des catastrophes naturelles, mais aussi un biais en défaveur des plus importantes, puisqu'il semble que le rapport entre coûts directs et indirects diminue fortement lorsque la taille de l'événement augmente (Hallegate, 2008). Les années récentes, toutefois, ont vu l'éclosion d'une littérature sur les conséquences économiques à long terme des catastrophes. En faisant appel à des hypothèses théoriques supplémentaires, souvent incorporées dans des modèles économétriques, ces travaux ont cherché à estimer l'impact macro-économique des catastrophes (Freeman et al., 2002; Hallegate, 2008), la charge qu'elles créent pour les finances publiques (Mechler et al., 2006), ou encore leur influence sur la dynamique de développement urbain (Vigdor, 2008).

Par ailleurs, même une fois décomposé en autant de parties, le concept de conséquences dépend fortement de facteurs extérieurs, de sorte qu'il ne peut être employé à des usages prospectifs de mesure du risque ou d'identification d'actions susceptibles de le réduire. Il y aurait peu de sens, à titre d'exemple, de prévoir l'impact d'un séisme sur l'habitat dans un lieu à partir des dommages observés dans un autre lieu. Il est ainsi apparu nécessaire de représenter les dommages causés par un aléa comme la résultante de deux variables : l'exposition et la vulnérabilité (UNDRO, 1979).

1.2.1.3 Exposition

L'*exposition*, que l'on appelle également les *enjeux* du risque, désigne les biens humains, au sens le plus large, qui peuvent être affectés par l'aléa considéré.

Pour des raisons similaires à que celles que nous venons d'évoquer, la mesure des enjeux est le plus souvent cantonnée à un sous-ensemble des biens susceptibles de subir des dommages physiques directs du fait de l'aléa, particulièrement dans les analyses générales où le manque d'information est une contrainte forte. À titre d'exemple, l'indice mondial des risques de catastrophe développé par le Programme des Nations Unies pour le développement (PNUD) est fondé sur la seule prise en compte de la population vivant dans la zone concernée (PNUD, 2004). Dans l'évaluation des « points chauds » mondiaux pour les risques de catastrophe naturelle conduite il y a quelques années sous l'égide de la Banque Mondiale, les enjeux comprenaient, outre la population, le Produit Intérieur Brut (PIB) de la région affectée, comme mesure des revenus susceptibles d'être amputés (Dilley et al., 2005). Pour des applications plus spécifiques, cependant, l'ensemble des enjeux peut être étendu et affiné.

1.2.1.4 Vulnérabilité

La *vulnérabilité* est le degré de dommages encouru par les biens exposés en cas de survenue de l'aléa. En isolant une telle variable, on cherche à rendre compte de la variabilité de l'impact d'événements d'ampleur comparable selon la région et l'époque où ils se produisent. La notion est

cependant restée vague, et son pouvoir explicatif limité, tant qu'elle n'a pas été reliée à des facteurs concrets. Dans les travaux précités du PNUD et de la Banque Mondiale, à titre d'exemple, elle est définie comme une variable résiduelle, estimée à partir du ratio entre dommages et exposition observé dans le passé, ce qui ne permet guère d'en tirer des conclusions opérationnelles de façon rigoureuse.

Une littérature récente a cherché à identifier des déterminants de la vulnérabilité, en balayant un vaste spectre de facteurs allant de la qualité des infrastructures au capital humain ou social¹². Dans la mise à jour de l'indice des risques de catastrophe du PNUD, Peduzzi et al. (2009) ont effectué des régressions de la vulnérabilité passée par rapport à différentes variables socio-économiques, et lui ont substitué, dans l'expression finale de l'indice, celles dont l'effet ressortait le plus significativement. Il s'agit du revenu (PIB) par habitant et de plusieurs indicateurs environnementaux, tels que le taux de croissance urbaine ou le pourcentage de terres arables.

Mais le concept semble malgré tout trop vaste pour pouvoir faire l'objet d'analyses précises, et une tendance marquée de la littérature récente a été de chercher à le décomposer. Le glossaire du quatrième rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat présente ainsi la vulnérabilité d'un système comme une combinaison de sa *sensibilité*, qui mesure son degré d'affectation *a priori* par le changement climatique, et de sa *capacité d'adaptation*, c'est-à-dire de transformation des effets du changement climatique à son avantage (Parry et al., 2007). Des concepts voisins, tels que la résilience ou la capacité à surmonter¹³ ont également été proposés¹⁴ sans que leur contenu soit toujours très bien défini, ce qui a contribué à faire régner une certaine confusion dans ce domaine. Ce courant de littérature a cependant ouvert un champ de réflexion important sur les facteurs qui permettent de limiter le degré de dommage supporté par la société. L'une des questions soulevées, qui nous intéresse particulièrement ici, est l'évaluation de l'influence des institutions et politiques publiques *existantes* sur le risque (Lahidji, 2010).

1.2.2 Santé publique

Le domaine de la santé publique s'est formé par intégration successive de sujets de préoccupation relative à l'état de santé de la population : cantonné aux maladies infectieuses et à certaines maladies professionnelles au XIXe siècle, il s'est étendu aux pathologies chroniques avec la montée des maladies cardio-vasculaires et des cancers à partir du milieu du XXe siècle¹⁵, puis aux problèmes de santé liés à l'environnement, notamment à la pollution, à partir des années 1970 et enfin,

12. Pour une revue désormais quelque peu datée, on pourra se reporter à Birkmann (2006).

13. Traduction de l'anglais *coping capacity*.

14. Voir par exemple (Briguglio et al., 2009).

15. L'acte de renaissance de l'épidémiologie fut la double étude de Richard Doll et Austin Bradford Hill sur les relations entre la consommation de tabac et d'une part le cancer du poumon, de l'autre les maladies cardio-vasculaires (Doll et Hill, 1950 ; Doll et Hill, 1954). Le retentissement de ces études, notamment de la seconde (connue sous le nom de *British doctors study*) ouvrit un nouveau champ à l'investigation épidémiologique.

récemment, aux comportements alimentaires et aux habitudes de vie.

Les analyses et les documents de référence portant sur l'ensemble de ce vaste domaine sont rares. La loi du 9 août 2004, qui est la première loi générale à être adoptée dans ce domaine en France depuis celle sur l'hygiène publique du 15 février 1902, en est un exemple. Elle a notamment instauré l'adoption par l'Etat d'objectifs de santé publique quinquennaux, et lancé cinq plans stratégiques nationaux tels que le plan national de lutte contre le cancer. Le champ de la politique de santé publique y est divisé en dix parties, dont trois répondent à notre caractérisation provisoire du risque social¹⁶ (article 2) : « la lutte contre les épidémies » ; « la prévention des maladies, des traumatismes et des incapacités » ; « l'identification et la réduction des risques éventuels pour la santé liés à des facteurs d'environnement et des conditions de travail, de transport, d'alimentation ou de consommation de produits et de services susceptibles de l'altérer »¹⁷. La loi précise qu'elle vise la santé à l'échelle de la population, auquel il est indispensable « de partir de l'analyse des problèmes de santé et de l'identification des éléments qui déterminent leur survenue ou leur aggravation, de spécifier clairement les résultats attendus, de mettre en œuvre les stratégies d'action qui sont les plus susceptibles d'être efficaces compte tenu des moyens disponibles et de suivre les résultats obtenus » (annexe contenant le rapport d'objectifs de santé publique). La loi ne relie pas ces « problèmes de santé », « leur survenue ou leur aggravation » à la notion de risque, auquel elle fait pourtant appel pas moins de 106 fois, par exemple pour énoncer la mission de veille et de vigilance sanitaires de l'Institut National de Veille Sanitaire (INVS) : « De rassembler, analyser et actualiser les connaissances sur les risques sanitaires, leurs causes et leur évolution [...] » (article 15 alinéa 2.a).

L'analyse du texte montre une diversité d'influences disciplinaires qui peut expliquer l'ambivalence de sa terminologie. Bien que le domaine de la santé publique corresponde fidèlement au champ d'étude de l'épidémiologie, d'autres disciplines et approches y ont contribué à la définition des problèmes de santé et des modalités d'action. C'est notamment le cas de la toxicologie et de la gestion de crise, dont l'empreinte est visible dans la loi française sur la santé publique.

1.2.2.1 Épidémiologie : facteurs de risque et fréquence des maladies

Les rares considérations terminologiques ou méthodologiques de la loi sont inspirées par l'épidémiologie. Ainsi : « Il est difficile actuellement de connaître la part des facteurs d'environnement collectif (agents physiques, chimiques et biologiques) dans les phénomènes morbides [...]. L'importance de chaque déterminant est appréciée en termes de risque attribuable : la fraction (ou proportion) du risque attribuable à un facteur de risque mesure l'impact de ce facteur de risque dans la population, c'est-à-dire la fraction de tous les cas d'une maladie qui sont dus au facteur

16. Les autres points ne sont pas étrangers au risque, mais traitent des modalités de sa prise en charge par la collectivité.

17. Loi n° 2004-806 du 9 août 2004 relative à la politique de santé publique, *Journal Officiel de la République Française*, n°185 du 11 août 2004, p. 14277.

concerné ou, en d'autres termes, la proportion qui peut être évitée si le facteur de risque n'était pas présent. Ainsi, on estime qu'environ 80 % des cancers bronchiques sont attribuables au tabac » (annexe contenant le rapport d'objectifs de santé publique).

Le risque correspondrait ainsi à la fréquence de la maladie, *c'est-à-dire qu'il serait homogène à une probabilité*, et les facteurs de risque seraient ses causes identifiées. Dans leur ouvrage de référence en épidémiologie, Rothman, Greenberg et Lash (2008) confirment cette interprétation tout en nous mettant en garde contre la pluralité des usages du terme dans leur discipline :

"Unfortunately, in epidemiologic parlance, even in the scholarly literature, 'risk' is frequently used for many distinct concepts : rate, rate ratio, risk ratio, incidence odds, prevalence, etc." (p.9).

Même si l'on s'en tient à cette première définition, la mesure du risque conduit à des questions délicates : la fréquence de la maladie doit-elle être mesurée par le nombre de nouveaux cas observés à partir d'un point de référence donné (*incidence*) ou à partir du nombre total de cas (*prévalence*) ? La première option est celle qui se prête le mieux à la recherche de causes. À titre d'exemple, si l'on retient la seconde, alors une augmentation du taux de survie des malades se traduit, à incidence donnée, par une augmentation du risque. Pourtant, nombre d'études épidémiologiques se fondent sur la prévalence et, dans certains cas, l'information sur l'incidence n'est pas accessible.

Les *facteurs de risque* sont des déterminants biologiques (âge, sexe, tension artérielle), génétiques, comportementaux (régime alimentaire), environnementaux (lieu de résidence ou de travail) ou encore psychologiques, que l'on peut considérer aux niveaux individuel ou collectif, et dont on cherche à mettre en évidence les effets par des méthodes statistiques. Parmi celle-ci, la plus importante est l'essai contrôlé, que nous avons brièvement décrite dans l'introduction.

Les deux notions complétant la mesure du risque sont l'*exposition*, qui désigne le degré de présence d'un facteur de risque, et l'*effet*, défini comme la variation de la fréquence de la maladie associée à une variation de l'exposition.

À partir de l'ensemble de ces éléments, la modélisation épidémiologique peut permettre de rendre compte de l'évolution de la fréquence de la maladie et de simuler l'effet de différentes actions.

Il faut noter, enfin, que le domaine de l'épidémiologie s'est considérablement étendu au cours des dernières décennies, et que cette expansion s'est accompagnée d'une dérive lexicale marquée. La notion d'épidémie, d'abord étendue des maladies infectieuses aux maladies chroniques, a progressivement été appliquée aux conditions de santé, comme dans le terme « épidémie d'obésité », et même aux comportements. C'est ainsi qu'à l'occasion de la journée mondiale sans tabac, trois administrateurs de l'Organisation Mondiale de la Santé ont récemment appelé à « contrer l'épidémie mondiale de tabagisme » (Bettcher, Smith et Perucic, 2008). De façon similaire, l'usage du substantif risque s'est transmis des causes (*facteurs de risque*) aux catégories de population parmi lesquelles

ces causes sont particulièrement présentes (*groupes* à risque pour des maladies infectieuses), puis à des comportements d'exposition, qu'ils soient délibérés (*usages* à risque d'alcool, sports à risques, *conduite* à risque) ou non (consommations à risque pour le cancer). Cette évolution dénote une tendance à chercher toujours plus en amont les causes de la pathologie, qui renvoie à la question difficile et ancienne de séparation du normal et du pathologique (Canguilhem, 1966).

1.2.2.2 Gestion des crises sanitaires : menace et impact

La terminologie de la loi sur la santé publique change lorsqu'elle traite de situations où le danger sanitaire est imminent. Ainsi :

« L'institut informe sans délai le ministre chargé de la santé en cas de *menace* pour la santé de la population ou de certaines de ses composantes, quelle qu'en soit l'origine, et il lui recommande toute mesure ou action appropriée pour prévenir la réalisation ou atténuer l'*impact* de cette menace » (article 15).

« En cas de *menace sanitaire grave* appelant des mesures d'urgence, notamment en cas de menace d'épidémie, le ministre chargé de la santé peut, par arrêté motivé, prescrire dans l'intérêt de la santé publique toute mesure proportionnée aux *risques courus* et appropriée aux circonstances de temps et de lieu afin de prévenir et de limiter les *conséquences des menaces* possibles sur la santé de la population. [...] En cas d'urgence, notamment de *danger ponctuel imminent* pour la santé publique, le représentant de l'Etat dans le département peut ordonner l'exécution immédiate, tous droits réservés, des mesures prescrites par les règles d'hygiène prévues au présent chapitre » (article 18).

« Lorsqu'il s'avère nécessaire de prévenir ou de maîtriser des *risques graves* pour la santé humaine [...] » (article 21).

Ces extraits nous ramènent à une conception plus traditionnelle du risque social. La *menace* ou le *danger* semblent être des sources clairement identifiés, uniques et *externes* au corps social. Si les risques mentionnés peuvent éventuellement être encore interprétés comme des probabilités, il ne fait pas de doute que la prise en compte des *conséquences* possibles est un élément important du problème.

1.2.2.3 Toxicologie : agent et gravité

La loi sur la santé publique emprunte le vocabulaire de la toxicologie lorsqu'elle s'intéresse plus spécifiquement à la source du risque qu'à la population exposée, par exemple lorsqu'elle décrit le contenu du plan national de prévention des risques pour la santé liés à l'environnement : « Ce plan prend notamment en compte les effets sur la santé des agents chimiques, biologiques et physiques présents dans les différents milieux de vie, y compris le milieu de travail » (article 53).

Nous disposons, dans ce domaine, de nombreuses définitions du risque provenant de législations

et d'agences sanitaires. À titre d'exemple, le règlement européen n° 178/2002 connu sous le nom de « loi sur l'alimentation », énonce : « Aux fins du présent règlement, on entend par [...] "danger", un agent biologique, chimique ou physique présent dans les denrées alimentaires ou les aliments pour animaux, ou un état de ces denrées alimentaires ou aliments pour animaux, pouvant avoir un effet néfaste sur la santé [...] "risque", une fonction de la probabilité et de la gravité d'un effet néfaste sur la santé, du fait de la présence d'un danger »¹⁸.

Pour l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES), le risque est « la probabilité de survenue du danger (c'est-à-dire de l'effet toxique) ; il est caractérisé en confrontant les résultats de l'estimation de la relation dose-réponse avec ceux de l'estimation des expositions »¹⁹.

L'Agence américaine de protection de l'environnement (EPA) décrit le risque comme « une mesure de la probabilité que des dommages à la vie, à la santé, aux biens ou à l'environnement se produisent comme conséquence d'un danger », ce dernier comme « la capacité potentielle d'une radiation, d'un produit chimique ou d'un autre polluant à entraîner une maladie ou blessure humaine » et l'agent comme « toute entité physique, chimique ou biologique qui puisse être dommageable pour un organisme »²⁰.

L'agence sanitaire anglaise, the Health and Safety Executive, propose des définitions similaires : « le risque est la probabilité qu'une personne ou qu'un bien soient affectés d'une manière spécifique par un danger » et le danger est « la capacité de dommage liée à une propriété intrinsèque ou la disposition d'une chose à causer des pertes » (HSE, 1999).

Au-delà des nuances et des imprécisions éventuelles, le *risque* apparaît ici comme une *probabilité* de survenue d'un *effet* bien spécifié, lié à la présence d'un *danger*, propriété intrinsèque d'un *agent* physique, chimique ou biologique.

Les effets visés sont clairement indiqués par l'EPA : « En général, un effet s'ensuit inévitablement d'un antécédent (cause ou agent). Un effet biologique est le résultat biologique de l'exposition à un agent causal. Ce terme est similaire à celui de réponse, mais il met l'accent sur l'agent actif (comme dans l'effet du cadmium) plutôt que le récepteur qui lui répond (comme dans la réponse de la truite) »²¹.

L'outil principal pour la détermination des effets toxicologiques est la relation dose-réponse

18. Règlement n° 178/2002 du Parlement européen et du Conseil du 28 janvier 2002 établissant les principes généraux et les prescriptions générales de la législation alimentaire, instituant l'Autorité européenne de sécurité des aliments et fixant des procédures relatives à la sécurité des denrées alimentaires, Journal officiel de l'Union européenne, L 031, 1 février 2002.

19. Il s'agit plus précisément de l'un des prédécesseurs de l'ANSES, l'Agence française de sécurité sanitaire de l'environnement et du travail (AFSSET). Source : fiche toxicologie, www.afssa.fr/ET/DocumentsET/fiche_afsset_toxicologie.pdf.

20. Environment Protection Agency, Terms of environment glossary, http://iaspub.epa.gov/sor_internet/registry/termreg/home/overview/home.do.

21. Environment Protection Agency, Causal analysis/diagnosis decision information system glossary, http://iaspub.epa.gov/sor_internet/registry/termreg/home/overview/home.do.

évoquée par l'ANSES, qui retrace la façon dont un organisme se comporte lorsque son exposition à une substance toxique change. L'un des enjeux importants des relations dose-réponse est de déterminer la dose qui peut être estimée sans effet pour l'organisme²². Leur évaluation peut reposer sur des études statistiques d'effets observés, comme dans le cas des analyses radio-toxicologiques des conséquences sanitaires des bombardements de Hiroshima et Nagasaki en 1945, qui fondent une grande partie des connaissances actuelles en matière de conséquences à long terme des rayonnements ionisants, ou sur des expérimentations *in vivo*. Dans les cas fréquents où l'on cherche à extrapoler à l'homme les résultats d'expérimentations animales, l'usage est d'appliquer à la dose un coefficient multiplicateur inférieur à un, dit facteur d'incertitude, pour tenter de tenir compte des différences entre espèces.

1.2.3 Accidents industriels et technologiques

Si l'attention à la sûreté des ouvrages et des installations dangereuses a fait de longue date partie des prérogatives des pouvoirs publics, elle a connu un très net regain avec le développement des industries chimique et nucléaire.

À titre d'exemple, la directive européenne 96/82/CE, dite « Seveso II », qui concerne les installations dangereuses autres que nucléaires, a constitué une étape importante dans le développement des politiques publiques des pays européens en matière de risque. Le *risque* y est défini comme « la probabilité qu'un effet spécifique se produise dans une période donnée ou dans des circonstances déterminées » (article 3)²³. L'*effet spécifique* dont il est principalement question dans le document est un accident majeur, défini comme « un événement tel qu'une émission, un incendie ou une explosion d'importance majeure résultant de développements incontrôlés survenus au cours de l'exploitation d'un établissement couvert par la présente directive, entraînant pour la santé humaine, à l'intérieur ou à l'extérieur de l'établissement, et/ou pour l'environnement, un danger grave, immédiat ou différé, et faisant intervenir une ou plusieurs substances dangereuses ». Les substances dangereuses visées par la Directive sont listées en annexe. Quant à la notion de *danger*, elle désigne « la propriété intrinsèque d'une substance dangereuse ou d'une situation physique de pouvoir provoquer des dommages pour la santé humaine et/ou l'environnement ».

Le risque est donc ici défini comme la probabilité de survenue d'un événement représentant lui-même un danger pour la santé humaine ou l'environnement, *indépendamment de ses conséquences finales*. La description détaillée des circonstances de survenue de l'événement montre par ailleurs que le risque est ici considéré à l'intérieur des limites d'une installation, c'est-à-dire d'un système technique.

22. Plus précisément, sans effet *observé*.

23. Directive 96/82/CE du Conseil du 9 décembre 1996 concernant la maîtrise des dangers liés aux accidents majeurs impliquant des substances dangereuses, Journal officiel de l'Union européenne, L 010, 14 janvier 1997.

Dans leur article de référence sur la terminologie du risque dans les installations nucléaires et plus généralement les systèmes techniques, Kaplan et Garrick (1981) définissent le risque comme la donnée de l'ensemble des *scénarios* décrivant les évolutions dommageables du système, accompagnés de leurs probabilités de réalisation et des conséquences qui leur seraient attachées. Ceci suppose, comme le notent les auteurs, que l'on dispose au préalable d'une description complète des états du système, ou en tout cas ceux qui peuvent conduire à un événement dommageable, que l'on construise à partir de ces états les scénarios formant le risque, et que l'on soit en mesure de leur attribuer des probabilités et des conséquences.

L'analyse des systèmes, que nous aborderons en détail au chapitre 3, est la discipline technique qui s'intéresse à l'identification d'états critiques résultant de la défaillance d'un composant physique, d'interactions anormales entre composants, d'erreurs humaines ou encore d'événements extérieurs. L'une de ses particularités, par rapport aux autres méthodes discutées, est que son objet est défini par des frontières relativement fermes. Elle a donc pu se donner pour ambition de construire des représentations fiables de tous les scénarios pertinents à l'intérieur de ces frontières. Elle a par la suite été progressivement conduite à élargir cette base en recherchant les causes des dysfonctionnements au-delà des frontières traditionnelles des systèmes techniques, par exemple dans les facteurs organisationnels et humain. En ce sens, l'approche a constitué une avancée importante pour l'analyse causale.

Le risque étant ainsi défini, il lui est fréquemment associé une courbe de risque, qui est la fonction de répartition cumulative des conséquences (pour l'ensemble des scénarios) (Helton, 1994).

1.2.4 Autres cas qui relèvent ou ne relèvent pas du risque social

1.2.4.1 Actes de malveillance, troubles sociaux et guerres

En 1914, les navires marchands américains engagés dans le ravitaillement des puissances alliées et soumis aux attaques des sous-marins allemands ne parvenaient pas à s'assurer auprès de compagnies privées. Pour palier cette défaillance, le Congrès adopta dès le mois de septembre 1914 le War risk insurance act, qui créait une assurance publique pour les navires et leurs équipages. L'assurance contre les détournements et actes de guerre se développa après la seconde guerre mondiale pour le transport aérien et maritime, tout en faisant l'objet de conditions particulières comme l'exclusion des pertes liées à un usage d'armes de destruction massive et une clause d'annulation unilatérale de la police par l'assureur avec un préavis de 7 jours (Hoeven, 2005). Pour Kunreuther et Michel-Kerjan (2005), ces restrictions traduisaient la difficulté des compagnies d'assurance à estimer le risque couvert.

Presque un siècle plus tard plus tard, les pertes sans précédent enregistrées suite aux attentats du 11 septembre 2001 amenèrent l'ensemble des compagnies d'assurance à faire jouer les clauses

d'annulation pour se retirer du marché des risques d'actes de guerre. Le secteur assurantiel dut ainsi être une nouvelle fois relayé par des États placés en situation d'« assureurs en dernier ressort ».

Des efforts ont néanmoins été engagés pour mieux évaluer ces risques. De nouveaux modèles d'évaluation du risque ont été développés, dont le trait commun est de limiter le recours aux probabilités subjectives collectées auprès d'experts en intégrant de façon plus systématique l'information existante et en explorant les relations causales en amont des événements (RMS, 2011). La terminologie a été clarifiée, différentes catégories d'événements ont été distinguées afin d'améliorer l'homogénéité des risques, et des liens ont été établis entre les structures des organisations terroristes et les formes et fréquences de leurs attaques. Par ailleurs, les modèles de théorie des jeux appliquée aux problèmes de stratégie militaire, développés à partir des travaux pionniers de (Schelling, 1960), ont été mis à contribution pour représenter les stratégies d'attaque et de défense intervenant dans la détermination du risque (Major, 2002).

En réaction à la réévaluation des risques liés au terrorisme à la suite des attentats du 11 septembre 2001, les autorités publiques ont également engagé une réflexion sans précédent sur leur représentation et leur estimation. Des études préliminaires ont cherché à élaborer des modèles estimatifs simples ou à appliquer les modèles de l'industrie de l'assurance au risque social posé par le terrorisme à travers les États-Unis (Willis et al., 2005). Le département américain de la sécurité intérieure (DHS) a lancé des programmes spécifiques d'évaluation des risques liés aux formes les plus graves d'actes terroristes. Les méthodologies mises sur pied dans ces programmes a été rendues publiques, et la communauté scientifique s'est investie dans leur analyse critique et leur amélioration (National Research Council, 2008 ; Piegorsch, Cutter et Hardisty, 2007).

Ces travaux posent pas à pas les fondements d'une analyse systématique des actes de violence comme sources de dommages incertains auxquels la société doit faire face. Il s'agit donc bien de risques sociaux.

1.2.4.2 Dérèglements socio-économiques

Les bouleversements économiques tels que les grandes faillites, les licenciements massifs, le déclin industriel, les dépressions, la déflation et les récessions, et enfin crises financières systémiques font également partie des phénomènes sociaux comportant à la fois une incertitude et des possibilités de perte substantielles.

La gestion des crises financières, à titre d'exemple, peut être analysée dans les mêmes termes que celle d'une crise épidémique (Freixas et al., 2000), à laquelle elle a d'ailleurs emprunté le terme de contagion. L'analyse de l'intervention nécessaire d'une autorité publique agissant en qualité de « prêteur en dernier ressort » dans cette gestion remonte à Thornton (1802) et Bagehot (1873), ce qui plaide pour la reconnaissance de la crise financière systémique comme un vieux risque social.

Dans l'analyse du risque et des moyens d'y faire face, les mécanismes causaux de la crise sont décrits de façon précise et font l'objet de recherches, depuis la transmission du mouvement de panique et son effet sur la masse monétaire jusqu'aux incitations négatives liées à l'intervention publique (ou aléa moral).

Le déclin d'une industrie, la faillite d'une grande entreprise nationale voire le retournement conjoncturel doivent-ils également être considérés comme des risques sociaux ? Il semble bien que la réponse ait varié selon les époques, les pays et les doctrines politiques et économiques. S'il semble aujourd'hui assez largement admis, au moins dans la plupart des pays de l'OCDE, que de tels événements constituent des risques privés qui n'engagent pas les responsabilités de l'État, c'est parce qu'ils sont produits par ce que l'on considère être le fonctionnement normal de l'économie et, plus spécifiquement, des marchés.

Une telle analyse lie l'appréciation du caractère social d'un risque économique à une représentation du fonctionnement de l'économie, de la même façon que l'évaluation d'un problème de santé publique dépend du jugement porté sur l'état de santé « normal » de la population. Sur des marchés complets où règne une concurrence pure et parfaite, il ne semble guère y avoir de place pour un risque collectif. Il est intéressant de noter que les récents plaidoyers en faveur d'actions plus marquées des autorités de régulation afin de prévenir les risques liés au fonctionnement des marchés financiers ont recouru à des arguments d'imperfection de ces marchés tels que l'imperfection de la concurrence ou l'omniprésence de comportements mimétiques, même en dehors des épisodes de crise systémique. Ainsi, la modification d'une hypothèse relative aux interactions micro-économiques transforme un risque privé en risque social, fournit une justification à l'intervention publique et décrit le levier sur lequel cette dernière peut s'appuyer.

1.2.4.3 Les risques sociaux au sens de la sécurité sociale

La dernière étape de notre tour d'horizon concerne les risques que l'on appelle sociaux au sens de la sécurité sociale.

Historiquement, la protection sociale s'est développée à partir de la couverture des risques de perte de revenu des travailleurs liés à la maladie, aux accidents du travail ou à la vieillesse. De ces trois risques, seuls les deux premiers comportent une réelle incertitude, les mécanismes de retraite s'apparentant simplement à une redistribution de revenus entre générations (régimes par répartition) ou entre périodes de la vie (régimes par capitalisation). Cependant, même pour les deux premiers risques, la question de la séparation entre risque privé et risque social nous mène une nouvelle fois à celle de la distinction entre le normal et le pathologique. Ainsi, pendant une grande partie du XIXe siècle, les accidents du travail et les problèmes chroniques de santé liés à l'exercice professionnel furent considérés comme des aléas relevant de décisions individuelles. Selon

l'esprit de l'époque, en choisissant leur travail, les employés en acceptaient les risques et étaient rémunérés pour cela (Tunc, 1990). Au cours du XXe siècle, au contraire, la protection sociale a progressivement étendu son champ à des risques qui n'avaient plus de lien direct avec le statut de travailleur : l'incapacité à répondre aux besoins de la famille (allocations familiales, soutien aux parents isolés, aides à l'éducation), la dépendance due au grand âge, etc.

Ainsi, le Système européen de comptes considère conventionnellement comme risque ou besoin social : la maladie ; l'invalidité et l'infirmité ; les accidents du travail et maladies professionnelles ; la vieillesse ; la survie ; la maternité ; la famille ; la promotion de l'emploi ; le chômage ; le logement ; l'éducation ; et l'indigence. Le Système définit alors la protection sociale comme l'ensemble des mécanismes de prévoyance en relation avec l'un de ces *risques ou besoins*²⁴. Ces mécanismes (ou régimes) peuvent être distingués selon qu'ils relèvent du principe de l'assurance sociale ou de celui de l'assistance sociale. Dans le premier cas, le régime est organisé pour le compte d'un ensemble de personnes affiliées, et il existe une correspondance entre les prestations qui leur sont délivrées et les contributions qui leur sont (ou ont été) demandées. Dans le second cas, les prestations ne sont conditionnées ni par l'affiliation ni par le versement de contributions.

Dans sa définition contemporaine, la protection sociale ne lie donc plus le risque à la réalisation d'un événement particulier de nature accidentelle, mais à la capacité de tout individu d'accéder à un niveau de vie considéré comme un minimum social. Seule une partie de son activité correspond donc à la définition que nous avons retenue ici.

1.3 Proposition de synthèse

Quelques observations tirées de cette vue d'ensemble nous permettent à présent de préciser notre idée du risque social à partir de quatre composantes : une source d'incertitude, des conséquences, des externalités et, comme préalable à ces éléments, une représentation.

1.3.1 Une représentation

Dans tous les domaines que nous avons passés en revue, la prise en compte de l'incertitude est relative à une description d'un contexte en termes de relations causales, ce que Cooke et Goossens (1999) appellent la « structure du cas » (*case structure*).

Comme nous l'avons vu, les représentations du risque ne se résument généralement pas à l'action directe d'un aléa exogène sur des conséquences déterministes. Ceci constitue un problème si l'on veut rester attaché au diptyque (aléa, conséquence)²⁵, mais n'en est plus un dès lors que l'on

24. Système Européen de Comptes SEC95, §4.84.

25. Sarewitz, Pielke et Keykhah (2003) proposent par exemple de distinguer un risque d'événement (*event*) et un risque de résultat (*outcome*).

reconnaît l'existence de différentes sources d'incertitude dans une représentation du risque. Pour être plus précis, il nous faut donc considérer l'ensemble des sources d'incertitude portant sur les différents maillons des chaînes causales constituant la représentation du risque : incertitudes relatives aux précipitations, au ruissellement des sols (à précipitations données) et au débordement d'un cours d'eau (à précipitations et ruissellement donnés). Dans ce cas, et dans ce cas seulement, la représentation de l'incertitude peut être considérée comme exogène.

Les limites de cette représentation ne peuvent pas être établis de façon générale. Celle-ci doit plonger aussi profondément que nécessaire et possible vers les causes pour parvenir à une évaluation satisfaisante de l'incertitude, et s'étendre aussi loin qu'il le faut pour intégrer les types de conséquences que l'on veut considérer. Les conséquences peuvent être éloignées du contexte temporel et spatial dans lequel un événement initial tel qu'un accident est considéré. Le lien qui les lie à l'événement initial peut être direct (dommages physiques directement causés) ou non (perte de revenu, impact économique à long terme, accidents provoqués par la survenue de l'aléa). Naturellement, plus les conséquences sont distantes de l'événement initial, plus les sources additionnelles d'incertitude sont nombreuses.

Cette structure causale est donc nécessaire à la compréhension des mécanismes par lesquels des facteurs extérieurs peuvent affecter des grandeurs qui nous intéressent. Il serait erroné de penser qu'elle le produit de l'activité d'évaluation du risque ; elle en est un préalable, en ce que sans elle, le risque ne peut pas être défini.

L'une des questions que nous rencontrerons fréquemment au cours des prochains chapitres est celle de l'incertitude relative à cette représentation elle-même, qui constitue une sorte de second niveau d'incertitude. Dans une interprétation possible, pour laquelle nous pouvons d'emblée avouer notre préférence, ces deux niveaux peuvent correspondre à la célèbre distinction opérée par Knight (1921) entre le risque, représentable par des distributions de probabilités, et l'incertitude, d'une nature plus générale. Comme nous le verrons, la prise en compte du risque social par le biais de représentations causales a été alimentée par un processus de production, de collecte et d'organisation de données et de connaissances permettant de transformer pas à pas l'incertitude knightienne en risque probabiliste.

1.3.2 Une source d'incertitude

L'une des composantes du risque sera donc un ensemble de sources d'incertitude, que nous pourrions désigner dans la suite comme aléa ou danger lorsque le contexte s'y prêtera, mais que nous considérerons plus généralement comme l'ensemble des variables exogènes à la représentation du risque. L'incertitude porte sur une réalisation particulière de ces variables au cours d'une période donnée. Les événements en question peuvent avoir une définition intrinsèque (comme une

explosion) ou relatives à un état de référence (comme une sécheresse ou une épidémie, qui représentent des écarts à des situations considérées normales). Dans les deux cas, les événements doivent être identifiables, c'est-à-dire que les situations qu'ils représentent doivent pouvoir être entièrement caractérisées.

1.3.3 Des conséquences

Les conséquences sont une mesure de l'action des variables exogènes sur un ensemble de grandeurs considérées, telles la vie et la santé d'individus, des actifs environnementaux, économiques ou culturels, etc. Lorsqu'il y aura lieu, nous pourrions considérer les conséquences comme le produit d'une exposition (ou d'enjeux) et d'une vulnérabilité.

1.3.4 Des externalités

La coexistence de sources d'incertitude et de conséquences possibles ne suffit cependant pas à caractériser le risque social. En particulier, la différence entre risque social et risque privé ne semble pas découler de la seule ampleur des conséquences redoutées. À titre d'exemple, comme nous l'avons vu, les fluctuations des marchés financiers peuvent se traduire par des transferts de richesse considérables, et parfois même par des pertes agrégées substantielles, sans que l'État intervienne, ni que les risques correspondants puissent être qualifiés de sociaux. Dans un domaine différent, la consommation de tabac a pu provoquer des millions de cancers jusqu'à la fin des années 1980 sans que les sociétés semblent considérer que ce problème relevait d'une action collective.

1.3.4.1 Deux formes principales

Outre les deux éléments initiaux, les risques sociaux mettent donc en jeu des dépendances entre agents que la théorie économique englobe dans la classe des externalités. Le cas le plus manifeste est celui où les décisions d'un agent influencent le risque auquel d'autres sont soumis. Lorsqu'il détermine ses investissements de sûreté, l'opérateur d'un site industriel dangereux détermine aussi le risque subi dans son voisinage. S'il ne tient pas compte de cette externalité, son effort de prévention est inférieur (et le risque est supérieur) à ce qui serait optimal du point de vue de la société. Des externalités similaires, mais positives, peuvent résulter de décisions d'agents également concernés par le risque. En votant en faveur d'un projet de construction de digue financé par les impôts locaux, chaque habitant d'une agglomération située en zone inondable choisit de contribuer à la réduction du risque d'inondations pour l'ensemble de l'agglomération.

Certains risques sociaux ne relèvent cependant pas de ce type d'externalités. De nombreux mécanismes assurantiels, comme par exemple le régime français d'indemnisation des victimes de catastrophes naturelles, regroupent des individus aux profils de risque très différents. Que des individus

faiblement exposés à un aléa acceptent de mutualiser leur risque avec celui d'individus beaucoup plus exposés voire, dans certains cas, prendre unilatéralement en charge les risques d'autres²⁶, peut être soit interprété comme le signe d'un manque d'information ou d'un comportement irrationnel, soit, et c'est l'explication mise en avant par l'État français, comme une expression de solidarité constitutive d'une société. Cette solidarité manifeste l'existence d'une autre forme d'externalité : celle où l'un des déterminants de l'utilité d'un agent (sa santé, son logement, etc.) influencent celle d'un autre.

1.3.4.2 Des tensions entre individu et collectivité

La prise en compte des externalités ne suffit pas toujours à établir une ligne de démarcation claire entre risques sociaux et risques privés et à offrir des fondements normatifs sûrs aux interventions publiques relatives aux premières. À titre d'exemple, la réglementation concernant le port de la ceinture de sécurité pour les conducteurs de voitures ou l'affichage obligatoire de messages dissuasifs sur les paquets de cigarettes visent des risques où seul le sort de l'individu semble en jeu. Dans la mesure où sa décision n'affecterait que sa propre utilité, l'individu ne ferait-il pas face à un risque privé ? Sur quel principe se fonderait, dans ce cas, l'action publique ? Le droit ne semble pas apporter, en la matière, de réponse définitive (Noiville, 2003). L'arrêt du Conseil d'Etat confirmant la légalité de l'obligation faite aux conducteurs d'automobiles de porter la ceinture de sécurité présentait comme seule justification « l'amélioration globale de la sécurité des occupants des voitures particulières ».

Si l'on laisse de côté les justifications fondées sur les limitations des individus dans l'appréciation du risque (mauvaise évaluation des probabilités, addiction, etc.), l'économie normative semble avoir deux arguments principaux en faveur d'une intervention du décideur public dans de tels cas. En premier lieu, il s'agit de sa capacité à s'abstraire de l'incertitude qui affecte les choix individuels. Si l'incertitude inhérente au risque pouvait être levée, les individus exprimeraient probablement d'autres préférences que celles qu'ils manifestent en présence de risque (Hammond, 1981). Le conducteur qui saurait qu'il va subir un accident grave choisirait spontanément de porter sa ceinture de sécurité. À l'instar des individus, le décideur social connaît à l'avance l'ampleur probable des pertes liées à un aléa ; mais à leur différence, la distribution de ces pertes ne lui importe pas. Lui seul peut donc exprimer d'emblée des préférences informées (Fleurbaey, 2010). La seconde piste est liée à la seconde forme d'incertitude que nous avons mise en avant : dès lors qu'un individu accepte implicitement la prise en charge de ses risques privés par la société, ces derniers deviennent sociaux et le décideur public est en droit d'intervenir pour prévenir les dommages associés.

Ces arguments incontestables doivent cependant être mis en regard des conséquences d'un tel

26. On peut penser au cas d'un sédentaire démunie acceptant d'indemniser, par le biais de mécanismes publics, un riche alpiniste.

interventionnisme pour la liberté de choix des individus face aux risques (Hammond, 1983 ; Kolm, 1998).

1.3.4.3 Une notion relative

Il faut enfin rappeler le caractère relatif de la notion d'externalités : les préférences évoluent et l'appréciation des externalités fait de même.

Nous avons observé à plusieurs reprises que la définition du risque dépendait d'un niveau de référence considéré comme « normal ». Ceci renvoie à une différence classique entre les approches du risque comme possibilité de perte et celles du risque comme incertitude concernant des conséquences, qui n'excluent pas la possibilité de gains. Au niveau individuel, il est maintenant bien établi que ce choix n'est pas neutre dans l'attitude des individus face au risque. Tversky et Kahneman (1981) ont montré que les individus évaluaient différemment une mesure de santé publique selon que ses résultats étaient présentés en termes de possibilités de gain ou de perte de vies ; ils en ont déduit que les choix individuels dépendaient du « cadrage » (*framing*) des problèmes. Leur théorie des perspectives (*prospect theory*) décrit un processus de décision dont la première étape est précisément la définition d'un point de référence (Kahneman et Tversky, 1979). Le comportement d'aversion pour la perte a depuis été mis en évidence dans un grand nombre de situations réelles, depuis les paris hippiques jusqu'aux heures travaillées par les chauffeurs de taxi²⁷. Il semble ne pas se résumer à la sensibilité des individus à la façon dont des perspectives incertaines sont présentées, mais relever plus fondamentalement d'une différence d'attitude entre ce que l'on considère comme acquis et ce qui serait à acquérir (Kahneman, Knetsch et Thaler, 1990).

Notre observation correspond donc à l'existence d'un point de référence dans la définition des risques sociaux. Des formes d'aversion pour la perte ont par ailleurs été observés chez les décideurs en charge de risques sociaux. Dans les procédures d'autorisation de mise sur le marché de médicaments, par exemple, la démonstration de l'innocuité du médicament pèse d'un poids beaucoup plus grand que celle de son utilité²⁸.

S'il apparaît donc que toute étude du risque social ayant une ambition descriptive doit rendre compte de ces attitudes, l'argument est moins convaincant sur le plan normatif où nous nous placerons, d'autant qu'il rencontre alors l'écueil considérable de la définition d'un point de référence collectif optimal. Nous laisserons donc là ces considérations pour nous en tenir à une approche contrefactuelle du risque où l'on compare un ensemble de possibles. Parmi ceux-ci, une proportion plus ou moins grande peut apparaître comme une perte en fonction du point de vue adopté, sur

27. Pour une revue, voir Camerer (2000)

28. Philipson et Sun (2008) analysent la procédure d'autorisation des médicaments par la Food and Drug Administration, et estiment que celle-ci privilégie largement la sûreté aux dépens de la disponibilité rapide des médicaments. Cette forme d'aversion pour la perte peut certes résulter de facteurs proprement institutionnels, comme l'asymétrie du coût des deux formes d'erreur en termes de réputation (Carpenter, 2004), mais ces facteurs traduisent alors, à leur tour, une aversion collective pour la perte.

lequel nous n'aurons rien à dire.

1.4 Conclusion

Le survol d'un ensemble de domaines répondant à notre définition provisoire nous a permis d'explorer diverses terminologies, démarches et représentations déployées pour comprendre le risque social, puis en rendre compte et éventuellement le mesurer. Il est apparu que les différences terminologiques ne se résument pas à des querelles de mots, ni à une volonté d'instrumentalisation de la part d'acteurs sociaux. Ils correspondent aux contraintes de chaque situation et aux méthodologies qui ont été développées pour en donner une explication en termes de causes et d'effets.

L'idée de représentation, ainsi que la notion économique d'externalités qui nous a permis, dans une mesure relative, de distinguer les risques sociaux des risques privés, ont enrichi notre approche initiale.

Avant de nous armer de ce concept pour le mettre à l'épreuve des problèmes réels de décision en situation de risque social, il convient de dire ce que nous n'en ferons pas.

Le point de vue que nous adopterons dans toute la suite sera celui d'un décideur individuel endossant des responsabilités collectives de prise en charge du risque. Nous critiquerons et tenterons de justifier ce point de vue au chapitre suivant.

Une première conséquence de ce parti-pris individuel sera l'absence de la notion d'externalités dans la suite de nos réflexions, même si elle constituera un élément essentiel de leur toile de fond.

Une seconde conséquence sera l'exclusion de toutes les situations relevant de la théorie des jeux, tels que les actes de malveillance ou la guerre. En nous limitant strictement au domaine de la théorie de la décision, nous ne considérerons donc qu'une partie des risques que nous avons ici caractérisés comme sociaux.

Chapitre 2

L'action publique face au risque social

*Starting at the right,
proceed to the left
passing expected utilities to chance nodes ;
passing maximums to decision nodes ;
until the root is reached.*

Neapolitan (2004)

2.1 Introduction

Après avoir précisé notre approche du risque social, nous cherchons dans ce chapitre à dégager une problématique de l'action publique dans ce domaine.

Le principe fondamental de l'action publique relative aux risques sociaux est d'offrir une protection minimale aux membres de la société à la fois *a priori*, par la maîtrise du risque, et *a posteriori*, par la compensation des dommages subis. Plus précisément, quatre objectifs principaux peuvent être distingués, correspondant à autant de champs de l'intervention publique¹ :

1. la connaissance du risque destinée à orienter les interventions dans l'ensemble des quatre champs (*évaluation du risque*) ;
2. l'action avant l'occurrence de l'aléa, et alors que rien ne l'annonce, afin d'abaisser sa probabilité ou de réduire l'intensité de l'aléa, l'exposition ou la vulnérabilité (*prévention du risque*) ;

1. Les deuxième et troisième terme sont parfois regroupés au sein de la *gestion du risque*. Outre à ces quatre champs, on évoque fréquemment celui de communication du risque qui, pris au sens le plus large, est équivalent au processus de construction d'une décision sociale évoqué dans la dernière partie de ce chapitre.

3. l'action contingente à l'occurrence de l'aléa afin de l'anticiper, lorsque cela est possible, pour en limiter les conséquences, puis de secourir les personnes affectées (*alerte, mitigation des dommages et gestion de crise*) ;
4. enfin l'organisation, après que les dommages sont réalisés, de l'indemnisation des victimes et la réhabilitation des zones ou des biens affectés (*indemnisation*).

Chacun de ces domaines comporte des logiques et des instruments d'intervention très différents selon la nature des risques. L'évaluation d'un risque local, d'une grande stabilité en termes de fréquence, comme certaines inondations, et celle de risques émergents, présentant de multiples facettes et posant des défis scientifiques importants, comme ceux liés aux nanotechnologies, n'ont guère que le nom en commun. La gestion de crise d'un risque à cinétique lente, tel que la sécheresse, a peu à voir avec celle d'une pandémie grippale. Le problème de l'indemnisation se pose en des termes très différents selon que la responsabilité des dommages est attribuable, comme dans le cas d'un accident industriel, ou non, ce qui est presque toujours le cas pour les catastrophes naturelles.

De multiples guides méthodologiques à usage des décideurs publics ont en outre été proposés dans les différents domaines couverts au chapitre précédent depuis trois décennies, dont des exemples classiques sont la formalisation du processus d'évaluation des risques en santé-environnement (National Research Council, 1983) et de la liaison entre évaluation du risque et décision publique (National Research Council, 1996) par l'Académie des sciences aux États-Unis.

Il serait donc vain de prétendre répertorier ici les modalités de l'action publique relative à l'ensemble des risques sociaux.

Pour différents qu'ils soient, cependant, ces domaines de l'action publique gagnent à faire l'objet d'une analyse commune. En premier lieu, des tendances communes ont été à l'œuvre dans l'ensemble de ces domaines, qui nous semblent constituer la toile de fond de la décision publique en situation de risque social. Il s'agit, en premier lieu, du développement spectaculaire de la législation sur le risque, qui semble marquer, comme dans d'autres domaines, le passage de formes d'intervention publique directe vers l'établissement de « règles du jeu social ». Nous y consacrerons la première partie de ce chapitre. Nous nous intéressons ensuite aux conséquences de la décentralisation des décisions à partir du cas des risques d'inondation, qui est illustratif de l'autonomie des forces sociales et des difficultés que peut avoir le décideur public à les diriger. Le second avantage de porter un regard transverse sur les risques sociaux est qu'il s'agit de cas de décision publique en situation d'incertitude ; une réflexion normative sur les modalités de cette décision peut donc être pertinente pour tous, de même qu'elle peut avoir avantage à considérer la diversité des situations. Nous en discuterons dans la dernière partie de ce chapitre. Les considérations précédentes nous amèneront à porter un regard circonspect sur la figure du décideur social bienveillant chère à la théorie économique normative, d'autant qu'elle se heurte également à d'importantes objections théoriques dans le cas du risque

social. Nous plaiderons néanmoins pour le maintien et le renouvellement des modèles normatifs, non comme instruments de décision centralisée mais plutôt comme outils d'analyse, de débat et de construction de décisions démocratiques.

2.2 La société de la réglementation du risque ?

La notion de risque a été au centre d'une activité législative et réglementaire considérable en France au cours des quinze dernières années. Cette évolution, suscitée pour partie par des initiatives de l'Union Européenne et par la jurisprudence des tribunaux européens, est fréquemment justifiée par des motifs de transparence et d'efficacité des politiques publiques. Des juristes tels qu'Alemanno (2008) considèrent ainsi que la réglementation européenne sur le risque est un outil à la fois contre le protectionnisme et pour assurer la « rationalité scientifique » de la réglementation. L'argument sous-jacent est que les méthodologies d'évaluation et de gestion du risque qui sont ainsi introduites dans le droit rendront les décisions publiques concernées à la fois mieux fondées et mieux proportionnées.

Le développement de la législation sur le risque s'est doublé de mutations institutionnelles importantes. La loi du 1er juillet 1998², complétée par celle du 9 mai 2001³, a profondément réorganisé le système d'évaluation et de surveillance des principaux risques pour la santé humaine, qu'elle a fonctionnellement séparé de la gestion du risque, en plaçant au centre du dispositif d'évaluation cinq nouvelles agences sanitaires : l'Institut national de veille sanitaire (INVS), l'Agence française de sécurité sanitaire des aliments (AFSSA), l'Agence française de sécurité sanitaire des produits de santé (AFSSAPS), l'Agence française de sécurité sanitaire environnementale (AFSSE)⁴ et l'Institut de radioprotection et de sûreté nucléaire (IRSN).

Nous commençons, dans cette partie, par décrire la montée de la législation sur le risque et de revenir à cette occasion sur la question de la terminologie. Nous illustrons ensuite la conversion rapide de l'appareil législatif et réglementaire français aux concepts liés au risque à partir du cas de la législation sur les sites classés pour l'environnement. Nous consacrons la dernière section à une discussion du principe de précaution.

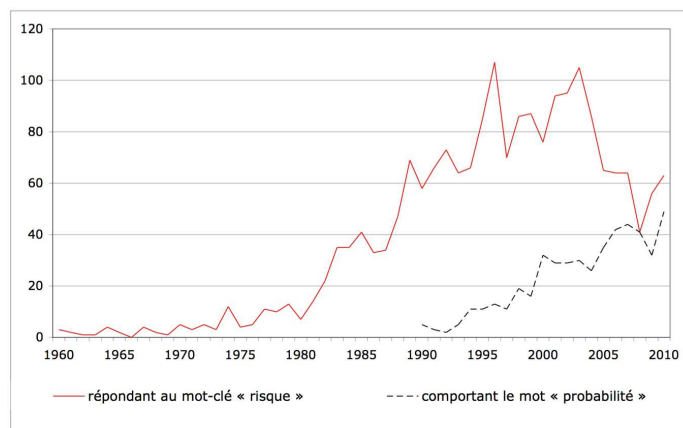
2.2.1 L'émergence de la probabilité dans la législation

La notion de risque fit son apparition dans le droit français en 1681 avec l'Ordonnance sur la Marine de Colbert, dans les termes suivants :

2. Loi du 1er juillet 1998 relative au renforcement de la veille sanitaire et du contrôle de la sécurité sanitaire des produits destinés à l'homme (J.O. du 2 juillet 1998)

3. Loi du 9 mai 2001 (J.O. du 10 mai 2001)

4. L'AFSSE a été par la suite transformée en Agence française de sécurité sanitaire de l'environnement et du travail (AFSSET), puis fusionnée avec l'AFSSA au sein de l'Agence nationale de sécurité sanitaire de l'alimentation, de l'environnement et du travail (ANSES).

FIGURE 2.2.1 – Textes législatifs et réglementaires français comportant les mots *risque* et *probabilité*

Source : base Légifrance (<http://www.legifrance.org>). La recherche par contenus n'est pas disponible pour les textes antérieurs à 1990.

« Sont aux risques des assureurs toutes pertes et dommages qui arriveront sur mer par tempêtes, naufrages, échouements, abordages, changements de route, de voyage ou de vaisseau, jet, feu, prise, pillage, arrêt de prince, déclaration de guerre, représailles, et généralement toutes fortunes de mer. »

Cette énumération d'événements fut reprise dans le Code de commerce de 1807 (Voidey, 2005). Le terme resta par la suite réservé à un usage très ciblé jusqu'à la fin des années 1970. Il a connu depuis un succès fulgurant : entre 1995 et 2005, près de 90 textes législatifs et réglementaires traitant du risque se sont ajoutés au droit français chaque année (voir la figure 2.2.1). Qu'ils traitent de thèmes nouveaux ou qu'ils développent la référence au risque au sein de réglementation existantes, ces textes ont accentué le caractère technique du droit, comme en témoigne également l'usage accru du terme probabilité.

Il en est d'autant plus remarquable que le risque ne fasse l'objet d'aucune définition en droit français. Certes, un ouvrage de référence en matière de vocabulaire juridique propose : « le risque est l'anticipation de la survenue d'un événement défavorable dont la réalisation ou la date de réalisation est incertaine ; se dit aussi bien de l'éventualité d'un tel événement en général, que de l'événement spécifié dont la survenance est envisagée » (Cornu, 2007). Mais une telle acception est trop générale pour être véritablement opérationnelle. Les textes législatifs, et en particulier ceux cités dans le chapitre précédent, laissent pour leur part la notion de risque indéfinie.

En la matière, la loi française procède généralement sur la base de nomenclatures (de substances, d'activités, etc.) et par énumération de sources de danger. À titre d'exemple, le livre V du code de l'environnement, relatif à la prévention des pollutions, des risques et des nuisances, commence par désigner les installations relevant de ses dispositions dans les termes suivants : « Sont soumis

aux dispositions du présent titre les usines, ateliers, dépôts, chantiers et, d'une manière générale, les installations exploitées ou détenues par toute personne physique ou morale, publique ou privée, qui peuvent présenter des dangers ou des inconvénients soit pour la commodité du voisinage, soit pour la santé, la sécurité, la salubrité publiques, soit pour l'agriculture, soit pour la protection de la nature, de l'environnement et des paysages, soit pour l'utilisation rationnelle de l'énergie, soit pour la conservation des sites et des monuments ainsi que des éléments du patrimoine archéologique »⁵.

Nous avons certes plaidé au chapitre précédent que la diversité des situations de risque et les difficultés pratiques de mesure du risque justifiaient une certaine ambivalence dans l'usage des termes. L'argument est connu des juristes, en France (Voidey, 2005) comme au Royaume-Uni (Fisher, 2010).

Le risque, cependant, est qu'en évitant de s'engager dans une clarification des termes autrement que par énumération, la loi se prive des moyens et des opportunités de traiter de questions importantes telles que les formes de prise en compte de l'incertitude ou de pondération des différents enjeux. La loi a tendance à définir les responsabilités face au risque par référence à des notions telles que les études de danger et les évaluations de risque, renvoyant de ce fait à des lexiques et des modes de gestion techniques et à l'arbitrage des tribunaux. Il peut être intéressant, à cet égard, de reconstituer la chronologie de l'apparition du concept de risque dans un cas comme celui de la réglementation sur les installations industrielles dangereuses.

2.2.2 Une réglementation toujours plus technique : le cas des installations classées

La notion d'étude de danger a fait son apparition dans la réglementation sur les installations classées avec le décret du 21 septembre 1977⁶, qui exigeait la réalisation d'une « étude exposant les dangers que peut présenter l'installation en cas d'accident et justifiant les mesures propres à en réduire la probabilité et les effets ». Le décret évitait donc une formulation en termes de risque, et bien qu'il mentionnait les probabilités et conséquences d'accident, ne mettait pas en relation ces termes et n'exigeait pas leur quantification. Il renvoyait par ailleurs à une nomenclature des installations classées établie par le décret du 20 juin 1953, qui reste la référence légale définissant les activités industrielles dangereuses de nos jours. Le décret de 1953 énumère 413 types d'installations industrielles selon leur production, en citant leurs « inconvénients » au nombre desquels l'« altération des eaux », les « émanations nuisibles accidentelles », et les dangers d'incendie ou d'explosion, mais aussi le « danger des mouches ».

Le droit français a par la suite connu une évolution importante lors de la transposition de la

5. Code de l'environnement, article L511-1.

6. Décret n°77-1133 du 21 septembre 1977 pris pour l'application de la loi n° 76-663 du 19 juillet 1976 relative aux installations classées pour la protection de l'environnement.

Directive européenne Seveso II. La Directive a réformé le cadre juridique européen de prévention et de gestion des accidents industriels, notamment en la fondant sur la notion de risque alors que, de façon remarquable, la première Directive Seveso⁷ ne comportait qu'une seule mention du terme risque et aucune du terme probabilité. Elle requiert notamment que les mesures de prévention et de gestion des accidents majeurs soient proportionnées aux risques d'accident propres à chaque établissement mais, comme nous l'avons vu au chapitre précédent, sans référence aux conséquences possibles d'un tel accident.

L'un des principaux textes de transposition de la directive dans le droit français est le décret du 20 mars 2000, qui exige pour les installations industrielles classées la réalisation d'« une étude de dangers qui, d'une part, expose les dangers que peut présenter l'installation en cas d'accident, en présentant une description des accidents susceptibles d'intervenir, que leur cause soit d'origine interne ou externe, et en décrivant la nature et l'extension des conséquences que peut avoir un accident éventuel, d'autre part, justifie les mesures propres à réduire la probabilité et les effets d'un accident, déterminées sous la responsabilité du demandeur ». Le décret précise que « le contenu de l'étude de dangers doit être en relation avec l'importance des dangers de l'installation et de leurs conséquences prévisibles [...] ».

L'arrêté du 10 mai 2000 définit l'accident majeur dans des termes identiques à ceux de la Directive, et la politique de prévention des accidents majeurs comme la mise en œuvre de mesures en relation avec l'étude de dangers. Il mentionne par ailleurs l'identification et l'évaluation des risques d'accidents majeurs en ces termes : « Des procédures sont mises en œuvre pour permettre une identification systématique des risques d'accident majeur susceptibles de se produire en toute configuration d'exploitation des installations. Ces procédures doivent permettre d'apprécier les possibilités d'occurrence et d'évaluer la gravité des risques d'accidents identifiés ». L'arrêté prévoit que cette information soit utilisée pour orienter la gestion des procédures d'urgence, sans la lier à l'étude des dangers.

Le même texte décrit dans les termes suivants l'analyse des risques : « au sens de l'article L. 512-1 du code de l'environnement, [elle] constitue une démarche d'identification et de réduction des risques réalisée sous la responsabilité de l'exploitant. Elle décrit les scénarios qui conduisent aux phénomènes dangereux et accidents potentiels. Aucun scénario ne doit être ignoré ou exclu sans justification préalable explicite. Cette démarche d'analyse de risques vise principalement à qualifier ou à quantifier le niveau de maîtrise des risques, en évaluant les mesures de sécurité mises en place par l'exploitant, ainsi que l'importance des dispositifs et dispositions d'exploitation, techniques, humains ou organisationnels, qui concourent à cette maîtrise ».

Cependant, la circulaire du 10 mai 2000 va beaucoup plus loin dans la spécification technique

7. Directive 82/501/CEE du Conseil du 24 juin 1982 concernant les risques d'accidents majeurs de certaines activités industrielles (Journal officiel n° L 230 du 05/08/1982)

de l'analyse des risques : « L'analyse des risques des installations porte sur toutes leurs conditions d'exploitation (phases transitoires et d'arrêt incluses). Elle nécessite l'utilisation de méthodes systématiques (HAZOP, AMDEC, *what-if*, arbres de défaillances, par exemple). Ces méthodes peuvent également faciliter l'étude des scénarios d'accident, faire apparaître l'importance pour la sécurité du respect de certaines conditions prises comme hypothèses (délai de réaction des opérateurs, par exemple), envisager les défaillances de mode commun (situations où un événement tel qu'un séisme, un incendie, une erreur de maintenance, etc, est susceptible d'affecter simultanément plusieurs systèmes nécessaires à la sûreté de l'installation) et permettre une évaluation correcte des conséquences ».

Le décret du 13 septembre 2005 apporte un élément important d'appréciation des objectifs de l'étude de danger : celle-ci « justifie que le projet permet d'atteindre, dans des conditions économiquement acceptables, un niveau de risque aussi bas que possible, compte tenu de l'état des connaissances et des pratiques et de la vulnérabilité de l'environnement de l'installation. »

Depuis l'arrêté ministériel du 29 septembre 2005, « l'étude de dangers doit contenir, dans un paragraphe spécifique, le positionnement des accidents potentiels susceptibles d'affecter les personnes à l'extérieur de l'établissement selon la grille de l'annexe V du présent arrêté. » Cette grille croise cinq niveaux de probabilité, notés de À à E, et cinq niveaux de gravité, modéré, sérieux, important, catastrophique et désastreux, sans préciser les limites absolues ou relatives de ces niveaux.

En outre, l'arrêté modifie la définition d'un accident majeur : « un événement tel qu'une émission, un incendie ou une explosion d'importance majeure résultant de développements incontrôlés survenus au cours de l'exploitation, entraînant, pour les intérêts visés au L. 511-1 du code de l'environnement, des conséquences graves, immédiates ou différées et faisant intervenir une ou plusieurs substances ou des préparations dangereuses. » L'impact sur la santé humaine et/ou l'environnement a ici été remplacé par l'ensemble plus large d'intérêts établi dans l'article 511-1 du code de l'environnement.

Au total, en cinq années, de 2000 à 2005, le droit français a absorbé un ensemble de dispositions techniques qui conditionne inévitablement la façon dont les dangers relatifs aux installations classées sont pris en compte. Cette évolution peut être porteuse des progrès annoncés par Alemanno (2008) à trois conditions : que les exploitants d'installation dangereuses aient la capacité de réaliser des études de danger de bonne qualité ; que les autorités publiques aient la capacité effective de les contrôler ; et que ces autorités disposent des instruments nécessaires à la mise en place d'une politique d'ensemble en matière de risques industriels. Nous ne traiterons pas ici de la première condition. Au sujet des deux autres, il est à noter que la loi du 30 juillet 2003, cherchant à promouvoir une démarche commune à différents secteurs de risque, a créé des plans de prévention des risques technologiques (PPRT) sur le modèle des plans de prévention des risques naturels (PPRN). Plus

bas, nous analysons quelques problèmes de mise en œuvre des PPRN dans le cas particulier des risques d'inondation.

2.2.3 Le principe de précaution

La Charte de l'environnement⁸, adoptée par le Parlement réuni en congrès le 28 février 2005, a inscrit dans la Constitution française le droit de chaque personne de vivre dans un environnement équilibré, ainsi que le devoir des autorités publiques d'engager des mesures conservatoires en application du principe de précaution lorsque « la réalisation d'un dommage, bien qu'incertaine en l'état des connaissances scientifiques, pourrait affecter de manière grave et irréversible l'environnement ».

Le principe de précaution a été formulé la première fois en Allemagne à la fin des années 1970, a fait son entrée sur la scène internationale en tant qu'« approche de précaution » dans la deuxième conférence internationale sur la protection de la mer du Nord en 1987, puis en tant que principe dans la conférence suivante en 1990 (Ewald, 2002). Au cours des années 1990 et du 2000, il a fait l'objet de débats intenses et de prises de position tranchées. En tant que principe de droit, il semblait concerner des situations très spécifiques, dans lesquelles la science établissait l'existence possible d'un risque sans pouvoir en mesurer l'ampleur, et ne pouvait exclure des conséquences graves et irréversibles (Kourilsky et Viney, 2000). Du point de vue de l'économie, il marquait la reconnaissance d'une différence entre des situations où l'incertitude était probabilisable et d'autres où il ne l'était pas (Godard et al., 2002). La Communication de la Commission européenne sur le principe de précaution introduisait en outre la notion importante de proportionnalité, qui devait permettre d'encadrer les mesures se réclamant de son application (CCE, 2000).

Plusieurs années après son entrée dans la Constitution et alors qu'il a fait l'objet de plusieurs arbitrages dans des tribunaux internationaux tels que l'Organisation Mondiale du Commerce (OMC) ou la Cour de Justice des Communautés Européennes (CJCE), l'apport du principe a pu sembler plus incertain. Un grand nombre de décisions publiques s'y référant ont été contestées devant les tribunaux, généralement au motif que leurs fondements scientifiques étaient mal assurés. Ces tribunaux ont pu eux-mêmes avoir des lectures contradictoires des exigences et des cas d'application du principe⁹. Dans le débat public, le principe de précaution est resté une notion mal définie, fréquemment utilisée pour désigner des situations d'incertitude classique.

Les décideurs publics et, dans une moindre mesure, les tribunaux, semblent hésiter devant la signification à accorder à la forme d'incertitude visée par le principe. Une simple spéculation peut-elle en relever ? Ou, à l'extrême opposé, une situation où le risque a été entièrement évalué mais où l'un des paramètres de cette évaluation paraît incertain ?

Une jurisprudence internationale semble cependant se dégager, précisant graduellement le cadre

8. Charte de l'environnement, Journal officiel de la République française, 2 mars 2005.

9. Des exemples sont discutés dans CPP (2010).

d'application du principe de précaution. Elle estime que le principe ne permet pas de contourner l'obligation d'évaluation, et qu'un risque peut toujours être évalué au moins en partie. Les nombreuses décisions publiques cassées par les tribunaux se caractérisaient ainsi par l'effort insuffisant consenti pour apprécier ce qui pouvait l'être dans une situation incertaine : « le risque doit être "suffisamment documenté" par des indications scientifiques solides et précises qui, sans lever totalement les incertitudes, permettent néanmoins d'établir que son existence n'est pas improbable ; la mesure doit être "suffisamment étayée" par les données scientifiques ; des "indices sérieux et concluants" sont exigés, qui "permettent de conclure raisonnablement que la mise en œuvre du principe de précaution est nécessaire" » (Noiville, 2009).

Il apparaît ainsi que l'incertitude visée par le principe n'est ni un doute dénué de tout élément d'évaluation, ni un risque dont l'évaluation ferait l'objet d'incertitudes quantifiables, par exemple sous forme d'intervalles de confiance¹⁰, mais d'un risque dont l'évaluation est partielle en raison de l'incapacité à établir (ou à rejeter) une ou quelques étapes du raisonnement menant d'une source initiale à des dommages finaux (Alemanno, 2007). Le principe, tel que rendu progressivement opérationnel par la jurisprudence, permet de fonder ces étapes sur des éléments de doute sérieux, avec l'obligation de mesurer le risque supposé et de montrer que les mesures envisagées lui sont proportionnelles.

Un instrument important et nouveau dans la prise en charge des risques sociaux a donc commencé à être forgé, et devrait continuer à l'être dans les prochaines années, sans que son devenir ait véritablement semblé dans les mains du décideur public.

2.3 Décisions publiques ou collectives ? L'exemple des inondations

Nous analysons dans cette partie le cadre légal et institutionnel relatif aux risques de catastrophes naturelles, et en particulier aux inondations, pour montrer que l'action publique en situation de risque social ne le cède en rien, en termes de complexité, au risque social lui-même, et que le décideur public ne semble pas toujours en mesure de mettre en place un choix au nom de la société.

2.3.1 Le système français de gestion des risques liés aux inondations

La législation relative aux risques de catastrophe naturelle est la première, en France, à avoir été explicitement construite autour de notions de prévention et de partage d'un risque social.

Le Préambule de la Constitution de 1946 proclamait le principe de « solidarité de tous les

10. Il convient de noter ici que l'essentiel des risques sociaux décrits dans le chapitre précédent sont soumis à cette dernière forme d'incertitude.

Français devant les charges qui résultent des calamités nationales ». En expression de cette solidarité, la loi du 13 juillet 1982 relative à l'indemnisation des victimes de catastrophes naturelles instaura la prise en charge par la collectivité des « dommages matériels directs ayant eu pour cause déterminante l'intensité anormale d'un agent naturel » (article 1) ¹¹. Elle introduisit par ailleurs la notion de risque à travers les plans d'exposition aux risques : « L'État élabore et met en application des plans d'exposition aux risques naturels prévisibles, qui déterminent notamment les zones exposées et les techniques de prévention à y mettre en œuvre tant par les propriétaires que par les collectivités ou les établissements publics » (article 5).

Par la suite, le décret du 3 mai 1984 visa à renforcer la prévention en établissant des modalités précises de prise en compte des risques naturels ¹². En demandant qu'une zone rouge soit déterminée partout où le risque était « particulièrement redoutable », elle chercha à obtenir que toutes les zones présentant des dangers directs pour la vie humaine soient systématiquement localisées. Le décret y imposa des mesures conservatoires en dehors de toute considération de coût et d'avantage. Dans la zone bleue, où le risque devait être significatif sans porter sur la vie humaine, les mesures de prévention devaient être économiquement proportionnées. Le décret fut par la suite remplacé par celui du 15 mars 1993 qui en reprit les grands principes ¹³.

La loi du 2 février 1995 tenta de poser les bases d'une politique plus ambitieuse de prévention des risques, comme contrepartie du mécanisme public d'indemnisation des dommages, en créant les plans de prévention des risques (PPR) et le fonds pour la prévention des risques naturels majeurs ¹⁴. Elle fut prolongée et approfondie par la loi du 30 juillet 2003, qui créa de nouvelles obligations d'information sur le risque et réforma les PPR ¹⁵. La logique des lois de 1995 et de 2003 était de faire des PPR le fruit d'une élaboration commune entre l'État et les collectivités territoriales, sur la base d'une caractérisation de l'aléa local par les services du premier. Au travers de ces Plans, l'État espérait avoir un meilleur contrôle des efforts de prévention consentis par les communes et les particuliers. Lors de leur élaboration, son représentant définit les contraintes portant sur toute construction nouvelle, demande les aménagements nécessaires sur le bâti existant et impose des mesures de protection ou de sauvegarde. Les dispositions du Plan sont annexées au Plan d'occupation des sols de la commune, peuvent être imposées à toute personne publique ou privée au terme d'un délai de cinq ans.

Enfin, l'Union européenne adopta en 2007 la directive 2007/60/CE, qui concernait les inonda-

11. Loi du 13 juillet 1982 relative à l'indemnisation des victimes de catastrophes naturelles, Journal officiel du 14 juillet 1982, pp.2242-3

12. Décret n° 84-328 du 3 mai 1984 relatif à l'élaboration des plans d'exposition aux risques naturels prévisibles, Journal officiel du 6 mai 1984, p. 1338

13. Décret n° 93-351 du 15 mars 1993 relatif aux plans d'exposition aux risques naturels prévisibles, Journal officiel du 17 mars 1993, p. 4145

14. Loi du 2 février 1995 relative au renforcement de la protection de l'environnement, Journal officiel du 3 février 1995

15. Loi du 30 juillet 2003 relative à la prévention des risques technologiques et naturels et à la réparation des dommages, Journal officiel du 31 juillet 2003

tions de tous types pouvant survenir dans les pays membres¹⁶. La directive instaura l'obligation pour les États-membres de procéder à une évaluation préliminaire des risques d'inondation, d'en établir une cartographie, et d'engager des mesures corrélatives de gestion du risque. L'évaluation du risque devait être fondée sur des relevés historiques et sur l'étude des évolutions à long terme, « en particulier l'incidence des changements climatiques sur la survenance des inondations », et comprendre au moins des cartes des districts hydrographiques, une description des inondations passées ayant eu un impact significatif, et éventuellement une évaluation des impacts potentiels d'inondations futures permettant de prendre en compte les conditions dans lesquelles celles-ci surviendraient (occupations des sols, mesures de protection existantes, etc.). La gestion des risques devait mettre en évidence des objectifs de réduction de la probabilité de survenue d'une inondation et/ou de ses conséquences potentielles, ainsi qu'un ensemble de mesures de prévention, de protection ou de préparation permettant d'atteindre ces objectifs. La directive fut transposée dans le droit français par l'intermédiaire de la loi du 12 juillet 2010¹⁷ et du décret du 2 mars 2011¹⁸, qui renforcèrent l'évolution vers une gestion des risques d'inondation par bassin.

Ces grandes étapes législatives furent accompagnées par la réalisation de dizaines de rapports d'expertise, d'enquête et de retour d'expérience et de l'élaboration de milliers de PPR dont la grande majorité était relative aux risques d'inondation. Il faudrait certes également évaluer les ressources engagées par l'État dans ce domaine mais, à première vue, il ne semble pas déraisonnable d'estimer qu'un effort important a été consenti afin de mieux prévenir les risques d'inondation.

2.3.2 L'effort global de prévention est-il suffisant ?

À en juger par les dommages causés et les évaluations menées par les services de l'État eux-mêmes, cet effort n'a pas produit les résultats escomptés. Des inondations catastrophiques sont régulièrement subies, dont les dernières en date sont celles causées en Vendée et en Charente-Maritime par la tempête Xynthia le 28 février 2010 (53 décès) et celles survenues dans le Var les 15 et 16 juin 2010 (23 décès). Selon la mission d'inspection tirant les leçons de la tempête Xynthia :

« la politique actuelle en la matière n'[a] obtenu, dans les départements visités, que des résultats peu satisfaisants, moins d'ailleurs au niveau des concepts et principes développés dans la législation et la réglementation nationales qu'à celui de l'application qui en est faite sur le terrain. » (Mission d'inspection sur la tempête Xynthia, 2010, p.4)

16. Directive 2007/60/CE du Parlement européen et du Conseil du 23 octobre 2007 relative à l'évaluation et à la gestion des risques d'inondation, Journal officiel de l'Union européenne n°288, 6 novembre 2007, p.27

17. Loi n° 2010-788 du 12 juillet 2010 portant engagement national pour l'environnement, Journal officiel du 13 juillet 2010, p.12905

18. Décret n° 2011-227 du 2 mars 2011 relatif à l'évaluation et à la gestion des risques d'inondation, Journal officiel du 3 mars 2011, p.4130

Les réformes engagées depuis 1995 tendent à faire des Plans de prévention des risques la clé de voûte de la politique de prévention. Or les PPR ne semblent pas être, dans leur modalités d'élaboration et d'application actuelles, à même d'assurer une prévention suffisante des risques liés aux inondations pour trois raisons principales.

En premier lieu, les PPR semblent avoir été conçus afin de réduire autant que possible les risques pour les vies humaines. De ce fait, ils relèvent d'une démarche de prévention nécessaire mais non suffisante. Ils ont par exemple un rôle limité en matière de réduction de la vulnérabilité dans le bâti existant, à moins que des risques immédiats pour la vie humaine soient avérés.

Comme il a été observé lors des événements de septembre 2002 en Languedoc-Roussillon, les PPR n'interdisent pas la construction dans des zones inondables « bleues », c'est-à-dire ne présentant pas de risque direct pour les vies humaines (bien que des mesures de précaution soient généralement imposées)¹⁹. De ce fait, ils autorisent que les coûts prévisibles des dommages futurs liés aux inondations soient transférés des bénéficiaires de ces constructions au régime catastrophes naturelles, et donc à la communauté nationale. En résumé, à travers les PPR, la politique de prévention s'intéresse essentiellement aux risques pour les vies humaines, alors que le régime d'indemnisation couvre exclusivement les pertes matérielles.

De ce fait, les PPR ne constituent pas une garantie systématique de contrôle de l'urbanisation en zone inondable. Le problème semble particulièrement aigu dans les régions où la pression démographique ou économique impose une expansion continue des zones d'activité ou d'habitation. Dans les quatre départements de l'Aude, du Gard, du Hérault et des Pyrénées Orientales, 384 100 (17 % de la population) personnes vivaient en 2002 en zone inondable, dont 321 000 (14 % de la population) dans des zones où le risque inondation est qualifié de « fort à très fort »²⁰. Selon le rapport de retour d'expérience des crues de septembre 2002, l'urbanisation des zones inondables s'était poursuivie dans cette région (bien qu'à un rythme plus lent que précédemment), et à en juger par les projets existants, « quelques centaines d'habitations supplémentaires pourraient s'installer dans la région Languedoc-Roussillon, en zone inondable (...) ».

Deuxièmement, la protection des vies humaines elle-même reste dans certains cas insuffisante. Les expériences de crues ont montré de façon régulière que le niveau d'aléa retenu et la délimitation des zones bleues et rouges dans les PPR peuvent être exagérément optimistes²¹. L'imprécision dans le tracé des zones de moindre risque est due parfois à l'incertitude scientifique inhérente à ce type d'exercice, mais le plus souvent à la pression d'élus locaux et de riverains, qui dans certains cas parviennent à obtenir une modification des tracés d'origine. Un rapport des instances en charge du

19. Mission d'expertise sur les crues de septembre 2002 (2003, p.78)

20. Source : Direction régionale de l'équipement de Languedoc-Roussillon, citée par la Mission d'expertise sur les crues de septembre 2002 (2003, p.76).

21. Mission d'inspection sur la tempête Xynthia (2010) ; Mission d'expertise sur les crues de septembre 2002 (2003, p.86).

retour d'expérience rappelait en 2002 que « là où des vies humaines seraient menacées, (...) la non constructibilité doit être la règle dans les PPR », ce qui ne semblait donc pas être toujours le cas²². Il préconisait que les services de l'État rappellent « qu'à l'occasion des l'élaboration des PPR, l'aléa n'est pas négociable ». Huit ans plus tard, une autre mission d'inspection relevait « une négociation inopportune sur l'aléa de référence et la prise en compte de la vulnérabilité et des enjeux dans les "zones de danger" »²³.

Enfin, le respect des PPR est handicapé par l'absence de mécanismes incitatifs suffisants. La mise en œuvre effective des mesures prévues repose sur une capacité de contrôle de l'État qui peut s'avérer extrêmement coûteuse. En octobre 2001, le rapport de la Commission d'enquête du Sénat sur les inondations de la Somme émettait des doutes sur la capacité matérielle des pouvoirs publics à assurer la « mission de police administrative » concernant la mise en œuvre par les individus et les collectivités des mesures préconisées par les PPR. En 2010, la mission d'inspection sur la tempête Xynthia préconisait pourtant de « fixer par voie réglementaire : la définition des aléas de référence inondation et submersion marine ; la caractérisation de l'intensité de l'aléa : fort, moyen, faible ; la délimitation des zones en différenciant clairement les zones de danger et les zones de précaution ; les mesures générales de prévention, de protection et de sauvegarde applicables à ces zones, y compris en matière de constructibilité »²⁴.

2.3.3 L'importance du processus de décision

Une partie au moins de la difficulté semble donc provenir de ce que l'État persiste à se donner un rôle qu'il ne parvient pas à tenir. Cette évolution semble confirmer les craintes exprimées il y a quinze ans dans un rapport officiel :

« Les doutes viennent du fait que l'impulsion, venue du sommet, est restée essentiellement une affaire de l'État et n'est pas partagée par la société. [...] Le renforcement de l'action nationale de prévention, qui pourrait apparaître *a priori* comme un indicateur positif, s'identifie ainsi à un durcissement des gestes d'autorité de l'État ; il n'a pas été conçu pour être partagé ; il peut être le déclencheur d'affrontements plus que d'actions collectives, entraîner une grande confusion dans les responsabilités, et un défaut de lisibilité des objectifs et des moyens : l'État, en élargissant ses propres compétences n'a-t-il pas atténué la responsabilité des collectivités et des particuliers ? » (Bourrelier, 1997).

Conférant d'un côté à l'État des responsabilités trop extensives, le système en place multiplie par ailleurs les incitations négatives.

22. Inspection générale de l'administration et al. (2002)

23. Mission d'inspection sur les inondations du Var (2010).

24. Mission d'inspection sur la tempête Xynthia (2010, p.10).

En rendant les conditions d'assurance (prime, franchise, clauses d'exclusion) indépendantes du degré d'exposition et de vulnérabilité de l'assuré, le mécanisme d'indemnisation des victimes entraînait, selon les termes du rapport de la Cour des Comptes en 1999 « une déresponsabilisation des assurés et une indifférence des assureurs », ces derniers estimant, « non sans fondement, qu'une telle tâche [de contrôle] relève de la mission de police administrative de l'État [...] : elles n'interviennent en fait que comme de simples intermédiaires financiers » (Cour des comptes, 1999). Malgré une réforme du mécanisme en 2005, ce constat reste entièrement valide (Mission d'inspection sur la tempête Xynthia, 2010, p.7). L'aléa moral individuel a son pendant dans les relations entre l'État et les collectivités territoriales, puisque l'existence d'un mécanisme d'assurance national permet à ces dernières d'externaliser en partie les coûts liés à leurs choix risqués.

Enfin, tout en affirmant l'importance d'un processus de construction collective de la décision, les PPR paraissent en retenir une conception très étroite. Alors que les services administratifs multiplient les appels à la « culture du risque » et à la connaissance de l'aléa, ils ne produisent, à titre d'exemple, pas de donnée concernant l'ampleur de l'effort national de prévention des risques d'inondation. Ceci exclut toute évaluation de cet effort, en regard par exemple du coût des désastres pour la collectivité.

Le processus de concertation et de négociation, qui se donne pour objectif de déterminer le risque acceptable localement dans les limites fixées par l'État, est focalisé sur l'aléa et sur la mise en place de restrictions déterminées de façon centralisée. La prise en compte des enjeux et de la vulnérabilité est sommaire. Les alternatives possibles aux actions envisagées (en matière de construction, d'usage et/ou de prévention) ne sont pas intégrées. Il n'existe ainsi pas de considération systématique d'une partie des déterminants du risque acceptable. Ne faut-il pas, à titre d'exemple, reconnaître dans l'élaboration même de la politique de prévention que la pression foncière ou économique (activités hôtelières et touristiques, etc.) et le coût d'opportunité de la non constructibilité varient selon les régions ? N'est-il pas également souhaitable d'accompagner la maîtrise de l'urbanisation des zones inondables par une exploration des alternatives existantes dans des zones non exposées au risque (terrains agricoles, zones classées, habitations délaissées, etc.), et une prise en compte des coûts afférents ?

Sur ces points, les Plans de Prévention des Risques Technologiques, homologues des PPR pour les risques industriels introduits par la loi du 31 juillet 2003, diffèrent de ces derniers puisqu'ils se fondent sur une estimation des enjeux, une superposition des aléas et des enjeux, et des investigations complémentaires, notamment en matière de coûts de mise en œuvre des mesures foncières. L'objectif, dont pourrait s'inspirer l'évolution future des PPR, est d'apporter « toutes les informations nécessaires aux différents acteurs concernés [...] afin de choisir les différentes orientations du plan » (Ministère de l'écologie, 2005).

2.4 Le point de vue normatif

Après avoir illustré la complexité de la gestion des risques sociaux, nous adoptons dans cette partie un point de vue normatif. Trois approches sont possibles pour tenter de déterminer ce que la société devrait faire face à une situation de risque social. La première est celle de la théorie du choix social, qui considère les conditions dans lesquelles un décideur social peut représenter les préférences d'un groupe d'individus dans une situation d'incertitude. La seconde est celle de la théorie de la décision classique, appliquée aux problèmes sociaux en supposant simplement que l'on se trouve en présence d'un décideur représentatif de la société. La troisième, enfin, consiste à supposer que certaines grandeurs observables sont la résultante de choix sociaux, et à en dériver des préférences sociales sous-jacentes. Nous passons ici en revue quelques principes et résultats de chaque approche et quelques critiques qui lui ont été adressées, en allant de la plus théorique à la plus pratique.

2.4.1 L'agrégation des préférences individuelles

2.4.1.1 Des résultats d'impossibilité

La théorie du choix social en présence d'incertitude, qui constitue un point de départ naturel d'une réflexion sur les décisions collectives relatives aux risques sociaux, est dominée par la tension entre l'exigence de rationalité collective (telle que définie, par exemple, par les axiomes de Savage) et celle de représentation des préférences individuelles (critères de Pareto).

Mongin (1995) a démontré que les préférences d'un groupe d'individus ayant des goûts et des croyances divers ne pouvaient être agrégées sans contrevenir soit aux critères de rationalité bayésienne, soit à l'une des conditions de représentativité parétienne.

Gilboa, Samet et Schmeidler (2004) ont cependant montré que l'agrégation était possible dès lors que l'on limitait l'application des critères de Pareto aux seuls cas où les individus ont les mêmes croyances (*principle of probabilistic unanimity*). Cette restriction permet en outre d'écarter les cas de fausse unanimité (*spurious unanimity*) où l'apparente concordance des préférences individuelles résulte en fait d'appréciations différentes tant en termes de goûts que de croyances (Mongin, 1997).

Toutefois, dans les cas où l'incertitude ne peut pas être résumée de façon satisfaisante par une distribution de probabilités (ce que l'on conviendra d'appeler une situation d'*ambigüité*), Gajdos, Tallon et Vergnaud (2008) ont établi que l'agrégation était possible si et seulement si la société ne montrait pas d'aversion à l'ambigüité, ce qui semble constituer une exigence forte.

2.4.1.2 La faiblesse normative de l'agrégation des croyances

Il existe des arguments importants contre l'usage des conditions de Pareto, même limité aux cas d'unanimité probabiliste, et plus généralement contre l'agrégation des préférences individuelles en

situation de risque social. Ils sont notamment d'ordre méthodologique, cognitif et éthique.

En premier lieu, la révélation des probabilités subjectives par les choix observés, déjà problématique en situation de risque individuel (Aumann et Savage, 1971), s'avère encore plus hasardeuse dans le cas des risques sociaux. En effet, dans la mesure où l'utilité même du décideur serait modifiée dans certains états qu'il doit prendre en compte (par exemple, une catastrophe nucléaire), ses choix risquent d'intégrer un comportement de couverture (*hedging*) contre ces états, ce qui fausse les résultats de la théorie de la décision.

En deuxième lieu, de nombreuses études empiriques, auxquelles nous ferons mention dans la suite, indiquent que les jugements individuels en matière de risque sont affectés d'un certain nombre de biais, variables selon les individus et selon les risques, qu'il n'est pas nécessairement souhaitable de reproduire dans les décisions publiques.

En troisième lieu et plus fondamentalement, la grande faiblesse de l'agrégation en présence d'incertitude est qu'elle repose sur une extension du « principe de souveraineté du consommateur », selon lequel il ne peut y avoir de hiérarchie des goûts, aux croyances, c'est-à-dire à des évaluations de faits (Mongin et d'Aspremont, 1998). L'application du principe d'anonymat dans une situation de risque social revient à donner le même poids à une opinion bien informée qu'à une autre fondée sur une totale ignorance. Il faut noter que l'un des rôles essentiels des institutions démocratiques est d'éviter une telle situation, notamment en faisant que l'apprentissage collectif privilégie les sources les mieux informées, ou à défaut d'en limiter les conséquences.

2.4.2 Le décideur social représentatif et rationnel

2.4.2.1 Le risque dans les modèles d'utilité espérée

Deux modèles de décision individuelle sont à l'origine de l'essentiel des applications de la théorie de la décision à des problèmes de choix sociaux, et notamment de risques sociaux : le modèle d'utilité espérée de von Neumann et Morgenstern (1944), également appelé d'utilité espérée « objective » (UEO) parce que les probabilités y sont pré-déterminées, et celui de Savage (1954), dit d'utilité espérée subjective (UES) parce que les probabilités y sont dérivées des préférences du décideur.

Le modèle d'UEO est de très loin celui qui l'emporte dans les développements théoriques comme dans les applications pratiques, en raison de sa simplicité d'usage. De nombreuses variantes en ont été dérivées afin d'étudier les comportements statiques et dynamiques face au risque (Gollier, 2001). L'essentiel des analyses coûts-bénéfices et des évaluations plus ciblées telles que les calculs de valeur de la vie reposent sur lui²⁵.

L'un des avantages supposés du modèle UEO dans un cadre social est le résultat établi par Arrow

25. Pour une revue des applications à la réglementation des risques sanitaires et environnementaux, on pourra se reporter à Viscusi (2007).

et Lind (1970), selon lequel un décideur représentatif d'un groupe d'individus suffisamment grand pourrait se montrer neutre face au risque en raison de la faiblesse de la part de risque porté par chaque individu. La grande majorité des applications aux risques sociaux font appel à ce résultat, qui permet d'éviter la tâche délicate de l'estimation d'un degré d'aversion collective pour le risque. Or le résultat d'Arrow et Lind dépend du rapport entre l'importance des pertes encourues et le nombre d'agents et, de ce fait, ne peut être supposé valide pour des risques catastrophiques tels que ceux d'une pandémie ou d'un accident nucléaire.

Une autre application du modèle UEO qui doit être mentionnée ici est le traitement du principe de précaution proposé par Gollier, Jullien et Treich (2000). En caractérisant le principe en termes de dynamique d'acquisition de l'information, ces auteurs parviennent à en donner une interprétation dans le cadre d'UEO. Ils se placent ainsi à contre-courant de la vision la plus courante en économie, selon laquelle le principe s'applique à des cas d'incertitude ne pouvant pas être résumés par une seule distribution de probabilités. Mais leur modèle ne reflète aucune différence entre des situations de prévention et de précaution, et n'a pas donné lieu à des développements ultérieurs.

2.4.2.2 Les critiques du décideur social représentatif et rationnel

Quasiment depuis leur origine, les modèles d'utilité espérée ont fait l'objet de critiques empiriques fortes prenant pour cible leurs représentations de l'incertitude et des comportements face à l'incertitude. Les premières mises en cause du cadre axiomatique de la théorie bayésienne provenaient d'expériences, réelles ou mentales, d'individus placés en situation d'incertitude (Allais, 1953; Ellsberg, 1961; Tversky et Kahneman, 1973, 1974). Ces travaux ont suscité une littérature importante qui a cherché à relâcher certains postulats des modèles d'utilité espérée afin d'obtenir une représentation plus générale. Nous serons amenés à discuter une partie de cette nouvelle théorie de la décision sous un angle théorique dans les deux derniers chapitres de ce travail. En matière d'application à des risques sociaux, cependant, ses apports sont encore limités²⁶.

En s'attaquant au sujet des choix sociaux, cependant, les modèles d'utilité espérée rencontraient d'autres difficultés considérables. Sans prétendre passer en revue la littérature pléthorique qui s'est développée autour de ces sujets, nous proposons d'évoquer ici trois sources de problèmes pour les modèles d'utilité espérée dans un cadre social.

Une première catégorie de critiques s'est portée vers la définition du problème de décision, particulièrement chez von Neumann-Morgenstern où il prend la forme d'un choix entre loteries objectives. Dans le cas de risques sociaux plus encore que celui de risques individuels, une telle simplification n'occulte-t-elle pas des aspects importants du problème ? En analysant des procédures

26. Citons tout de même quelques travaux traitant de décisions publiques fondées sur des probabilités de deuxième ordre (Viscusi, 1998), l'attitude individuelle face à un risque social ambigu (Riddell et Shaw, 2006), le consentement à payer (Viscusi, Magat et Hubert, 1991) ou la valeur d'une vie statistique (Treich, 2010) en présence d'ambiguïté.

d'évaluation des risques (particulièrement en santé-environnement), un courant de littérature a montré qu'il était impossible de parvenir à élaborer un protocole objectif, débarrassé de toute appréciation de valeur, et donc ne faisant pas *déjà* appel aux préférences lors de l'estimation des probabilités associées aux loteries²⁷. Il fallait donc ouvrir la question de l'évaluation du risque au choix social. Ces critiques rejoignent certaines de nos observations du chapitre précédent, et nous semblent particulièrement intéressantes car elles n'affectent pas la seule capacité descriptive de la théorie bayésienne, mais aussi sa portée normative.

Une deuxième catégorie de critiques s'est concentrée sur un autre aspect de la mesure du risque : sa dimension sociale. Puisque les coûts sociaux d'un événement dommageable dépendent des comportements, une mesure complète du risque devrait intégrer une représentation des interactions sociales pertinentes. À titre d'exemple, en raison de la sensibilité des croyances probabilistes aux effets de « cadrage » (Tversky et Kahneman, 1981), des institutions telles que les médias, les autorités publiques, les entreprises et différents groupes sociaux ont la capacité d'amplifier ou au contraire d'atténuer le risque social (Pidgeon, Kasperson et Slovic, 2003). Or la notion de rationalité véhiculée par les axiomes de von Neumann et Morgenstern ou les postulats de Savage ne laisse aucune place à de tels aspects du contexte social.

Une troisième catégorie, enfin, était relative aux approximations de la fonction d'utilité. Alors que la question de la mesure du risque était déjà passablement complexe, celle de sa valorisation par la société ouvrait des sujets de controverse inépuisables. Des études dites psychométriques décrivaient, à titre d'exemple, des appréciations individuelles du risque où entraient en compte que le caractère volontaire ou subi de la situation, ou encore l'angoisse liée à l'hypothèse d'un accident (Slovic, 2000). Les débats au sujet de la valorisation de la vie ont permis d'affiner les méthodes de l'analyse coûts-bénéfices, l'approche fondée sur le capital humain étant peu à peu délaissée au profit de mesures fondées sur le consentement à payer ou à recevoir²⁸. De même, la contestation de l'usage d'une unité de mesure (généralement monétaire) unique a conduit au développement de l'analyse multi-critères (Keeney et Raiffa, 1976).

2.4.3 Les préférences sociales révélées

2.4.3.1 Risque accepté et risque acceptable

Deux travaux publiés à la fin des années 1960, qui simplifiaient à l'extrême l'approche de la théorie de la décision, ont exercé une influence considérable sur les décisions publiques en situation de risque social, notamment dans le domaine des risques industriels et technologiques.

En 1969, Starr publia une étude dans laquelle il faisait figurer sur un diagramme (bénéfices,

27. Voir par exemple Shrader-Frechette (1991, ch. 4).

28. Pour une discussion synthétique et assez complète des méthodes de l'analyse coûts-bénéfices, on pourra se reporter à Adler et Posner (1999).

risques) une vaste gamme de situations de risque, allant des catastrophes naturelles à la guerre du Vietnam, en passant par le ski ou les déplacements en voiture (Starr, 1969). Les risques étaient mesurés par une probabilité de décès par heure d'exposition, et les bénéfices par une estimation de surplus économique. La position des points délimitait une plage, bordée en bas par une courbe d'aspect logarithmique regroupant les situations subies et en haut par une deuxième courbe reliant les situations choisies. Ces courbes, selon Starr, reflétaient les préférences de la société dans ses choix de situations à risque, notamment de technologies risquées, à la manière de la méthode des préférences révélées en économie. Elles permettaient, disait-il, de déterminer si les risques posés par une technologie étaient acceptables au regard des bénéfices qui en étaient attendus.

Starr élaborait ainsi une méthode de dérivation d'une notion proposée deux années plus tôt par Farmer dans le domaine du risque nucléaire à partir d'un raisonnement encore plus simple. Farmer avait proposé de définir le risque acceptable par le biais d'une droite tracée de façon relativement arbitraire dans le plan (probabilité, conséquences) (Farmer, 1967). À sa suite, l'acceptabilité du risque, généralement traduite par un seuil de probabilité annuelle (pour un accident d'un certain type ou un décès causé), est devenue une notion centrale pour la gestion des risques sociaux²⁹.

En 1992, un rapport de l'autorité en charge de la santé publique et de la sûreté au Royaume-Uni, le *Health and Safety Executive*, connu également un certain retentissement. Le rapport proposait de parler de tolérabilité dans tous les cas où les risques induits par une activité semblent justifiés au regard des avantages qu'elle procure, sans pour autant être acceptables, et de gérer les risques tolérables en conformité avec le principe *ALARA*³⁰ (HSE, 1992). À travers la notion de « raisonnablement réalisable », ce principe a l'avantage d'ouvrir le processus de décision à l'ensemble des considérations pertinentes, sans pour autant déterminer la manière dont elles doivent être prises en compte. Ne rompant pas tout à fait avec les pratiques en vigueur, le rapport recommandait toutefois au gouvernement de se donner un seuil de tolérabilité, qu'il fixait à une probabilité de 10^{-4} de provoquer une mort par année d'activité pour « toute grande usine industrielle », et de 10^{-5} pour les nouvelles centrales nucléaires.

2.4.3.2 Insuffisance du risque acceptable

Les critiques des idées d'arbitrage entre bénéfices et risques et de niveau de risque acceptable furent bien sûr nombreuses, et consistèrent le plus souvent à souligner le caractère réducteur des représentations proposées, qui transformaient un ensemble de situations très disparates et évolutives à un échantillon de cas statiques décrits à partir d'un, deux ou trois arguments. Dans une réplique célèbre à Starr, Fischhoff et al. (1978) menaient une étude psychométrique auprès d'un groupe de femmes et montraient que si le niveau de risque estimé acceptable était effectivement corrélé aux

29. Voir à ce propos le chapitre 4.

30. *As Low As Reasonably Achievable*, ou aussi bas que raisonnablement réalisable.

bénéfices attribués à une activité, il était également toujours inférieur au niveau de risque effectif perçu par les individus.

2.5 Conclusion : la nécessité d'un cadre d'analyse

Les évolutions retracées dans les deux premières parties de ce chapitre montrent l'ampleur de la décentralisation des décisions en matière de risques sociaux dans la société française au cours des deux dernières décennies : une réglementation au contenu toujours plus technique, mise en œuvre par des agences réglementaires et des opérateurs privés ; de nouveaux outils juridiques dont la portée effective est définie par les tribunaux ; des collectivités territoriales et des citoyens qui parviennent dans une large mesure à influencer la mise en œuvre des décisions publiques dans un sens qui leur soit favorable. La figure du décideur public représentatif faisant, au nom de la société, des choix optimaux face au risque, semble peu à sa place dans ce tableau³¹.

C'est dans un tel contexte que nous devons replacer les modèles normatifs présentés dans la dernière partie du chapitre et les applications qui en sont faites. Le cas du risque acceptable semble emblématique. Les fortes critiques dont la notion fait l'objet depuis quarante ans ne semblent pas avoir nuit à sa capacité à s'ancrer dans les pratiques des décideurs publics. Il se peut même qu'elles l'aient renforcée puisque, à l'image de l'étude de Fischhoff et al., elles reprenaient souvent l'idée d'acceptabilité. De nombreuses agences réglementaires ont ainsi défini des objectifs probabilistes qui correspondent au niveau de risque acceptable dans les activités placées sous leur surveillance.

Une explication possible de ce succès est la simplicité du risque acceptable et son utilité en tant que signal dans le jeu à décideurs multiples que constitue la gestion collective du risque : signal de l'agence réglementaire à l'entreprise exploitante au sujet de l'effort minimal de sûreté à réaliser, signal de l'industrie à la population exposée concernant le risque maximal qu'elle court. Dans une telle optique, le risque acceptable n'est plus un critère de décision, mais l'instrument de comportements stratégiques.

Faut-il, dès lors, renoncer à toute ambition normative ?

Nous ne le pensons pas. Il nous semble au contraire que dans le jeu complexe où nous nous situons, le modèle du décideur unique a des vertus indispensables et encore largement inexploitées.

La première de ces vertus est que les modèles d'utilité espérée et leurs variantes semblent d'irremplaçables outils de synthèse de l'information nécessaire à la régulation des interactions sociales. Il s'agit donc d'aider non pas tant la décision que l'analyse des situations de risque social et la structuration du débat public³².

La seconde vertu est que le modèle du décideur unique reste d'une grande force normative

31. Grelot et al. (2001, p.55) faisaient déjà cette observation à propos de la gestion des risques d'inondation.

32. Bouyssou et al. (2006) analysent une vaste gamme de modèles de décision multi-critères dans une telle optique.

pour peu que son usage soit adapté à l'état de la société. Certes, s'il est appliqué pour prescrire une solution optimale, alors il s'adresse effectivement à un décideur unique qui n'existe plus. Les contestations dont les applications « fermées » de l'analyse coûts-bénéfices font régulièrement l'objet en fournissent une indication. Mais si le modèle est mis en œuvre afin de décomposer un problème de décision complexe en éléments simples, en faisant la part des grandeurs objectives et des appréciations, des utilités et des probabilités, et enfin, comme nous le proposerons, de l'incertitude non probabilisable, alors son point de vue est celui, abstrait, de la société, que chacun peut adopter afin de contribuer à construire une décision sociale.

L'impossibilité sans doute définitive de parvenir à une formulation satisfaisante de problèmes de choix sociaux dans les termes de la théorie de la décision ne signifie donc pas, de notre point de vue, qu'il faille abandonner toute ambition normative pratique dans ce domaine.

Ayant clarifié notre position, nous nous cantonnerons dans toute la suite à l'intérieur des limites du modèle de décideur unique, pour tenter d'enrichir la synthèse d'une situation de risque social qu'il propose à la société.

Deuxième partie

Risques et sûreté nucléaires

Chapitre 3

L'analyse de la sûreté des centrales nucléaires

*Celui qui ne sait pas, et sait qu'il ne sait pas,
finit par mener sa mule à destination.
Celui qui ne sait pas, et ne sait pas qu'il ne sait pas,
trempe à jamais dans l'encre de l'ignorance.*
Proverbe persan

3.1 Introduction

Les risques nucléaires sont liés aux effets des rayonnements ionisants émis par les substances radioactives sur l'homme et sur l'environnement. Bien sûr, ces effets dépendent largement de la dose effective absorbée. Ainsi, pour l'homme, on estime aujourd'hui qu'une exposition aux rayonnements ionisants au-delà d'un certain seuil (1 Gy)¹ entraîne des effets à court terme, spécifiques, déterministes (c'est-à-dire certains) et dont la *gravité* est fonction de la dose absorbée, alors qu'une exposition plus faible a des effets à long terme, probabilistes et dont la *probabilité* est fonction de la dose absorbée. Les effets de l'exposition aux faibles doses sont cependant notoirement mal connus et constituent un domaine important de recherche en radiobiologie et en épidémiologie².

Les risques nucléaires sont de deux types. D'une part, des rayonnements ionisants sont émis non

1. Le gray (Gy) est l'unité de mesure de la dose absorbée, c'est-à-dire l'énergie de radiation absorbée par unité de masse, dans le Système international (SI). Un gray est défini comme la dose d'énergie absorbée par une masse d'un kilogramme soumise à un rayonnement ionisant apportant une énergie d'un joule ($1\text{ Gy} = 1\text{ J/kg}$). Pour apprécier les effets biologiques des radiations, il faut en outre tenir compte des différences d'effets entre rayonnements et des différences de sensibilité entre tissus. Le sievert (Sv) est l'unité de mesure de la dose effective, homogène au gray, qui tient compte de ces deux facteurs.

2. Voir, par exemple, Estève (2009).

seulement par les activités nucléaires proprement dites, mais aussi par l'imagerie médicale et même par le milieu naturel, de sorte qu'il semble inévitable qu'une part importante de la population soit exposée à de faibles doses de rayonnements. D'autre part, l'utilisation de substances radioactives engendre des risques d'accident. L'évaluation et la gestion des risques nucléaires s'est donc naturellement divisée en deux champs. Le premier, celui de la radioprotection, concerne l'exposition des personnes et des milieux ; il se situe en quelque sorte au point d'arrivée des rayonnements. L'autre, celui de la sûreté nucléaire, s'intéresse aux conditions de fonctionnement des installations (et des dispositifs de transport) nucléaires, c'est-à-dire à la source des rayonnements.

Par ailleurs, l'industrie nucléaire civile se décompose en une chaîne d'activités qui vont de la fabrication des combustibles, leur stockage et leur transport au stockage des déchets et au traitement des effluents, en passant bien sûr par la génération d'électricité dans des centrales nucléaires. Les risques liés à ces activités sont différents ; il doivent être évalués et pris en compte de façon spécifique. Nous nous concentrerons ici sur le cas des centrales nucléaires, qui constituent, du fait de leur complexité et des quantités considérables de produits de fission hautement radioactifs qu'elles renferment, la première source de risques accidentels d'origine nucléaire. Nous nous intéresserons plus spécifiquement aux centrales nucléaires à eau pressurisée (REP), qui sont le type de centrale nucléaire le plus répandu au monde, notamment en France (où l'intégralité des 58 réacteurs en activité, ainsi que l'EPR³ en construction sont des REP) et aux États-Unis (où il existe 69 REP en activité).

Dans la première partie de ce chapitre, nous décrirons à grands traits le fonctionnement d'une centrale de type REP et ses dangers, afin de préciser ce que l'on entend par sûreté. Dans les deux parties suivantes, qui forment le corps du chapitre, nous examinerons tour à tour les approches déterministe et probabiliste de la sûreté d'un point de vue profane. Largement inspirée à l'origine par les méthodes déterministes de calcul de fiabilité, l'analyse de la sûreté nucléaire a par la suite été marquée par l'apparition de modèles probabilistes, l'accumulation d'une expérience opérationnelle dans différentes parties du monde, enfin la remise en cause et le retour d'expérience provoqués par les accidents de Three Mile Island, de Tchernobyl et, plus près de nous, de Fukushima. Cette évolution est sans doute encore loin d'être achevée, comme en témoignent par exemple les interrogations persistantes sur la prise en compte des facteurs humains ou de l'incertitude. Davantage qu'une description détaillée des analyses de sûreté actuelles, nous proposerons dans ces deux parties de mettre l'accent sur la transformation progressive des méthodes et sur les avancées importantes qui semblent encore se dessiner dans certains domaines. La quatrième partie, enfin, comparera la façon dont les deux approches traitent de l'incertitude, ce qui nous permettra d'analyser la question de leur complémentarité.

3. *European Pressurized water Reactor.*

3.2 Préliminaires : risques et sûreté dans une centrale nucléaire

3.2.1 Description sommaire d'une centrale de type REP

L'électricité produite par un réacteur nucléaire provient de la chaleur dégagée dans son cœur par la fission d'atomes d'uranium ou de plutonium. Le combustible (dioxyde d'uranium UO_2 ou de plutonium PuO_2) est disposé dans le réacteur sous forme de pastilles formant des crayons, entourés de gaines qui sont constituées d'un alliage de métaux résistant à la corrosion, tel que le zirconium. La fission est causée par la collision entre un neutron et un noyau d'uranium ou de plutonium. Il en résulte une libération d'énergie, l'émission de deux ou trois neutrons et la production de deux fractions de noyau. La maîtrise de ces trois produits de la fission constitue le fondement de la sûreté des centrales nucléaires.

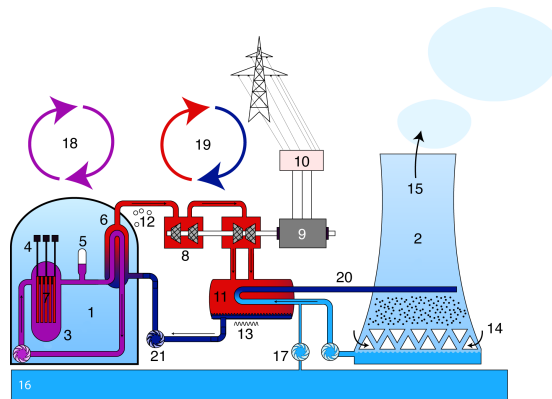
3.2.1.1 Le refroidissement du cœur

La chaleur est extraite par un fluide primaire* qui circule dans le cœur et à l'intérieur d'un circuit fermé (voir le schéma de la figure 3.2.1). Il s'agit, dans le cas des réacteurs en fonctionnement en France, d'eau que l'on maintient sous pression afin qu'elle puisse atteindre une température de 285° , qui assure un rendement suffisant. Le circuit primaire est constitué de la cuve qui contient le combustible, d'un pressuriseur et d'une ou plusieurs boucles, équipées d'une pompe afin de faire circuler le fluide. Chaque boucle est reliée à un générateur de vapeur. Le fluide, dont la température approche 300° au contact des gaines, échange sa chaleur à l'intérieur de ces générateurs avec l'eau d'un circuit séparé. C'est cette eau secondaire qui, en se vaporisant, alimente une turbine et un alternateur, et produit l'électricité. Après son passage dans la turbine, la vapeur est condensée dans un échangeur à tubes horizontaux, que l'on place sous vide pour en améliorer le rendement thermodynamique. L'échangeur fonctionne lui-même grâce à un circuit de refroidissement tertiaire. À sa sortie de l'échangeur, l'eau du secondaire est reconduite vers les générateurs de vapeur.

3.2.1.2 Le contrôle de la réaction en chaîne

Les neutrons émis lors de la fission entretiennent la réaction en chaîne. Afin de réduire la fréquence des opérations de renouvellement, on dépose dans le cœur une quantité de combustible largement supérieure au seuil (dit masse critique) nécessaire à l'entretien d'un niveau constant de réaction. Les neutrons sont donc régulés de façon à ce qu'à puissance constante, un seul neutron émis par chaque noyau cassé aille à son tour frapper un nouveau noyau. Des barres de commande introduits dans le cœur du réacteur et le bore présent dans l'eau primaire permettent de capter ou de ralentir les neutrons surnuméraires.

FIGURE 3.2.1 – Schéma d'une centrale nucléaire de type REP



Légende : (1) Bâtiment du réacteur ; (2) Tour réfrigérante ; (3) Cuve ; (4) Grappes de commande ; (5) Pressuriseur ; (6) Générateur de vapeur ; (7) Cœur du réacteur ; (8) Turbine ; (9) Alternateur ; (10) Transformateur ; (11) Condensateur ; (12) Vapeur ; (13) Fluide ; (14) Air ; (15) Air humide ; (16) Fleuve ou mer ; (17) Appoint ; (18) Circuit primaire ; (19) Circuit secondaire ; (20) Circuit tertiaire ; (21) Pompe.

3.2.1.3 Le confinement des produits radioactifs

Les fractions de noyau, que l'on appelle produits de fission, sont des éléments radioactifs artificiels. Il s'agit d'atomes instables, possédant un nombre de protons et de neutrons très variable, et dont la durée de demi-vie peut aller de quelques secondes à des millions d'années. Les produits de fission, qui constituent l'essentiel des déchets nucléaires, doivent être constamment confinés pour empêcher la dispersion de radioactivité. Un premier niveau de confinement est assuré par le gainage du combustible ; d'autres sont prévus dans la conception de la centrale. Au fur et à mesure de l'usage du combustible, la masse de produits de fission radioactifs présente dans le cœur augmente, dégageant une chaleur qu'il faut continuer à évacuer même lorsque le réacteur est arrêté. Les éléments de combustible usés sont ensuite retirés du cœur et entreposés dans une piscine de désactivation, où l'on attend qu'ils perdent une partie de leur radioactivité.

La description sommaire à laquelle nous venons de procéder suffit à mettre en lumière les trois sources fondamentales de risque dans une centrale nucléaire de type REP : le refroidissement insuffisant du combustible, qui conduirait à la fusion du cœur du réacteur ; la perte de contrôle sur la réaction en chaîne, qui peut conduire à une augmentation exponentielle de la puissance et à un accident de criticité*, dont la cinétique dépasse les capacités de réaction de l'homme ; enfin une défaillance du confinement des produits radioactifs. En outre, l'ensemble des circuits d'eau et de vapeur d'une centrale se caractérise par des différences de température et de pression considérables, qui sont également source de dangers.

Les centrales nucléaires renferment des radionucléides dans leur enceinte et en relâchent régulièrement à l'extérieur (dans l'air, dans des cours d'eau ou dans la mer) dans des conditions normales

de fonctionnement dans des quantités infinitésimales. En cas d'accident, ces relâchements peuvent devenir incontrôlés et d'ampleur beaucoup plus importante. Les risques liés à l'activité des centrales nucléaires sont donc liés d'une part à l'exposition du personnel, de la population extérieure ou de l'environnement de la centrale aux radiations et aux rejets chimiques dans des conditions normales de fonctionnement, de l'autre à la survenue d'accidents.

3.2.2 La sûreté nucléaire et son rapport avec le risque

L'Agence Internationale de l'Énergie Atomique définit la sûreté comme :

« [l']obtention de conditions d'exploitation correctes, [la] prévention des accidents ou [l']atténuation de leurs conséquences, avec pour résultat la protection des travailleurs, du public et de l'environnement contre des risques radiologiques indus » (AIEA, 2007).

En se référant à l'absence de risques radiologiques indus, cette définition implique d'une part que ces risques doivent être analysés de façon suffisamment précise et, de l'autre, qu'ils doivent être contrôlés et, le cas échéant, réduits. La sûreté a ainsi trait à la fois à l'évaluation et à la gestion des risques radiologiques.

Par ailleurs, telle que décrite ici, la notion repose sur un ensemble d'appréciations normatives concernant les conditions d'exploitation normales, le degré de prévention des accidents et celui d'atténuation de leurs conséquences et, en fin de course, l'ampleur des risques imposés aux travailleurs, à la population et à l'environnement. Ces appréciations ne sont pas librement établies par les acteurs en charge de la conception ou l'exploitation des centrales, comme le dit la même agence dans un document destiné à fournir des éléments de « bonne pratique » en la matière⁴ :

« Il faut procéder à une analyse de sûreté de la conception de la centrale en appliquant à la fois des méthodes déterministes et des méthodes probabilistes. À partir de cette analyse, il faut établir et confirmer la base de conception des constituants importants pour la sûreté. Il faut aussi démontrer que la centrale telle que conçue est capable de satisfaire à l'une quelconque des limites prescrites pour les rejets radioactifs et des limites acceptables concernant les doses potentielles de rayonnements pour chaque catégorie d'états de la centrale, et qu'une défense en profondeur a été assurée » (AIEA, 2005).

Il s'agit donc, pour les concepteurs et les exploitants, de mettre en œuvre une méthodologie d'ensemble, à partir de raisonnements pour partie déterministes et pour partie probabilistes, afin de montrer⁵ que des normes fixées par ailleurs sont respectées : des limites de rejets et d'exposition aux rayonnements dans certains états donnés de la centrale, ainsi que les conditions constitutives

4. Il s'agit ici spécifiquement de sûreté de la conception des centrales. Des principes généraux s'appliquent de la même façon à l'exploitation, aux modifications, à l'arrêt et au démantèlement des installations.

5. L'Agence emploie le verbe *démontrer*, qui fait effectivement partie de la terminologie usuelle du secteur. Nous reviendrons sur ce choix sémantique en conclusion du chapitre suivant.

d'une *défense en profondeur**. C'est à des autorités extérieures, en l'occurrence celles en charge de la réglementation de la sûreté, qu'il revient de s'assurer de la suffisance de ce dispositif pour protéger les travailleurs, le public et l'environnement contre des « risques indus » et d'envisager, si nécessaire, des actions correctrices.

Nous nous concentrerons dans la suite de ce chapitre sur la première de ces deux questions, celle des moyens mis en œuvre pour analyser les risques induits par les installations nucléaires et montrer que certaines exigences techniques sont satisfaites. Ce faisant, nous serons amenés à évoquer la façon dont ces exigences sont elles-mêmes déterminées. Cependant, nous remettrons pour l'essentiel la discussion de la seconde partie de la définition, celui du choix des critères de sûreté et de leur lien avec un niveau de protection de la société, au chapitre suivant.

3.3 L'approche déterministe de la sûreté

L'approche déterministe vise à introduire assez de marges dans la conception et la conduite d'une installation et assez de redondance dans ses dispositifs de protection pour qu'elle reste sûre dans les conditions les plus défavorables que l'on puisse raisonnablement envisager. Elle se caractérise par le calcul et le maintien de marges de sûreté dans l'utilisation des matériaux et des composants, et par une analyse de la sûreté de l'installation sous l'hypothèse d'occurrence d'événements contre lesquels on souhaite se prémunir⁶. On ne s'intéresse donc ni aux causes nécessaires ou suffisantes à la survenue de ces événements dommageables, ni à sa probabilité. Lorsque l'on estime que la *seule* survenue de l'un d'entre eux ne suffirait pas à causer un accident grave, on dit que l'installation est dimensionnée* pour cet événement.

Si l'approche déterministe de la sûreté nucléaire est reliée à l'idée de risque, c'est donc de manière relativement indirecte et largement focalisée sur la source de celui-ci, c'est-à-dire les trois dangers précités : la perte de contrôle de la réaction en chaîne, un refroidissement insuffisant du combustible et la perte d'étanchéité du confinement des produits radioactifs. C'est initialement le dernier de ces dangers qui a focalisé l'attention, conduisant notamment à l'élaboration de la méthode dite des barrières en France. L'expérience a toutefois rapidement révélé l'insuffisance de cette méthode, qu'il a fallu renforcer par une prise en compte plus complète des systèmes sur lesquels repose la sûreté d'une centrale. Ceci a donné naissance au concept de défense en profondeur, qui a lui-même été adapté pas à pas au progrès des connaissances relatives au fonctionnement des installations nucléaires (et continue à l'être).

Cette partie propose un survol de ces méthodes destiné à mettre en évidence certaines de leurs particularités. Nous y verrons notamment que si l'intégration systématique de marges de sûreté,

6. Nous consacrons principalement cette section à l'analyse déterministe de la sûreté des installations, renvoyant nos quelques remarques sur la fiabilité mécanique et les marges de sûreté à une discussion ultérieure (voir la sous-section 3.4.2.3).

la stipulation des incidents et accidents et le recours à des hypothèses conservatrices caractérisent l'approche déterministe, le qualificatif ne doit cependant pas tromper : l'approche déterministe n'exclut totalement ni la prise en compte de probabilités de réalisation des événements, ni celle de leurs conséquences ultimes. Il serait donc plus approprié de parler d'outils *principalement* déterministes.

3.3.1 Le principe de confinement et les fonctions de sûreté

Depuis la conception des premières centrales nucléaires à usage industriel, le fondement de la démarche de sûreté a été de confiner la matière fissile, combustible et produits de fission, par une succession de barrières physiques.

Ces barrières sont au nombre de trois dans la plupart des centrales aujourd'hui en fonctionnement. Dans les réacteurs à eau pressurisée, la première barrière est formée par les gaines qui entourent le combustible et l'isolent du fluide réfrigérant. Si les gaines viennent à fuir, des radionucléides se mêlent au fluide réfrigérant. Elles sont alors contenues dans le circuit primaire où circule le fluide. En cas de perte de l'étanchéité de cette deuxième barrière, une partie des particules radioactives se retrouve dans le bâtiment du réacteur, à l'intérieur d'une enceinte de confinement qui forme la troisième barrière à leur échappement. Dans certaines centrales de conception récente telles que l'EPR, une enceinte supplémentaire a été ajoutée pour disposer, au total, de quatre barrières de protection.

Pour les responsables de la sûreté, l'existence d'une telle succession de barrières devait permettre, dans une centrale par ailleurs bien construite et conduite avec prudence et vigilance, d'exclure l'hypothèse d'accidents majeurs⁷. Cependant, il est rapidement apparu que le principe de confinement ne pouvait, à lui seul, constituer une approche suffisante de la sûreté des centrales nucléaires.

En premier lieu, la définition des barrières, qui est sans équivoque dans des installations plus simples, est entachée d'exceptions dans une centrale. L'enveloppe du fluide primaire n'est pas limitée au circuit primaire, ni d'ailleurs située dans la seule enceinte de confinement. Elle se trouve aussi des bâtiments annexes où l'on épure le fluide, ajuste son volume et règle sa teneur en bore. La piscine de désactivation, qui est en contact avec l'air, est également considérée comme une partie de l'enveloppe du fluide primaire. Les lignes du circuit secondaire sont équipées de soupapes de sûreté et de vannes de décharge à la sortie des générateurs de vapeur et à l'extérieur de l'enceinte

7. Le principe de confinement a trouvé sa formalisation la plus poussée dans la méthode dite des barrières, que des responsables français de la sûreté nucléaire avaient élaborée au cours des années 1950 en enrichissant les approches anglo-saxonnes de leurs propres observations sur la prévention des évasions dans les camps de prisonniers de la Seconde guerre mondiale, et qu'ils ont mise en œuvre dans l'examen des premières centrales françaises (Bourgeois et al., 1996). La méthode des barrières comporte trois volets : l'incorporation de marges de sûreté suffisantes dans la conception des centrales ; la surveillance et la détection de toute entrée à l'intérieur de ces marges ; enfin des actions de sûreté destinées à leur rétablissement. L'examen de sûreté consistait à s'assurer que même dans des conditions de défaillance d'un matériel ou d'erreur humaine, l'étanchéité d'ensemble des barrières pourrait être préservée. Voir également la sous-section 4.1.1.

de confinement, afin de pouvoir relâcher de la vapeur dans l'atmosphère s'il y a surpression. Les tubes des générateurs de vapeur qui, à eux seuls, ont une surface totale supérieure à $10\,000\,m^2$, jouent donc à la fois un rôle d'enveloppe du fluide primaire et d'enceinte de confinement : en cas de brèche suffisamment importante dans un tube, les vannes et les soupapes de la ligne secondaire s'ouvrent et il n'existe pas de troisième barrière pour s'opposer au relâchement de radionucléides.

De surcroît, l'efficacité de la méthode reposait sur la redondance des barrières, qui suppose leur indépendance : la défaillance d'une barrière ne devait pas s'accompagner d'une augmentation significative des risques de défaillance des autres. Or l'expérience d'exploitation des centrales de recherche et des premières centrales industrielles a montré que diverses séquences accidentelles contrevenaient à l'hypothèse d'indépendance des barrières, de même qu'elle a prouvé la possibilité de deux types d'événements jusque là négligés : les pertes de réfrigérant primaire, qui conduisent rapidement à une fusion du cœur, et les accidents de criticité.

Le principe de confinement a alors dû être intégré dans une démarche plus complète, où l'on commence par identifier l'ensemble des *fonctions de sûreté** d'une centrale, c'est-à-dire celles dont l'accomplissement assure que les dangers inhérents à la manipulation d'une réaction de fission sont maîtrisés. Comme nous l'avons expliqué, au niveau le plus général, ces fonctions sont le contrôle de la réaction en chaîne, une évacuation suffisante de la puissance par refroidissement du combustible et le maintien de l'ensemble de la matière radioactive dans un confinement efficace. On peut parvenir à des classifications plus détaillées en remplaçant chacune de ces fonctions générales par plusieurs autres qui en sont des conditions nécessaires (et en constituent ensemble une condition suffisante). À titre d'exemple, le refroidissement du combustible peut être remplacé par la circulation du fluide primaire et le refroidissement de celui-ci.

Cette démarche, que nous allons étudier à présent, reprenait les barrières de protection passives du principe de confinement, mais considérait en outre les systèmes actifs de sauvegarde dont les centrales allaient être progressivement dotées afin d'assurer la permanence de leurs fonctions de sûreté.

3.3.2 Principes généraux de la défense en profondeur

Davantage qu'une méthode, la défense en profondeur est une idée directrice, que l'on peut ainsi résumer : toute installation nucléaire doit être dotée d'un ensemble de lignes de défense organisées de sorte que la protection des hommes et de l'environnement contre les rayonnements ionisants ne dépende jamais d'une ligne seule. Elle consiste à mettre en place, par des interventions à chaque stade de la vie d'une centrale (conception, construction, mise en service, exploitation, mise à l'arrêt, démantèlement), des équipements et des procédures destinés à prévenir tout événement qui pourrait menacer l'une des fonctions de sûreté, à le détecter précocement s'il devait malgré tout se produire

et à empêcher sa progression, enfin à en limiter les conséquences néfastes. Ces équipements et procédures doivent être organisés en une succession de lignes, chaque ligne étant conçue dans le but de limiter les conséquences d'une défaillance au niveau inférieur et d'éviter des dérèglements supplémentaires qui nécessiteraient une réponse au niveau supérieur (voir le tableau 3.1).

TABLE 3.1 – Objectifs et moyens de la défense en profondeur

Niveau de défense	Objectifs	Moyens principaux
1	Prévention du fonctionnement anormal et des défaillances	Conception selon des hypothèses conservatrices, qualité de réalisation et d'opération
2	Maintien de l'installation dans le domaine autorisé et détection des défaillances	Systèmes de régulation, de contrôle et de protection et autres mécanismes de surveillance
3	Maîtrise des accidents à l'intérieur des hypothèses de conception	Systèmes de sauvegarde et procédures de gestion des accidents
4	Prévention de la progression des séquences accidentelles et maîtrise de leur conséquences	Mesures complémentaires et gestion des accidents graves
5	Mitigation* des conséquences radiologiques de rejets importants	Plans d'intervention d'urgence extérieure au site

Source : à partir de Libmann (1996)

Le premier niveau de la défense en profondeur consiste d'abord à identifier et décrire les situations normales d'opération du réacteur (fonctionnement normal, *transitoires** normaux, certaines situations d'arrêt), ainsi que les scénarios d'*agression interne** (incendie, inondation) ou *externe** (secousse sismique, conditions météorologiques extrêmes, etc.) dont on estime que l'installation doit pouvoir les subir sans que son fonctionnement soit perturbé *ou* sans que cela conduise à des rejets importants de radionucléides. Il s'agit ensuite de dériver les contraintes physiques (mécaniques, thermiques, de pression, etc.) maximales auxquelles les différents constituants de l'installation sont soumis dans ces situations. Il faut enfin s'assurer de l'existence d'une marge de sûreté suffisante entre ces niveaux de contrainte et les limites techniques de chaque constituant, tel qu'il est conçu, mis en œuvre et exploité. L'ensemble des états de l'installation couverts par les marges de sûreté forme son *domaine de fonctionnement autorisé*. Le premier niveau comprend des contraintes strictes de qualité dans la conception, la construction et l'exploitation des installations visant à assurer qu'en situation normale, l'installation se trouve toujours à l'intérieur de ce domaine⁸.

L'objectif du deuxième niveau de défense est de maintenir l'installation à l'intérieur de ce domaine, grâce à des systèmes de régulation, de contrôle et de protection censés corriger toute évolution

8. En France, ces contraintes sont définies par l'arrêté ministériel et la circulaire du 10 août 1984.

anormale avant qu'elle ne conduise à une sollicitation excessive d'un matériel. Ces systèmes sont ceux qui permettent de réguler l'activité du cœur (température, pression et puissance neutronique ou thermique), de protéger celui-ci de l'effet des dérèglements lorsque la régulation s'avère insuffisante (soupapes de sûreté, procédures d'arrêt du réacteur), de surveiller l'efficacité des barrières et des systèmes d'épuration (mesures de la radioactivité et des conditions atmosphériques) et de contrôler périodiquement l'évolution des matériels (recherche de fissures, essai des matériels, etc.).

Le troisième niveau est conçu sous l'hypothèse de sortie du domaine autorisé. Il se fonde sur un ensemble hypothétique d'incidents et d'accidents, que l'on postule pour concevoir les systèmes automatiques de *sauvegarde* : injection d'eau de secours dans le circuit primaire, alimentation de secours en eau des générateurs de vapeur, aspersion à l'intérieur de l'enceinte pour faire baisser la pression, isolement automatique des traversées, etc. Le but, ici, est de ménager aux opérateurs un temps suffisant pour établir un diagnostic fiable avant d'intervenir sur l'installation, et de préserver l'intégrité de la structure du cœur afin de permettre un refroidissement ultérieur et de limiter les rejets dans l'environnement. Les accidents pris en charge à ce niveau peuvent conduire à des rejets incontrôlés, mais non à une fusion du cœur.

Le quatrième niveau est placé sous l'hypothèse d'accidents particulièrement graves (tels qu'un accident de manutention du combustible ou la rupture complète d'une tuyauterie primaire) ou de défaillances multiples (comme la perte totale d'alimentation électrique ou la défaillance de l'arrêt d'urgence du réacteur lors d'un transitoire). Comme au niveau précédent, les événements de référence sont déterminés *a priori*. En présence d'une défaillance multiple, il est fait appel à des systèmes alternatifs qui permettent de prévenir un grave endommagement du cœur du réacteur. Dans un cas d'accident grave, en revanche, il y a fusion plus ou moins complète du cœur, les deux premières barrières de confinement sont franchies et la défense repose donc principalement sur la tenue de l'enceinte de la centrale. Il s'agit alors, à travers le Plan d'Urgence Interne, de prolonger le confinement assuré par l'enceinte aussi longtemps que possible afin de protéger les populations extérieures.

Enfin, le cinquième niveau examine l'hypothèse où, les mesures précédentes ayant failli, des rejets importants peuvent avoir lieu. Des Plans Particuliers d'Intervention mettent en jeu des mesures de protection civile telles que l'évacuation ou le confinement de la population, la distribution prophylactique de tablettes d'iode, ou les restrictions de consommation. Ces mesures, qui relèvent de décisions des autorités de protection civile et non de l'exploitant, ne sont pas nécessaires lorsque le confinement de la matière radioactive dans l'enceinte de la centrale s'avère efficace.

Les trois premiers niveaux de la défense en profondeur sont censés assurer la maîtrise d'un ensemble d'événements pris en compte dans le dimensionnement de l'installation. Ce sont les seuls que la défense en profondeur comptait à son origine (et jusqu'au milieu des années 1970), lorsque

l'ambition était de parvenir à maîtriser tout événement à l'intérieur des limites de la centrale (voir la sous-section 4.1.1). Les concepteurs des centrales nucléaires étaient alors convaincus que les systèmes de sauvegarde et, en dernier recours, l'enceinte de confinement permettraient d'éviter des rejets importants de matières radioactives (sauf événement *a priori* extrêmement improbable et, de fait, non pris en compte).

La mise en échec des trois premiers niveaux (c'est-à-dire la survenue d'un accident entraînant de graves dommages du cœur) représente donc en soi une défaillance générale de l'installation. Le quatrième niveau de défense est constitué de mesures additionnelles (c'est-à-dire non intégrées à la conception) visant à limiter les rejets. Le cinquième niveau représente le passage de l'accident hors de l'enceinte, à partir duquel la défense relève de la responsabilité des autorités publiques. Sous l'hypothèse d'un dépassement des trois premiers niveaux, les objectifs de la défense en profondeur deviennent plus limités, et l'appréciation de son efficacité dépend largement du contexte et des conséquences finales de la séquence. Il existe donc une discontinuité importante, en termes de logique et de méthodes, entre le troisième et le quatrième niveaux de la défense en profondeur.

De ce fait, nous commençons dans la suite par analyser les méthodes et les conditions d'efficacité de la défense en profondeur dans les trois niveaux inférieurs avant de nous tourner vers les arguments et les choix méthodologiques qui ont présidé à la transformation du concept d'origine.

3.3.3 Méthodes et conditions théoriques d'efficacité de la défense en profondeur

Les dispositifs de la défense en profondeur sont développés pour répondre à des scénarios d'accident particulièrement défavorables, dits *accidents enveloppe**. Pour sélectionner ces derniers, on commence par identifier de la façon la plus complète possible les causes de perturbation de chacune des fonctions de sûreté. Puis l'on classe ces événements par degré de gravité. Ceux qui se rapportent au deuxième niveau de défense, par exemple, entraînent de simples variations singulières de paramètres concernant la criticité (retrait incontrôlé de grappes de régulation, dilution excessive mais limitée d'acide borique, etc.), le refroidissement (arrêt d'une pompe, baisse de la pression dans le circuit primaire, ouverture intempestive d'une soupape du circuit secondaire, etc.) ou le confinement (solicitation excessive de la cuve par la mise en service intempestive de l'injection de sécurité, etc.). Au troisième niveau, les événements retenus sont des défaillances sérieuses, telles que l'éjection d'une grappe de contrôle, la rupture d'une tuyauterie ou celle de tous les crayons d'un assemblage combustible en cours de manutention.

Une fois les événements pertinents ainsi répertoriés, les moins vraisemblables d'entre eux sont écartés. Dans l'approche d'origine, la vraisemblance des événements et le seuil en-dessous duquel elle pouvait être estimée négligeable étaient établis par jugement d'experts, sans que l'on cherche à

TABLE 3.2 – Conditions de fonctionnement dans les réacteurs nucléaires

Catégorie	Fréquence annuelle	Conséquences maximales admissibles ^a
Fonctionnement normal	quasi-permanent	$\sim 10 \mu Sv$ par an
Incidents	10^{-2} à 1	$\sim 10 \mu Sv$ par incident
Accidents	10^{-4} à 10^{-2}	5 mSv pour l'organisme entier 15 mSv pour la thyroïde
Accidents importants	10^{-6} à 10^{-4}	150 mSv pour l'organisme entier 450 mSv pour la thyroïde

a. Exposition maximale d'une personne située à cinq cents mètres de la centrale pendant deux heures.

les quantifier. Par la suite, les autorités en charge de la gestion de la sûreté ont défini des critères d'acceptabilité des incidents et accidents fondés sur des considérations de probabilité et de conséquences. L'autorité de sûreté nucléaire française⁹ considère ainsi, depuis 1975, quatre catégories de *conditions de fonctionnement**, qui correspondent chacune à une plage de fréquence estimée et une valeur limite des conséquences en termes de dose reçue par une personne située à cinq cents mètres de la centrale pendant deux heures : les « situations normales d'exploitation », les « incidents mineurs mais fréquents », les « accidents peu probables » et les « accidents importants mais hypothétiques » (voir le tableau 3.2)¹⁰. Le seuil d'acceptabilité de chaque catégorie, défini en termes de probabilités, est d'autant plus faible que les conséquences associées sont sérieuses. Ces conséquences sont toutefois considérées sous l'hypothèse formelle qu'aucun système ou barrière supplémentaire ne vient s'opposer au rejets de radionucléides. Elles constituent donc une mesure théorique de la gravité de l'accident et non une estimation de ses effets réels. De ce fait, *la relation entre probabilité et conséquences des conditions de fonctionnement ne peut être interprétée en termes d'acceptabilité du risque*¹¹. Elle constitue plutôt une norme de bonne conception et de bonne conduite de l'installation au sein de la démarche déterministe.

Il est important de noter que la méthodologie de la défense en profondeur distingue bien les deux phases de la sélection des conditions de fonctionnement pertinentes : en premier lieu, une identification aussi exhaustive que possible ; après seulement, un tri des événements les moins probables. Naturellement, la phase d'identification ne peut pas prétendre à l'exhaustivité, particulièrement dans une installation aussi complexe qu'une centrale nucléaire, qui compte des kilomètres de tuyau-

9. L'usage conventionnel veut que l'on désigne les instances publiques en charge de la sûreté nucléaire en France comme « l'autorité de sûreté nucléaire », y compris avant la création de l'Autorité du même nom.

10. Le terme officiel de condition de fonctionnement est quelque peu euphémistique puisque dans certaines de ces conditions, le fonctionnement de l'installation est définitivement arrêté. L'Agence Internationale de l'Énergie Atomique considère de son côté quatre catégories d'« états de l'installation » : l'exploitation normale, les déviations anticipées sans conséquence significative pour la sûreté, les accidents de dimensionnement et les accidents hors dimensionnement (nous utilisons ici une traduction approximative de la terminologie officielle de l'Agence, dont on trouvera la synthèse dans son glossaire de sûreté (AIEA, 2007)).

11. L'acceptabilité du risque, dont la détermination éventuelle relève des responsabilités du décideur public, est examinée plus en détail au chapitre suivant, notamment dans la sous-section 4.3.1.3.

terie, des milliers de vannes et des centaines de pompes. De plus, dans la phase de tri, l'évaluation des probabilités (et éventuellement des conséquences) est sujette à erreur¹².

En retenant les atteintes jugées les plus graves par des experts, on rassemble un très grand nombre d'incidents et d'accidents que l'on répartit en familles en fonction de leurs conséquences. À l'intérieur de chaque famille, on retient la séquence d'événements qui aurait les conséquences les plus défavorables ; c'est l'accident enveloppe. Lorsqu'il ne vérifie pas les conditions d'acceptabilité réglementaires, il faut réduire sa probabilité d'occurrence ou ses conséquences par l'intégration d'un dispositif supplémentaire (respectivement de prévention ou de maîtrise) dans la défense en profondeur. Dans le cas d'une défaillance d'une gravité extrême comme une rupture de la cuve, les exigences réglementaires en matière de qualité de conception sont si strictes que l'on considère, en pratique, l'événement impossible.

Les dispositions prises pour assurer la fiabilité de chaque ligne de défense sont de nature déterministe. La première est le *critère de défaillance unique*. Selon celui-ci, les systèmes nécessaires à la prise en charge d'incidents ou d'accidents (systèmes de protection et de régulation du niveau incidentel, systèmes de sauvegarde des niveaux accidentels maîtrisés ou graves) doivent pouvoir supporter la défaillance de l'un de leurs composants sans que cela n'entrave leur fonctionnement. Pour cela, les systèmes adoptés comportent généralement deux files, c'est-à-dire deux sous-systèmes étant capables d'assurer seuls la fonction d'ensemble¹³. L'efficacité du critère de défaillance unique dans un système redondant repose sur une condition d'*absence de défaillance de cause commune* (ou de « *modes communs* ») des sous-systèmes. Comme nous le verrons¹⁴, il est impossible de garantir le respect d'une telle condition. Pour s'en rapprocher, toutefois, il a été convenu de séparer géographiquement ou physiquement les deux voies, de diversifier les provenances de matériels similaires, voire même de confier les interventions de maintenance à des équipes différentes.

La seconde disposition principale est celle des cumuls conventionnels, qui consistent à se prémunir contre la survenue conjointe et fortuite de plusieurs événements défavorables. Le critère de défaillance unique impose en soi une forme de cumul, puisqu'il revient à supposer la survenue d'un incident ou accident et la défaillance d'un composant du système de sauvegarde concerné. La règle supplémentaire adoptée en France pour accroître le caractère conservateur de la défense en profondeur consiste à cumuler les effets d'une grosse brèche sur le circuit primaire ou secondaire, qui est l'un des accidents les plus sérieux considérés au quatrième niveau de la défense en profondeur, avec ceux du séisme de référence et d'une perte totale des alimentations électriques externes. Le cumul de ces aléas permet de traiter de façon implicite du caractère de cause commune d'un séisme, susceptible d'affecter à la fois les fonctions de sûreté de l'installation et le réseau d'électricité nécessaire

12. Les méthodes d'évaluation des probabilités de défaillance sont analysées dans la suite.

13. Dans d'autres pays, on a préféré des systèmes à trois ou quatre files, chacune d'entre elles remplissant alors les deux-tiers ou la moitié de la fonction.

14. Voir en page 95.

aux interventions de secours.

Enfin, on évoque souvent une condition d'indépendance des niveaux de défense pour décrire le principe selon lequel tout débordement d'un niveau de la défense devrait être traité au suivant. Il s'agit d'une dénomination impropre puisque les niveaux sont dépendants par construction, la défaillance de l'un étant une condition nécessaire de celle du suivant. La condition que l'on cherche à qualifier est plutôt que les *défaillances de cause commune de systèmes de protection et de sauvegarde* assignés à des niveaux différents de défense doivent relever du risque résiduel. À titre d'exemple, la probabilité annuelle de l'ensemble des modes communs à des systèmes engagées dans les trois premiers niveaux ne doit pas dépasser 10^{-4} . Comme précédemment, on ne peut pas s'assurer du respect général de cette condition, mais tout au plus veiller à ce que les modes communs identifiables soient spécifiés et pris en compte dans la conception de la défense. Il s'agit essentiellement des agressions externes, en particulier les aléas climatiques et les séismes.

En résumé, il est possible de dégager trois conditions fondamentales à l'efficacité des trois premiers niveaux de la défense en profondeur : une prise en compte aussi complète que possible des incidents et accidents pouvant survenir, la fiabilité des dispositifs placés à chaque niveau pour prévenir ou répondre à leur survenue et l'indépendance entre niveaux. Du fait de ces conditions, plusieurs limites s'imposent à l'efficacité de la défense en profondeur, que l'on peut chercher à repousser mais qui ne peuvent être abolies : l'incomplétude de l'ensemble d'incidents et d'accidents identifiés, l'incertitude dans l'évaluation de la probabilité d'occurrence de ces événements et l'existence possible de causes communes à la défaillance d'au moins deux composants ou systèmes intervenant dans la défense.

3.3.4 L'évolution de la défense en profondeur

Ces limites sont devenues de plus en plus évidentes à partir de la fin des années 1960, à la lumière de travaux montrant d'une part qu'une fusion complète du combustible formant le cœur du réacteur ne pouvait être exclue, et contestant d'autre part la présomption d'inviolabilité de l'enceinte de confinement dans une telle situation. Cette dernière remise en cause a eu un impact considérable sur l'analyse de la sûreté puisqu'elle amenait, dans une démarche déterministe, à établir une équivalence entre fusion du cœur et rejets massifs de substances radioactives. Il devenait donc impératif de réviser les fondements de la défense en profondeur afin de mieux identifier, prévenir et gérer les situations d'endommagement important du cœur. L'expérience d'exploitation des centrales américaines, puis l'accident de Three Mile Island, venant confirmer certaines craintes des spécialistes, les années 1970-1980 ont vu des efforts importants se déployer en ce sens.

3.3.4.1 Prise en compte de nouveaux facteurs de risque

De toute évidence, la gamme d'événements pris en compte dans les défenses existantes était insuffisante, puisqu'elle se limitait aux défaillances internes, que l'on complétait par la perte totale des alimentations électriques externes, un petit nombre d'agressions internes (projectiles provenant des équipements à haute pression) et une agression externe (séisme)¹⁵. À partir du milieu des années 1970, cette liste fut progressivement étendue à d'autres agressions internes (incendies, inondations internes, chutes de charge, etc.) et externes (inondations externes, chutes d'avion, risques industriels, etc.). Pour tenir compte de ces événements dans le dimensionnement des centrales, on fit parfois référence aux précédents historiques, sans aucune considération de probabilité (comme c'était déjà le cas pour les séismes). Dans d'autres cas où l'on estimait pouvoir approcher les probabilités d'occurrence par des fréquences empiriques, celles-ci servirent de critère de sélection des événements de référence. C'est ainsi que l'on évalua la probabilité de chute d'un avion sur les éléments sensibles d'une centrale durant une année à 10^{-8} pour l'aviation commerciale, 10^{-7} pour l'aviation militaire et 10^{-6} pour l'aviation de tourisme. Sur cette base, il fut décidé de protéger les centrales contre les chutes d'avions de tourisme et, dans certains cas, d'avions militaires, et de négliger les risques liés aux avions de ligne¹⁶.

Par ailleurs, l'arrivée à maturité de méthodes probabilistes permit d'étudier de façon plus précise les conditions de survenue d'accidents que l'on avait exclus du dimensionnement des centrales et qui conduisaient directement à une fusion du cœur. Les événements les plus sérieux pris en compte jusqu'alors (les conditions de fonctionnement de quatrième catégorie) n'excluaient certes pas un endommagement significatif du cœur et des rejets incontrôlés de substances radioactives, mais ils devaient se solder, après l'entrée en jeu automatique des systèmes de protection et de sauvegarde, par le retour à une situation stable. Ce rétablissement ne pouvait pas être assuré pour les accidents de la nouvelle catégorie, qu'il fut convenu d'appeler « conditions de fonctionnement supplémentaires ». Il s'agissait généralement d'accidents résultant de défaillances multiples, telles que la perte totale des alimentations électriques, dont la probabilité fut estimée proche de 10^{-5} en dépit de la redondance des sources¹⁷, des tableaux électriques et des voies. Dans certains cas, il s'avéra que des systèmes importants pour la sûreté, comme celui d'arrêt d'urgence du réacteur, étaient exposés à des modes communs. Au total, l'analyse probabiliste montra que la probabilité de ces accidents graves, dont l'ordre de grandeur était estimé à 10^{-5} voire, dans certains cas, 10^{-4} , était sensiblement plus élevée

15. Comme nous l'avons vu dans la sous-section précédente, il a été décidé de cumuler conventionnellement l'une des défaillances internes les plus graves au séisme de référence et à la perte totale des alimentations électriques externes.

16. Il est remarquable que, dans ce cas spécifique et contrairement aux critères appliqués aux conditions générales de fonctionnement, les conséquences de la chute ne semblent pas avoir joué de rôle dans le choix des événements pertinents.

17. Pour son alimentation électrique, une centrale REP dispose de deux sources externes, deux groupes électrogènes et, en outre, de possibilités d'ilotage*.

que celle de conditions de fonctionnement d'une gravité comparable, telles que les grosses brèches du circuit primaire ou secondaire.

Ceci conduisit l'autorité française de sûreté nucléaire à définir pour la première fois, en 1977 et 1978, un objectif probabiliste relatif non plus à l'occurrence d'événements particuliers à l'intérieur de l'installation, mais au risque général engendré par elle. Il s'agissait, en l'occurrence, de veiller à ce que la probabilité de survenue de « conséquences inacceptables » ne dépasse pas 10^{-6} par an pour une tranche donnée, et 10^{-7} pour chaque famille d'accidents en son sein. Comme dans le cas des conditions de fonctionnement, cette dernière valeur définissait *de facto* un seuil de probabilité en-deçà duquel une famille d'accidents pouvait être ignorée. Lorsque cette limite n'était pas respectée pour un certain type d'accident, les concepteurs furent amenés à revoir le dimensionnement des installations et, dans certains cas, à le compléter par des dispositifs particuliers visant à réduire la probabilité ou les conséquences de l'accident. Cependant, ces améliorations n'ont pas toujours été suffisantes pour assurer le respect des objectifs probabilistes, qui n'avaient d'ailleurs qu'une valeur indicative selon l'autorité de sûreté elle-même (voir la sous-section 4.3.2).

Il faut noter que les objectifs probabilistes affichés par les autorités de sûreté amenaient à estimer la probabilité des principaux scénarios accidentels, et encourageaient ainsi à leur tour l'usage de méthodes probabilistes.

3.3.4.2 Three Mile Island

C'est dans ce contexte de changements méthodologiques et pratiques que survint, le 28 mars 1979, l'accident de Three Mile Island (TMI), qui ne fit que renforcer les évolutions en cours. L'accident, où il y eut fusion du cœur mais pas de rejets importants, leva en effet de façon spectaculaire l'incertitude des spécialistes de la sûreté sur plusieurs questions importantes¹⁸.

En premier lieu, il apparut clairement qu'une succession de dysfonctionnements et d'erreurs humaines en apparence mineurs pouvaient conduire à un accident plus grave que le plus grave des accidents enveloppe retenus dans le dimensionnement des centrales (à savoir la perte du réfrigérant primaire par une grosse brèche instantanée doublement débattue, c'est-à-dire un sectionnement du tuyau des deux côtés). Plus spécifiquement, il fallait consacrer une attention particulière aux accidents déclenchés par une petite brèche sur le circuit primaire, certes moins graves mais aussi plus fréquents que les grosses brèches.

Deuxièmement, TMI prouva que les conséquences d'une fusion du cœur n'étaient pas toujours catastrophiques pour l'environnement, et que l'enceinte de confinement pouvait jouer un rôle crucial dans la protection de la population et du personnel. La tenue de la centrale à court et à long terme dans un scénario d'accident grave devint donc un paramètre crucial de la sûreté.

18. Nous n'évoquons pas ici la somme d'enseignements techniques spécifiques qui firent partie du retour d'expérience de TMI et vinrent par la suite modifier la conception et la conduite des installations nucléaires.

Troisièmement, la conduite des opérations durant les premiers jours de crise, alors que la crainte de rejets massifs était forte, fut estimée très insatisfaisante ; il s'avéra par la suite que les autorités en charge des décisions d'urgence avaient manqué à la fois de pratique et d'informations, en particulier au sein de l'agence de sûreté, la Nuclear Regulatory Commission (NRC). Il fallait donc mieux se préparer à la gestion d'accidents graves, ce que l'on chercha à faire en adoptant formellement deux niveaux supplémentaires de défense en profondeur.

Quatrièmement, le facteur humain apparut comme un rouage essentiel des dynamiques accidentelles, dans le sens négatif (erreur) comme positif (récupération). Il devint donc nécessaire de le prendre en compte dans l'approche de la sûreté. Sur le plan pratique, ceci conduisit notamment à des modifications importantes dans la conception des salles de commande et à la création d'une fonction d'ingénieur de sûreté-protection, chargé du contrôle et du soutien des équipes en situation accidentelle.

Certains de ces enseignements allaient dans le même sens que les conclusions de la première analyse probabiliste complète de la sûreté d'une centrale, l'étude WASH-1400 publiée quelques années auparavant (NRC, 1975). L'étude avait estimé la probabilité de fusion du cœur à $3 \cdot 10^{-5}$ pour les réacteurs à eau bouillante et $6 \cdot 10^{-5}$ pour les REP (c'est-à-dire entre un et deux ordres de grandeur au-delà de ce que l'on pensait précédemment) et indiqué, pour les secondes, l'importance des accidents de perte de fluide primaire suite à des petites brèches, événements qui étaient alors considérés secondaires. Elle avait aussi montré que conditionnellement à une fusion, la probabilité de perte de confinement était non négligeable et avait identifié un mode de ruine du confinement jusqu'alors ignoré dans les analyses déterministes¹⁹ (Beckjord, Cunningham et Murphy, 1993). Dans le même temps, elle avait décrit un nombre significatif de scénarios d'accident qui ne conduisaient pas à des rejets importants.

Mais WASH-1400 avait aussi soulevé de vives critiques, qui avaient conduit son commanditaire, la Nuclear Regulatory Commission, à se distancier de ses conclusions. En confirmant plusieurs enseignements de l'étude, TMI contribua beaucoup à montrer la pertinence de l'approche qui y avait été mise en œuvre²⁰.

3.3.4.3 Un concept en mouvement

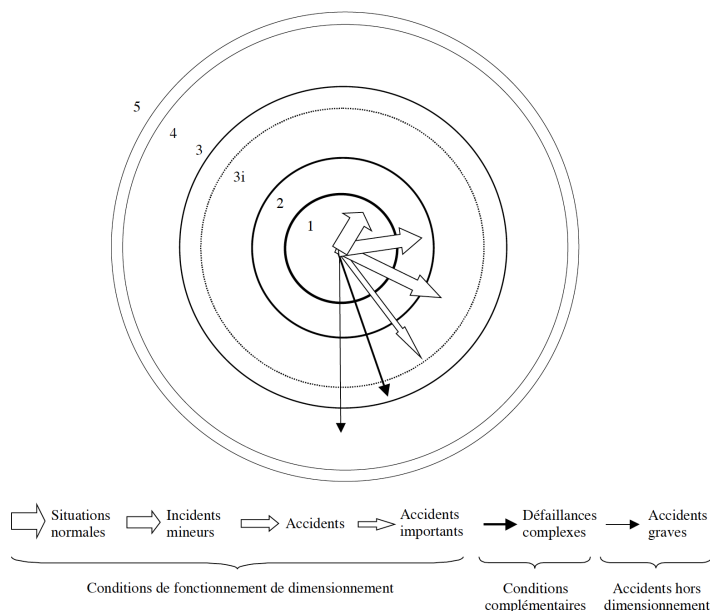
L'ensemble de ces influences ont contribué à transformer le concept de la défense en profondeur à partir du milieu des années 1970, notamment en y faisant entrer une part croissante d'éléments probabilistes et en l'adaptant à la possibilité de survenue d'un accident majeur.

En premier lieu, il s'est agi, comme nous l'avons vu, de se donner un critère de sélection des dé-

19. Il s'agit d'un scénario de perte de réfrigérant primaire par une brèche située à l'extérieur de l'enceinte de confinement (dit de bipasse).

20. Voir à ce propos la sous-section 4.2.1.2.

FIGURE 3.3.1 – Défense en profondeur et catégories de fonctionnement



faillances retenues pour le dimensionnement. Cette approche a été ensuite étendue aux événements externes, particulièrement lorsqu'un traitement déterministe conduisait à des contraintes de dimensionnement jugées excessives et trop coûteuses (comme dans le cas des chutes d'avions). Enfin, pour étendre l'approche à des événements d'une gravité extrême qui n'avaient pas été pris en compte dans le dimensionnement initial, on a défini des critères probabilistes pour la survenue de « conséquences inacceptables ». Ces critères ont à leur tour encouragé l'usage de méthodes probabilistes d'évaluation des séquences accidentelles.

Le dimensionnement des centrales a été adapté afin de tenir compte de ces « conditions de fonctionnement complémentaires ». Cela s'est traduit, en France, par certaines modifications d'installations ou de matériels, ainsi que par le développement des procédures de conduite dites H et U, qui permettaient de prévenir ou de faire face à la survenue de défaillances particulièrement graves.

La reconnaissance d'un risque résiduel de perte du confinement, en partie quantifié par des estimations de probabilités, rendait également nécessaire la création de deux niveaux supplémentaires de défense visant à limiter et retarder les retombées extérieures à l'installation. La figure 3.3.1 montre l'évolution des niveaux de défense en profondeur et leur rapport avec les catégories de fonctionnement. Le niveau 3i correspond au dimensionnement initial des installations, et le niveau 3 à celui obtenu après la prise en compte du risque d'accidents graves liés à des défaillances complexes (défaillances multiples, causes communes)²¹.

21. En toute rigueur, les niveaux 1 et 2 devraient également être modifiés pour refléter l'évolution du dimensionnement, puisque certaines défaillances complexes sont prises en charge dès ces niveaux.

L'élaboration des actions de sûreté des quatrième et cinquième niveaux de défense posait de nouvelles difficultés à la démarche déterministe. Le quatrième niveau était constitué de dispositifs non intégrés à la conception des centrales ; le cinquième niveau relevait d'interventions extérieures planifiées et menées par les autorités publiques. Dans un cas comme dans l'autre, la question du coût financier des mesures se posait de façon plus claire que lors du dimensionnement initial des installations, puisqu'un nombre important de centrales de puissance élevée existait déjà ou était en construction. Or la démarche déterministe, qui fut suivie dans un premier temps, avait tendance à conduire à un surdimensionnement manifeste des nouvelles mesures de protection. Le principe même de la démarche étant de majorer à la fois les probabilités et les conséquences d'un accident, toute notion de proportionnalité entre le risque et les actions correctrices entreprises lui était étrangère. Il fut finalement décidé de calibrer les mesures de réduction des risques de rejet des accidents envisagés au quatrième niveau en conformité avec le principe *ALARA*²² (AIEA, 2001). Après les probabilités d'accidents conduisant à une fusion du cœur et une perte du confinement, c'était donc leurs conséquences qu'il s'agissait d'estimer de façon plus réaliste.

La nécessaire prise en compte des probabilités et des conséquences de différents types d'événements a ainsi progressivement érodé la place du raisonnement déterministe dans l'outil déterministe par excellence qu'était la défense en profondeur. Le terrain de la sûreté était ainsi préparé à l'éclosion de méthodes probabilistes plus ambitieuses, qu'avait préfigurées l'étude WASH-1400.

Le concept de défense en profondeur a cependant continué à être adapté aux enseignements tirés de l'expérience ou des analyses, notamment probabilistes.

L'accident de Tchernobyl, dont les aspects techniques n'étaient guère transposables aux centrales occidentales en raison des différences technologiques importantes, montra cependant certaines difficultés pratiques de la conduite accidentelle et l'ampleur des moyens nécessaires à la gestion de crise. Il occasionna également le premier exercice de communication de crise nucléaire dans les pays européens, notamment en France où l'épisode mit en lumière des faiblesses institutionnelles sérieuses²³. Les niveaux quatre et cinq de la défense en profondeur furent ainsi consolidés à la lumière de la catastrophe.

L'accident de Fukushima a déclenché une remise en cause encore plus profonde, dont nous étudierons en détail les premières étapes dans la dernière partie du chapitre 6. S'il est tout à fait prématuré d'essayer de porter une appréciation générale sur le retour d'expérience de l'accident, il semble néanmoins certain que la défense en profondeur est entrée dans une phase d'évolution importante, qui devrait se traduire en particulier par un nouvel élargissement de la gamme des événements considérés (principalement des agressions extérieures) ainsi que par la prise en compte

22. Le principe *ALARA* (*As Low As Reasonably Achievable*) constitue un critère de mise en regard implicite des coûts et des bénéfices d'une mesure.

23. Voir la sous-section 4.3.2.1.

des possibilités d'accidents simultanés dans plusieurs réacteurs d'une centrale.

3.4 L'approche probabiliste de la sûreté

La différence fondamentale entre les outils d'analyse de la sûreté dits déterministes et ceux dits probabilistes ne réside pas dans l'usage des probabilités, qui n'est pas rare dans les premiers, comme nous venons de le voir, alors que les seconds peuvent aussi faire appel à des raisonnements déterministes. La particularité de l'approche probabiliste est la recherche d'une représentation explicite de la chaîne d'événements susceptibles de conduire à un certain résultat, qu'il s'agisse de la panne d'une composante, de la survenue d'un certain type d'accident suite à un événement initial ou de la sûreté d'ensemble de l'installation. La philosophie de l'approche probabiliste est résumée par la définition du risque proposée par Kaplan et Garrick (1981), que nous avons déjà évoquée : le risque est la donnée de tous les triplets (S_i, p_i, x_i) relatifs à un système donné, où S_i est un scénario (ou séquence) d'accident, p_i sa probabilité de réalisation et x_i ses conséquences. L'approche probabiliste se donne pour objectif d'identifier ces séquences de façon aussi complète que possible, d'évaluer leur probabilité de réalisation et de décrire leurs conséquences²⁴. Son intérêt, outre de quantifier les risques de défaillance, est de construire un modèle du système étudié *fondé sur les relations de cause à effet*, ce qui nécessite d'entreprendre un effort majeur d'acquisition de connaissances générales relatives au système et d'identification des interactions possibles en son sein.

L'inférence peut procéder par induction, comme lorsque sont élaborés des *arbres d'événements** pour représenter les séquences accidentelles déclenchées par des événements initiateurs, ou par déduction, comme lorsque l'on reconstitue les combinaisons d'événements pouvant conduire à la défaillance d'un système par l'intermédiaire d'*arbres de défaillance**²⁵. L'approche probabiliste s'intéresse non seulement aux systèmes importants pour la sûreté au sens traditionnel, mais également à la physique des phénomènes consécutifs à des événements particuliers, aux interventions humaines et aux systèmes informatiques. L'approche probabiliste combine ces outils en fonction du caractère spécifique du problème à étudier.

3.4.1 Les évaluations probabilistes de la sûreté

Lorsqu'elle est étendue à la sûreté générale d'une installation nucléaire, l'approche probabiliste prend le nom d'évaluation (ou étude) probabiliste de la sûreté (EPS)²⁶. L'agrégation de toutes les

24. Comme dans la partie précédente (voir note 6), nous laissons ici temporairement de côté les questions de fiabilité mécanique, pour nous concentrer sur l'analyse des systèmes et des séquences accidentelles.

25. Nous reprenons ici l'affirmation assez commune selon laquelle ces deux méthodes relèvent respectivement de démarches inductive et déductive (voir, par exemple, NRC (1981, section I p.8)). Dans la suite, nous montrerons toutefois que parler de déduction, dans le cas de l'analyse des systèmes, constitue un abus de langage typique (voir le paragraphe 3.4.1.1).

26. L'évaluation probabiliste de la sûreté correspond à l'anglais *Probabilistic safety assessment*, qui est le terme de référence des organismes internationaux tels que l'AIEA et l'AEN. Dans les pays anglo-saxons, notamment à des fins

séquences accidentelles permet alors de mesurer le risque total engendré par l'installation.

Il existe trois niveaux d'EPS : le premier porte sur l'évaluation du risque de fusion du cœur du réacteur ; le deuxième prolonge l'hypothèse de fusion en étudiant ses conséquences pour la centrale, et notamment les rejets de matière radioactive en dehors de son enceinte ; enfin, le troisième considère, pour un scénario de rejets donné, les conséquences pour le monde extérieur. Idéalement, les trois analyses devraient être articulées pour permettre de suivre l'enchaînement de causes et d'effets allant des prémisses d'un accident jusqu'à ses conséquences finales. En pratique, toutefois, l'accent a été surtout mis, jusqu'à ce jour, sur les EPS de niveau 1, pour des raisons que nous évoquerons dans la dernière section de ce chapitre. De telles évaluations ont été effectuées pour la grande majorité des réacteurs nucléaires en opération dans le monde, et constituent une exigence minimale de sûreté. Des EPS de niveau 2 ont été développées dans les pays les plus avancés en matière de sûreté (notamment en France), et apparaissent aujourd'hui comme une norme de bonne pratique (AIEA, 2001). Enfin, seul un faible nombre d'EPS de niveau 3 a été réalisé dans le monde (notamment aux États-Unis et en Suède).

Nous proposons dans cette section un survol de l'approche probabiliste de la sûreté, en prenant pour modèle les EPS françaises et en indiquant quelques alternatives méthodologiques lorsque cela était possible sans engager de longs développements techniques. Nos visées seront avant tout synthétiques et critiques ; pour des exposés complets et détaillés de ces outils, on pourra consulter notamment Bedford et Cooke (2001) et Villemeur (1988).

3.4.1.1 Le risque de fusion du cœur : l'EPS de niveau 1

L'analyse des séquences d'événements susceptibles de mener à une fusion du cœur du réacteur commence par la sélection d'événements *initiateurs**. Des arbres d'événements décrivent ensuite la progression de l'accident en considérant l'une après l'autre toutes les missions de systèmes ou d'opérateurs visant à l'arrêter.

L'usage combiné d'arbres d'événement et de défaillance à partir d'ensembles d'initiateurs et d'états critiques identifiés, l'analyse séparée des causes communes et des erreurs humaines, et l'évaluation des arbres de défaillance par l'intermédiaire de programmes informatiques (*codes*) forment ensemble la méthodologie de référence des EPS de niveau 1 des centrales nucléaires dans le monde.

Événements initiateurs

Un événement est initiateur lorsqu'il perturbe le fonctionnement normal de l'installation, nécessite le recours à un ou plusieurs systèmes de sûreté ou de sauvegarde et peut être considéré comme le début d'une séquence accidentelle. Les événements initiateurs peuvent avoir pour origine une

de réglementation, on lui préfère la dénomination de *Probabilistic risk assessment*, qui met l'accent sur la mesure du risque au prix d'un euphémisme. *Quantitative risk assessment*, plus précis, est également d'usage courant et semble progressivement gagner les faveurs des analystes.

défaillance interne (rupture de tuyauterie, panne de composant, erreurs humaines ou informatiques), une défaillance externe (perte des alimentations électriques), une agression interne (inondations et incendies) ou une agression externe (aléas naturels, accidents de transport).

L'identification des événements initiateurs vise, par principe, à l'exhaustivité. On ne peut bien sûr jamais être certain d'avoir atteint cet objectif et, dans la pratique, on en reste assez loin dans le cas de systèmes aussi complexes qu'une centrale nucléaire. Il s'agit alors de procéder, en considérant les ressources humaines et financières disponibles, à un arbitrage entre l'étendue des catégories d'événements considérés et la précision des séquences étudiées. La première étude probabiliste de la sûreté des réacteurs de 900 MWe en France, à titre d'exemple, a fait l'objet d'une attention particulière pour l'analyse des états transitoires et la prise en compte des défaillances de systèmes supports ; mais les agressions internes et externes en ont été exclues (IPSN, 1990b). Ce choix a été confirmé dans les mises à jour ultérieures de l'étude (IRSN, 2006), de même qu'il a été suivi dans l'EPS des réacteurs 1300 MWe (EdF, 1990).

Les événements initiateurs constituent le point de départ des séquences d'accident, c'est-à-dire celui où l'on choisit d'arrêter la démarche inductive de représentation des causes. Ce choix est, dans une large mesure, affaire de jugement et de contexte. Pour pouvoir considérer un événement comme initiateur, il faut d'une part que l'on puisse évaluer sa probabilité d'occurrence de façon satisfaisante, et de l'autre que l'effort supplémentaire de remontée à ses causes ne présente pas d'intérêt du point de vue d'autres séquences accidentelles. Si, à titre d'exemple, on choisit comme événement initiateur la rupture d'une tuyauterie primaire et non la fissure qui l'a amenée, c'est que la fréquence de la rupture est estimée suffisamment bien connue, et que la rupture est le point de passage unique de toutes les conséquences pertinentes de la fissure, quelle qu'en soient les causes (Villemeur, 1988, p. 272). Dans la terminologie des réseaux bayésiens causaux que nous introduirons au chapitre suivant, nous dirons que les grandeurs latentes, c'est-à-dire celles qui sont laissées en dehors du champ du modèle (la fissure), ne sont causalement liées qu'à une seule variable du modèle (la rupture), ou encore qu'il n'existe pas parmi elles de cause commune.

Les événements initiateurs, en revanche, ne sont pas indépendants : l'occurrence d'un événement peut accroître la probabilité de survenue d'autres, et la recherche de ces corrélations constitue un élément important de la démarche.

Les probabilités de survenue des événements initiateurs sont, dans la mesure du possible, estimées par des fréquences observées empiriquement. Ces estimations sont régulièrement révisées pour tenir compte de l'expérience d'exploitation. Si l'événement est rare et l'on ne dispose pas de fréquence fiable, sa probabilité d'occurrence est évaluée sur la base d'avis d'experts. Lorsqu'il s'agit d'une perte de système, ces avis peuvent être complétés, ou remplacés, par des outils d'analyse des systèmes tels que les arbres de défaillance²⁷. Lorsqu'il s'agit d'une erreur des opérateurs, on peut

27. Cependant, pour certains systèmes relativement élaborés, on dispose d'une base de données de fiabilité suffi-

faire de même appel à des méthodes d'analyse de la fiabilité humaine ou à des fréquences observées lors d'exercices de simulation.

Les événements initiateurs dont la probabilité de survenue est inférieure à un certain seuil (généralement 10^{-8}) sont considérés comme négligeables et écartés.

Analyses des séquences d'accident

Les arbres d'événements sont des diagrammes probabilistes issus de la théorie de la décision. Ils représentent une séquence accidentelle, depuis l'événement qui l'initie jusqu'aux différents événements qui peuvent la conclure, en décrivant toutes les interventions humaines ou de systèmes censées maintenir ou ramener l'installation à l'intérieur de son domaine de fonctionnement autorisé²⁸. Chacun des nœuds de l'arbre représente ainsi une mission : on parle de *mission de système* lorsque la séquence suppose en principe l'intervention d'un système de sûreté, et de *mission de conduite* lorsqu'il s'agit d'une intervention humaine. Deux branches sont issues de chaque nœud ; ils figurent la réussite et l'échec de la mission. Le succès d'une mission de système est défini par le respect de critères quantitatifs traduisant le maintien des fonctions de sûreté, et celui d'une mission de conduite par la réalisation d'une action appropriée et l'absence (ou la récupération) d'actions inopportunes dans les délais impartis. La succession des nœuds de l'arbre fournit une description de la séquence accidentelle. Pour partie, elle suit l'ordre de réalisation des causes et de leurs effets. La logique de progression causale ne suffisant cependant pas à déterminer l'arbre de façon univoque (puisque les événements considérés ne découlent généralement pas les uns des autres de façon linéaire), l'ordre des nœuds est aussi pour partie déterminé par l'analyste, avec pour principal souci de simplifier la structure de l'arbre. À titre d'exemple, entre deux défaillances qui ne seraient pas directement liées, celui-ci choisira de faire figurer en premier celle qui mènerait de façon certaine à une fusion de cœur ou, s'il y a égalité selon ce critère, celle qui y mènerait le plus rapidement.

Considérons, à titre d'illustration, la séquence accidentelle déclenchée par une perte totale de l'alimentation électrique²⁹. Dans une telle situation, la circulation du fluide primaire s'arrête et le cœur du réacteur n'est plus refroidi. Une série de missions doit alors être menée à bien pour éviter la fusion du cœur :

- la chute automatique des grappes d'arrêt doit placer le cœur en état sous-critique ; son échec entraîne la fusion à brève échéance ;
- le système d'alimentation de secours des générateurs de vapeur doit permettre d'évacuer la

samment large pour estimer bien établis les taux de défaillance et les durées d'indisponibilité qui en sont dérivés. C'est le cas, par exemple, des groupes électrogènes de secours.

28. Certaines études, en particulier WASH-1400 (NRC, 1975), choisissent de ne représenter que les interventions de systèmes dans l'arbre d'événements et de considérer les interventions humaines à l'intérieur de l'analyse des systèmes.

29. Nous présentons ici de façon succincte l'analyse de cette séquence dans les centrales 900 MWe françaises, telle qu'elle a été menée au début des années 1980. Comme l'explique Libmann (1996, p.180), des études ultérieures ont amené à en modifier certains scénarios (notamment à tenir compte de la possibilité d'une survenue de l'événement initiateur pendant un arrêt du réacteur) et à réévaluer certaines probabilités.

- puissance résiduelle du cœur ; sa défaillance conduit à une fusion au bout d'une heure ;
- le groupe turbo-alternateur de recours ultime (LLS) doit alimenter les batteries qui permettent le fonctionnement du contrôle-commande ; à défaut, ces batteries se déchargent en une heure, conduisant à la fusion au bout d'une heure supplémentaire ;
- le même groupe doit alimenter la pompe de test qui permet de continuer à injecter de l'eau froide aux joints des pompes primaires, évitant ainsi la perte de leur étanchéité qui conduirait à une fusion au bout de trois heures.

Dans chacun des trois derniers cas de défaillance, la récupération d'une source électrique permettrait en principe d'éviter la fusion du cœur.

L'arbre d'événements de la figure 3.4.1 représente une séquence accidentelle limitée aux trois premières heures d'absence totale d'alimentation électrique. Comme on le voit, la considération du système d'arrêt d'urgence comme premier nœud de l'arbre permet de ne faire figurer aucun autre nœud sur la branche « échec », puisque celle-ci mène de façon rapide et irréversible à la fusion.

On a également choisi ici de ne pas détailler les suites d'événements conduisant à des conséquences qui demeureraient incertaines dans la limite temporelle imposée à la séquence. L'une des branches de l'arbre agrège ainsi tous les scénarios où ni la fusion ni la sauvegarde du cœur ne sont obtenues au bout de trois heures. En pratique, lorsque l'on quantifie les probabilités d'accident à l'intérieur d'intervalles temporels donnés, ce type de branche est négligé.

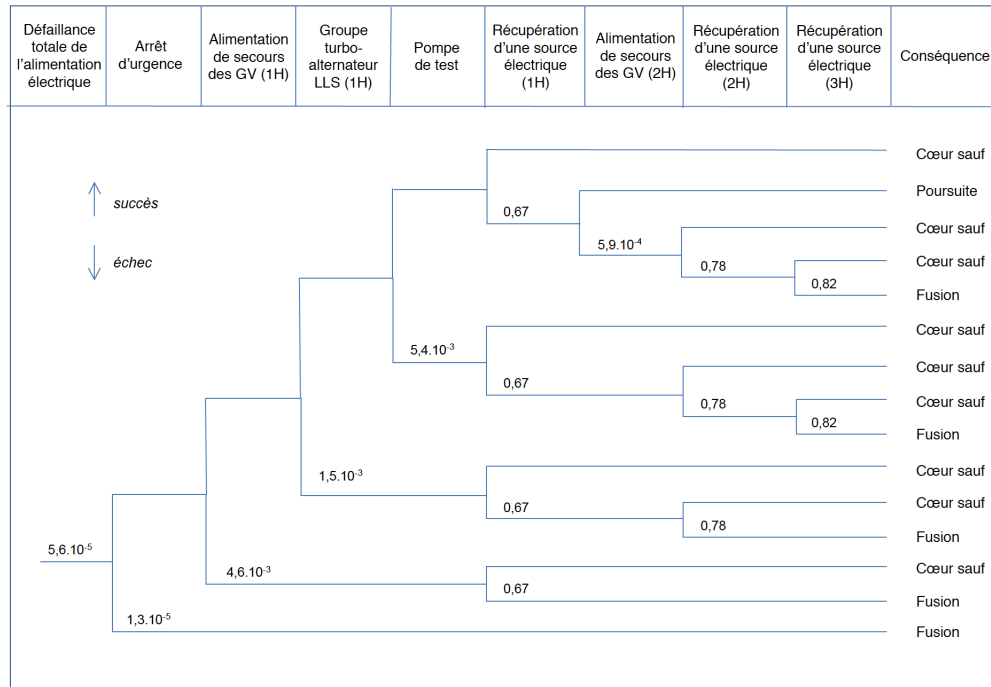
Le choix de la durée sur laquelle une séquence accidentelle est développée fait l'objet d'un arbitrage entre précision et simplicité. L'idéal, du point de vue de la représentation de l'incertitude, est de prolonger la séquence jusqu'à obtention d'états qui puissent être considérés définitifs dans tous les scénarios. Cette option, qui peut demander un effort considérable de modélisation, a été choisie dans l'actualisation de l'EPS française de niveau 1 publiée en 2006 (IRSN, 2006). Les séquences accidentelles s'y achèvent donc systématiquement par trois types d'événements :

- l'accident proprement dit, c'est-à-dire la fusion de tout ou partie du combustible du cœur, que l'on identifie en pratique à des causes estimées suffisantes pour le provoquer, telles que le dénoyage prolongé des crayons de combustible, des sollicitations de la cuve au-delà de ses conditions de tenue ou l'injection dans le cœur d'une eau dont la teneur en bore serait insuffisante ;
- une conséquence acceptable, compatible avec la préservation des fonctions de sûreté ;
- un événement initiateur d'une autre séquence.

D'autres EPS font le choix d'une plus grande simplicité en arrêtant l'analyse des séquences à un terme donné (généralement 24 h) et négligent de ce fait les scénarios accidentels à plus long terme.

Les probabilités de succès ou d'échec figurant sur les branches de l'arbre sont, comme dans le cas des événements initiateurs, estimées à partir de données de fiabilité, d'avis d'experts ou, dans le

FIGURE 3.4.1 – Arbre d'événements d'une séquence liée à la perte des alimentations électriques



Source : à partir de Libmann (1996)

cas de missions plus complexes, d'une analyse séparée du système ou de l'intervention humaine en question. Il s'agit, pour chacun des événements de l'arbre, de probabilités conditionnelles reflétant ses chances de réalisation sous l'hypothèse de survenue de tous les événements l'ayant précédé dans la séquence accidentelle. L'évaluation de ces probabilités suppose que l'on ait au préalable rendu compte de toutes les interactions possibles entre événements. Les différents systèmes en jeu dans une séquence accidentelle doivent donc faire l'objet d'une analyse aussi complète que possible en termes de rapports causaux. On cherche, en outre, à tenir compte des actions de réparation et de restauration qui peuvent être entrepris durant la période considérée (ici en vue de rétablir l'accès à une source électrique), en réponse soit à l'événement initiateur, soit à d'autres défaillances, et de leur impact sur la séquence accidentelle.

Analyse des systèmes

Pour analyser les conditions et probabilité d'échec d'une mission de système, on représente les combinaisons de défaillances de composants et d'erreurs qui peuvent y conduire par l'intermédiaire d'un arbre de défaillance. Les nœuds d'un arbre de défaillance sont constitués par des événements de trois types :

- l'événement de tête est une caractérisation de l'échec d'une mission spécifique d'un système ;

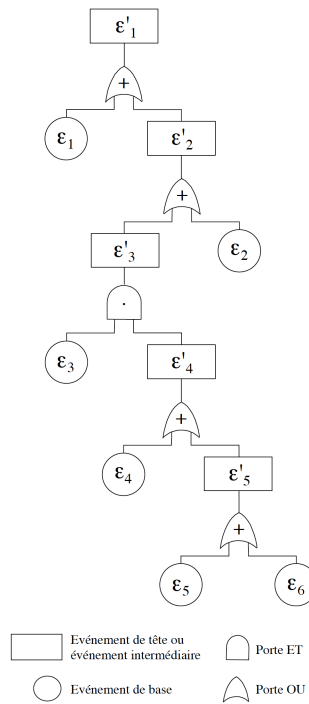
- les *événements intermédiaires* sont des défaillances de composantes du système ou des erreurs humaines dont on cherche à expliciter les causes comme pour l'événement de tête ;
- les *événements de base* sont des défaillances de composantes ou des erreurs humaines qui marquent les limites de l'analyse au sens où l'on choisit de ne pas en analyser les causes ; il faut pour cela que l'on puisse estimer leur probabilité de survenue de façon satisfaisante et raisonnablement supposer qu'ils n'admettent pas de cause commune, comme dans le cas des événements initiateurs.

Les branches de l'arbre relient des événements d'entrée à un événement de sortie, et décrivent l'action causale des premières sur le dernier. Elles sont jointes par des portes logiques traduisant la nature de la combinaison des événements d'entrée : *ET* s'il s'agit d'une conjonction, *OU* si c'est une disjonction, *ET* ou *OU* conditionnels à un événement particulier, combinaisons k/n (conjonction de k des n événements d'entrée), etc. Il s'agit donc, à chaque niveau de l'arbre, d'identifier les causes nécessaires et suffisantes à la survenue d'un événement³⁰.

La figure 3.4.2 présente un exemple d'arbre de défaillance : celui d'une cuve pressurisée et des systèmes de pompage et de contrôle associés. La cuve est initialement vide. Un interrupteur actionne un premier relais électrique, puis un second qui met le moteur de la pompe sous tension. La pompe injecte un fluide dans la cuve et le porte jusqu'à une pression limite en un temps estimé à 60 secondes. Une fois ce niveau atteint, un détecteur de pression (pressostat) permet, par le biais du second relais électrique, d'arrêter le moteur de la pompe. Afin de parer à une défaillance du pressostat, une minuterie a été introduite, qui est mise en marche dès que le premier relais est sous tension et qui interrompt automatiquement celui-ci s'il enregistre 60 secondes d'opération continue. L'événement de tête considéré (ε'_1) est la rupture de la cuve après le démarrage de la pompe. Les événements de base sont les *défaillances primaires** de la cuve (ε_1), des deux relais électriques (ε_5 et ε_2), du pressostat (ε_3), de l'interrupteur électrique (ε_4) et de la minuterie (ε_6). Les événements intermédiaires sont la *défaillance secondaire** de la cuve (ε'_2), le maintien sous tension du second relais électrique pendant plus de 60 secondes (ε'_3), le maintien de la tension pendant plus de 60 secondes en l'absence d'une action du pressostat (ε'_4) et le maintien sous tension du premier relais électrique pendant plus de 60 secondes en l'absence d'une action du pressostat (ε'_5).

30. Les combinaisons k/n (avec $k \leq n$) permettent d'expliciter de la façon la plus générale le rapport entre les portes logiques et les relations de causalité : dans une combinaison k/n , la conjonction de k événements d'entrée constitue une cause suffisante (mais non nécessaire dès lors que $k < n$) de l'événement de sortie, et chaque événement d'entrée est une cause nécessaire (mais non suffisante dès lors que $k > 1$) de cette conjonction. Chaque événement d'entrée est donc une cause de l'événement de sortie au sens de la condition INUS de Mackie : « une partie non suffisante mais nécessaire d'une condition qui est elle-même non nécessaire mais suffisante pour le résultat » (Mackie, 1965). Il faut noter en outre que les portes *ET* et *OU* sont des cas particuliers de combinaisons k/n , respectivement pour $k = n$ et $k = 1$.

FIGURE 3.4.2 – Arbre de défaillance d'une cuve pressurisée



Source : NRC (1981)

La logique de construction de l'arbre est détaillée dans l'ouvrage de référence de la NRC (1981). Elle consiste à tenter de répertorier de façon exhaustive les causes immédiates de chaque événement considéré, et de remonter ainsi aux causes de l'événement de tête que l'on choisit de considérer ultimes. À titre d'exemple, on commence par relier la rupture de la cuve à la double possibilité d'une défaillance primaire ou secondaire. La première, qui peut être notamment due à une fatigue accumulée, ne concerne que la cuve elle-même, que l'on a choisi de considérer comme une composante dans la description du problème ; de ce fait, cette défaillance constitue un événement de base. La seconde, en revanche, peut être causée par différentes combinaisons d'événements, que l'on détaille lors des étapes suivantes.

Comme nous l'avons souligné, cette logique d'identification des causes est souvent qualifiée de déductive. Il ne s'agirait cependant à proprement parler de déduction que dans la mesure où l'on disposerait d'un modèle explicatif complet du système en question, ce qui n'est jamais le cas. Il est ainsi admis que la recherche des causes ne peut prétendre à l'exhaustivité, et que seules les plus vraisemblables d'entre elles doivent être retenues (voir par exemple Villemeur (1988, p.191)). À titre d'exemple, on a écarté, dans la première étape de construction de l'arbre que nous venons d'exposer, la possibilité que la cuve ait été mal posée.

Un abus de langage voisin consiste à estimer que les rapports de cause à effet à l'intérieur du système, et plus généralement le système lui-même, sont déterministes, comme le fait par exemple l'ouvrage précité de la NRC :

« Un système est une entité déterministe comprenant un ensemble fini d'éléments en interaction » (NRC, 1981, p. I-4.).

Cette affirmation peut être comprise de deux manières. Soit elle se réfère à un déterminisme laplacien, qui nous est inaccessible car nos limites cognitives ne nous permettent pas d'appréhender l'ensemble des causes en jeu dans la détermination d'un effet particulier, et dans ce cas elle n'a pas d'implication pratique. Soit, comme le suggère l'emploi du qualificatif « fini », elle cherche à justifier le recours à un raisonnement déterministe pour décrire les interactions au sein du système, et alors elle se place elle aussi sous l'hypothèse implicite de complétude de la représentation choisie³¹.

Au total, c'est bien au prix d'une *exclusion de l'incertitude relative à l'action de causes ignorées* que les arbres de défaillance peuvent être construits selon un mode de raisonnement déductif et déterministe. Comme nous le verrons plus loin, les méthodes et doctrines de sûreté nucléaire ont eu une certaine disposition à l'oubli d'hypothèses simplificatrices de ce type et aux confusions sémantiques qui s'ensuivent.

Une fois l'arbre établi, sa *quantification* consiste à relier directement l'événement de tête à ses causes premières, les événements de base. En associant une variable booléenne à chaque événement, il est possible d'écrire la variable de tête en fonction des seules variables de base, puis de ramener son expression à une somme de produits de variables de base. Ces produits sont appelés *coupes minimales*. Chacune d'entre elles est donc une cause suffisante (mais non nécessaire) de l'échec de la mission de système, et les événements de base qui la composent sont autant de causes nécessaires (mais non suffisantes) à cet échec³².

Il ne reste plus alors qu'à calculer la probabilité de l'événement de tête à partir de celles des événements de base, qui sont données, ce qui engage généralement certaines approximations. Les méthodes les plus communes de dérivation des probabilités font l'hypothèse que les événements de base sont indépendants entre eux, ce qui nécessite qu'aucun événement de base ne soit la cause d'un autre, et qu'il n'existe aucun événement extérieur à l'arbre qui soit la cause commune de deux événements de base³³. La prise en compte de toutes les causes communes aux événements de base constitue, après celle de toutes les causes significatives des événements intermédiaires et de l'événement de tête, une seconde hypothèse causale de grande importance.

31. Nous reviendrons sur cette question dans le chapitre 5 en nous appuyant sur le formalisme des réseaux bayésiens causaux. Voir en particulier la sous-section 5.2.2.3.

32. Ces événements vérifient donc aussi la condition INUS (voir la note 30).

33. Nous supposons ici que pour qu'il y ait dépendance statistique entre deux variables, il faut qu'elles admettent une cause commune. Cette hypothèse, souvent appelée Principe de cause commune de Reichenbach, sera explicitée au chapitre suivant (voir la sous-section 5.2.2.2).

TABLE 3.3 – Probabilités de fusion du cœur d'un REP 900 MWe selon l'EPS « version APE »

Familles d'initiateurs	Probabilité
Défaillances de l'arrêt d'urgence lors d'un transitoire	$4,4 \cdot 10^{-8}$
Pertes de réfrigérant primaire	$1,2 \cdot 10^{-6}$
Ruptures de tubes des générateurs de vapeur	$2,6 \cdot 10^{-8}$
Ruptures de tuyauterie d'eau alimentaire	$1,3 \cdot 10^{-8}$
Ruptures de tuyauterie vapeur	$2,2 \cdot 10^{-8}$
Perte de la source froide ou des systèmes associés	$9,2 \cdot 10^{-7}$
Transitoires secondaires	$8,1 \cdot 10^{-7}$
Perte totale des alimentations électriques	$2,4 \cdot 10^{-6}$
Pertes de sources électriques	$2,2 \cdot 10^{-7}$
Pertes de réfrigérant primaire en dehors de l'enceinte de confinement	$5,0 \cdot 10^{-6}$
Transitoires primaires	$2,1 \cdot 10^{-6}$
Total	$1,3 \cdot 10^{-5}$

Source : IRSN (2006)

Nous détaillerons les méthodes de quantification et de calcul des probabilités au cinquième chapitre (section 5.3.3.2).

L'intégration du risque

Lorsque toutes les séquences accidentelles ont été quantifiées, il devient possible les regrouper selon différentes modalités : par famille d'initiateurs, par type de défaillance (défaillances d'un équipement spécifique, erreurs humaines, défaillances de mode commun), et bien sûr par type de conséquence (notamment fusion du cœur). Le tableau 3.3 présente la probabilité agrégée de fusion du cœur d'un REP 900 MWe pour toutes les familles d'initiateurs et pour une conduite accidentelle de type « approche par états » (APE)³⁴, selon la dernière version de l'EPS française de niveau 1 (IRSN, 2006). La probabilité annuelle totale de fusion du cœur par année d'opération y est évaluée à $1,3 \cdot 10^{-5}$.

Les résultats des EPS servent effectivement à orienter les décisions relatives à la sûreté des centrales, comme en témoigne la réévaluation du « référentiel de sûreté » adopté par les autorités françaises à la suite de l'actualisation de 2006. Cette réévaluation a notamment amené l'autorité de sûreté nucléaire à recommander des modifications de conception permettant de réduire le risque de fusion du cœur dans l'un des scénarios de perte de réfrigérant primaire avec bipasse de l'enceinte de confinement (Groupe Permanent Réacteurs, 2008). Cette famille d'initiateurs contribue à elle seule à hauteur de 40 % au risque total de fusion du cœur tel qu'il apparaît dans l'EPS.

Il n'est cependant pas possible de tirer, à partir des EPS 1 françaises, de conclusion au sujet du risque total de fusion du cœur. Comme il a été indiqué plus haut, l'EPS de niveau 1 française se concentre sur les défaillances de système et exclut de son champ les agressions internes et externes.

34. Pour une présentation de cette approche, voir la sous-section 3.4.2.1.

À titre d'illustration, l'EPS de cinq centrales américaines publiée par la NRC en 1990 incluait, pour deux des centrales, une analyse de certaines agressions externes. Or la contribution de celles-ci au risque total de fusion était, dans les deux cas, supérieure d'un ordre de grandeur à celle des initiateurs internes (NRC, 1990). En raison des différences de réalités physiques et de choix méthodologiques, ces résultats ne peuvent certes pas être extrapolés aux EPS françaises. Il n'en demeure pas moins vraisemblable que le risque lié aux agressions externes, loin d'être négligeable, représente en fait la portion la plus importante du risque total³⁵. Comme nous le verrons dans la section suivante, cette limitation ne constitue pas à proprement parler une incertitude affectant l'EPS, mais plutôt une partie de l'incertitude générale non prise en compte par l'EPS.

Par ailleurs, la modélisation des séquences accidentelles dans une EPS de niveau 1 ne doit pas être réduite à l'évaluation finale de probabilités de fusion du cœur pour des familles d'événements. L'EPS identifie en fait plusieurs milliers de séquences, dont chacune est caractérisée par un ensemble d'événements et de valeurs de paramètres qui décrivent de façon détaillée une situation accidentelle. Cette information est notamment nécessaire à l'élaboration de scénarios de progression de l'accident à l'intérieur de l'enceinte dans l'EPS de niveau 2.

3.4.1.2 Les rejets radioactifs en cas de fusion du cœur : l'EPS de niveau 2

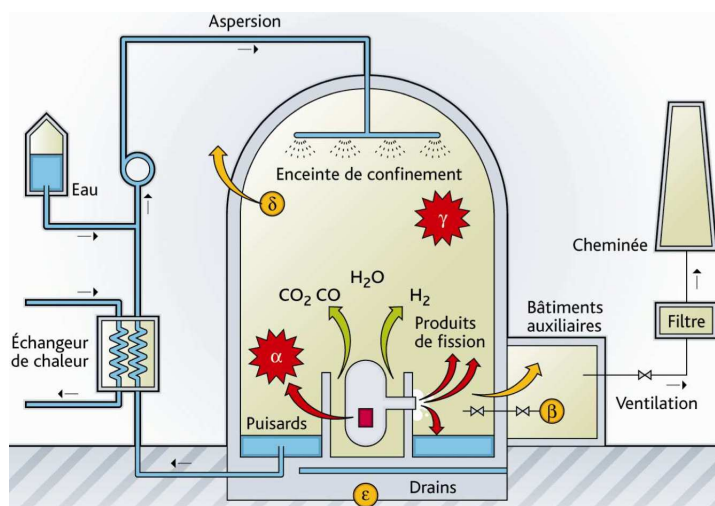
L'EPS de niveau 2 explore la progression des accidents par l'intermédiaire d'arbres d'événements intitulés *arbres de déroulement des accidents graves** (ADAG), jusqu'au rejet éventuel de produits radioactifs hors de l'enceinte de confinement. Elle a pour point de départ des états endommagés du cœur correspondant aux états finaux des scénarios d'accident élaborés par l'EPS de niveau 1, accompagnés des probabilités correspondantes et désormais considérés comme événements initiateurs. Le prolongement de ces scénarios par de nouveaux arbres entraîne cependant une augmentation exponentielle (fonction du nombre de niveaux de l'arbre, qui peut être important) du nombre de branches et des capacités de calcul requises. Il faut donc réduire le nombre d'événements initiateurs, et deux méthodes sont utilisées pour cela : la suppression des événements dont la probabilité est inférieure à un certain seuil et le regroupement des événements en fonction d'un ensemble de variables permettant de caractériser l'état de l'installation. La première méthode présente le double inconvénient de dépendre d'un seuil arbitraire et d'entraîner l'exclusion d'accidents aux conséquences très importantes au seul motif que leur probabilité est très faible. Aussi, les EPS récentes ont tendance à privilégier la seconde approche.

La fusion du cœur libère une énergie à l'intérieur de l'installation qui soumet celle-ci à des

35. Les deux incidents les plus graves (en termes d'accroissement de la probabilité conditionnelle de fusion du cœur) survenus dans des centrales françaises avaient pour cause des épisodes météorologiques extrêmes. Il s'agit de l'inondation de la centrale du Blayais lors des tempêtes de décembre 1999, à laquelle nous consacrons une partie du chapitre 6, et de l'obturation de la source froide par des matériaux charriés par la crue du Rhône à la centrale de Cruas en décembre 2009.

contraintes inhabituelles de température, de pression, de déformation ou de radiation. De même que les arbres d'événement des EPS de niveau 1 représentent le succès ou l'échec des missions visant à arrêter une séquence accidentelle, les arbres de déroulement des accidents graves décrivent les phénomènes physiques multiples découlant de l'accident, les concentrations de produits de fission dans différentes parties de l'installation³⁶ et l'action des systèmes de sauvegarde destinés à limiter les rejets³⁷. L'événement critique de l'EPS 2, à l'image de la fusion du cœur dans l'EPS 1, est la ruine ou la perte d'intégrité de l'enceinte de confinement. C'est à elle que tient la différence entre un accident grave, comportant éventuellement des rejets importants mais *a priori* maîtrisés, et un accident conduisant à un échappement massif et non maîtrisé de matière radioactive.

FIGURE 3.4.3 – Modes de ruine de l'enceinte de confinement



Cinq modes de ruine de l'enceinte de confinement ont été identifiés par l'étude WASH-1400, et continuent à structurer les EPS 2 à ce jour (voir figure 3.4.3). Le mode α représente une explosion de vapeur dans la cuve qui conduit à la rupture rapide de l'enceinte. Le mode β décrit un défaut d'étanchéité de l'enceinte préexistant à l'accident ou rapidement causé par lui. Le mode γ désigne une explosion d'hydrogène. Le mode δ est constitué par l'augmentation progressive de la pression jusqu'à la perte de confinement. Enfin, dans le mode ϵ , le *radier** de la centrale est traversé par le *corium**.

Les branches de l'ADAG étant extrêmement nombreuses, elles font l'objet d'un nouveau regroupement en Familles de Progression d'Accident (FPA). Chaque famille agrège des branches de l'arbre présentant des caractéristiques communes en termes de rejets de substances radioactives en dehors de l'enceinte. Le résultat final de l'EPS 2, associé à chaque FPA et à sa probabilité de réalisation,

36. Dans le circuit primaire, l'enceinte de confinement, les circuits secondaires et auxiliaires, et enfin dans le sol.

37. Système d'aspiration de l'enceinte, ventilation et filtration des bâtiments auxiliaires, éventage et filtration de l'enceinte.

est un *terme source**, c'est-à-dire une quantité de produits de fission rejetés, fraction de la totalité existant dans le cœur.

3.4.1.3 Les conséquences pour l'environnement en cas de rejets : l'EPS de niveau 3

L'EPS de niveau 3 est d'une nature différente des deux précédentes, puisqu'elle décrit les conséquences des accidents dans un environnement non confiné. Elle est constituée de modèles de dispersion de la matière radioactive (dans l'air, l'eau et la terre) et d'impact sur la santé humaine, sur celle des écosystèmes et sur les actifs économiques (notamment agricoles), culturels, sociaux, etc. Elle comprend en outre une évaluation du coût des mesures de gestion de crise (évacuation, soins d'urgence, etc.).

L'évaluation de ces éléments rencontre certains des problèmes bien connus de l'analyse coûts-bénéfices (AEN, 2000) : choix et justification des méthodes de valorisation de la vie et la santé humaines (consentement à payer, consentement à accepter, capital humain) ; valorisation d'autres biens non marchands ; choix d'un ou plusieurs taux d'actualisation en fonction de la nature et l'échéance des effets ; détermination du coefficient d'aversion pour le risque, etc.

Par leur nature même, les accidents nucléaires posent en outre des difficultés spécifiques. Leurs répercussions secondaires sur différents marchés ne peuvent pas être estimées négligeables, comme cela est habituellement supposé dans les méthodes de calcul de surplus. Dans une hypothèse d'accident majeur, les conséquences de la relocalisation de dizaines ou centaines de milliers de personnes pour les marchés du logement et du travail, ou celles de la perte des ressources productives d'une région agricole pour les marchés de l'alimentation illustrent ce point. Un accident majeur aurait des conséquences pour l'image de la région et même du pays affecté, se traduisant par des pertes commerciales ou une réduction de l'investissement. Enfin, comme l'illustre l'expérience japonaise depuis l'accident de Fukushima, un tel événement a de grandes chances d'avoir un impact dévastateur pour l'industrie nucléaire elle-même, entraînant notamment des surcoûts considérables liés à la gestion du parc de centrales existantes (améliorations de sûreté, fermetures anticipées, etc.).

Tous ces facteurs ne sont généralement pas pris en compte dans les EPS de niveau 3, rendant leurs résultats sujets à de très lourdes incertitudes.

3.4.2 Limites et incertitudes de l'approche probabiliste

L'objectif des méthodes que nous venons d'exposer est de parvenir à une quantification de l'incertitude, destinée à sa prise en compte explicite dans des décisions relatives à la sûreté. Ces méthodes reposent cependant sur des représentations de l'installation et de son environnement qui peuvent elles-mêmes être incertaines. De ce fait, elles ne mesurent qu'une partie de l'incertitude générale relative aux dangers d'une centrale nucléaire.

Nous commençons dans cette section par donner deux exemples de limitations importantes des EPS, avant d'analyser les réponses théoriques et le traitement pratique proposés pour tenir compte des incertitudes afférentes.

3.4.2.1 Deux aspects particuliers

Nous avons évoqué plus haut quelques enseignements tirés de l'accident de TMI au sujet de la défense en profondeur, et souligné l'importance de cet événement pour les évaluations probabilistes de sûreté, qu'il contribua à relancer quatre ans après la publication de l'étude WASH-1400. Cependant, tout en confirmant la pertinence d'une approche systématique fondée sur les probabilités, l'accident mit aussi en lumière les lacunes d'une représentation de la sûreté qui, à l'instar de celle de l'étude, ne rendait pas suffisamment compte de l'influence des actions humaines et des défaillances de cause commune. Ces deux questions difficiles firent l'objet de nombreux travaux dans les années suivantes.

Prise en compte du facteur humain

La prise en compte *a priori* des interventions humaines dans le fonctionnement des installations nucléaires est l'un des aspects les plus problématiques de l'approche probabiliste.

L'homme étant, par opposition à la machine, doué d'un certain degré de libre arbitre, ses actions sont moins prévisibles et il serait encore plus illusoire de prétendre les répertorier de façon exhaustive. Ceci est vrai pour des actes ou des comportements dommageables, tels que la négligence ou l'erreur, aussi bien que réparateurs, comme la restauration de systèmes défaillants ou la récupération d'erreurs. On se borne généralement, dans la pratique, à tenter d'identifier, *parmi les interventions humaines planifiées* au cours des séquences accidentelles, celles dont l'échec contribue le plus au risque d'accident³⁸ et à évaluer, pour ces interventions, la probabilité d'une erreur non récupérée. En d'autres termes, il ne s'agit pas de représenter les problèmes d'opérateurs doués de libre-arbitre et devant choisir leur conduite parmi un ensemble d'actions possibles, à la manière de la théorie de la décision, mais le succès ou l'échec d'interventions humaines codifiées et normalisées, prescrites dans les procédures de gestion des états incidentels, accidentels ou d'urgence, qui tiennent dans le fonctionnement de l'installation un rôle similaire à celui des missions des systèmes de sauvegarde.

Deux sources principales d'information sont exploitées à ces fins : le retour d'expérience, pour lequel il existe une classification des actions et une collecte systématique de données au niveau international ; et l'utilisation de simulateurs, qui est particulièrement précieuse pour l'étude des comportements humains en situation incidentelle ou accidentelle. Pour dériver des probabilités

38. La contribution au risque s'entend ici en termes de probabilité et de conséquences de l'erreur toutes choses égales par ailleurs ; plus précisément, il s'agit d'une espérance des conséquences négatives de l'intervention.

à partir des fréquences d'erreur observées auprès de ces deux sources, on procède fréquemment par analogie et par extrapolation (Swain et Guttman, 1983), notamment en considérant que le comportement observé en simulateur est identique à celui en situation réelle.

Or les modalités de conduite d'une action humaine dépendent de facteurs qui rendent délicates ces méthodes d'inférence. L'influence du contexte immédiat, qu'il s'agisse de l'information dont disposent les intervenants ou des conditions de stress qui leur sont imposées (délais sous lesquels l'intervention doit être réalisée, importance du danger, etc.), est difficile à mesurer. Mais celle de facteurs plus diffus tels que la culture de sûreté au sein de l'installation, l'importance des moyens économiques et humains consacrés à la sûreté ou la qualité de l'environnement réglementaire l'est encore davantage. En dépit du foisonnement de la littérature sur la fiabilité humaine au cours des deux dernières décennies, la prise en compte de ces aspects reste embryonnaire. Dans les EPS françaises, la seule disposition permettant de rendre compte de l'influence des facteurs contextuels est l'évaluation séparée des interventions, selon qu'elles sont menées en situation accidentelle ou dans des conditions normales de fonctionnement (IPSN, 1990c)³⁹.

Les interventions considérées dans ce dernier cas sont les opérations normales de conduite de l'installation, ainsi que des opérations de maintenance, de vérification et de test de ses composantes par les équipes dites de conduite. Des erreurs dans l'accomplissement de ces tâches peuvent, dans certains cas, constituer des événements initiateurs d'une séquence accidentelle ou rendre indisponible un système de sauvegarde. Leurs probabilités sont estimées à partir d'une probabilité d'erreur de base, représentant la fréquence d'erreurs de toutes formes (par omission, confusion, mauvaise réalisation) dans l'accomplissement d'une tâche générique familière à l'opérateur. La valeur généralement admise pour ce paramètre est de 3/100. Conditionnellement à l'erreur initiale, il s'agit alors de considérer l'éventualité de deux autres types d'événement :

- la récupération de l'erreur, c'est-à-dire sa détection et sa correction avant qu'elle n'ait eu de conséquence importante, dont on évalue la probabilité en fonction des moyens de détection requis (alarme, opérations de requalification, tests périodiques, etc.) ;
- le fait que d'autres erreurs aient été commises, dont on évalue la probabilité en se donnant un modèle de dépendance statistique entre actions. Dans le cas du maintien de composants montés en parallèle, par exemple, on considère que la dépendance est totale, c'est-à-dire que conditionnellement à une erreur sur une composante, la probabilité d'une erreur sur les autres est égale à un.

Les interventions en situation accidentelle nécessitent une analyse plus complexe, puisqu'elles consis-

39. L'analyse des systèmes elle-même n'ignore d'ailleurs pas tout à fait ce type de distinction, puisqu'elle tient compte des conditions dans lesquelles les équipements sont sollicités. La Règle fondamentale de sûreté relative aux études probabilistes de sûreté requiert ainsi que le taux de défaillance des matériels soit révisé (jusqu'à être éventuellement considéré égal à 1) lorsque ceux-ci sont sollicités au-delà de leurs conditions de qualification (ASN, 2002).

tent en une succession de phases (détection de l'anomalie, diagnostic de son origine, choix de la consigne de conduite, conduite proprement dite, récupération de situation dégradées, gestion de crise) faisant interagir des opérateurs différents (équipe de conduite, ingénieur de sûreté et de radioprotection, équipe de crise). L'enchaînement de développements défavorables lors de l'accident de TMI a illustré la complexité de la conduite accidentelle, qui nécessite de choisir, parmi un ensemble très étendu de combinaisons d'événements possibles, celle qui s'est effectivement déroulée, dans certains cas à partir d'informations partielles et dans des délais très limités. Afin de réduire cette difficulté, la conduite accidentelle a été par la suite analysée et planifiée en fonction non plus des événements singuliers s'étant produits dans l'installation, mais de l'état général de celle-ci, tel que défini par certains paramètres physiques caractéristiques. Même si ces états devaient être suffisamment détaillés pour déterminer de façon univoque le risque accidentel et la marche à suivre, leur nombre restait très inférieur à celui des combinaisons d'événements à prendre en compte. Ainsi l'*approche par états** a-t-elle été développée et mise en œuvre tout au long des années 1980-90 et constitue désormais le mode de conduite accidentelle de référence en France.

La méthode générale d'évaluation des probabilités reste, cependant, la même que pour la conduite en conditions normales. Les interventions des équipes font l'objet de calculs d'échec spécifiques dans la phase de détection/diagnostic/choix de conduite et dans celle de conduite, où les probabilités d'erreur sont pondérées par un facteur de contexte traduisant l'influence des conditions de stress et de danger. Les échecs de l'ingénieur de sûreté et de radioprotection (ISR) et de l'équipe de crise sont liées à leur absence (absence de la salle de contrôle dans le cas de l'ISR, délai de quatre heures pour être opérationnel dans le cas de l'équipe de crise) ou à la mauvaise application de leurs instructions. L'un et l'autre sont donc réputés ne pas se tromper dans leurs décisions.

Les événements liés au facteur humain peuvent être intégrés aux EPS de différentes façons. S'ils constituent des initiateurs, ils apparaissent en tant que tels au début de séquences accidentelles. S'ils se produisent dans une situation normale de fonctionnement et rendent un système inopérant, il est d'usage de les prendre en compte au sein de l'arbre de défaillance de celui-ci. Les erreurs en situation accidentelle, enfin, viennent modifier une séquence en cours et peuvent donc être soit intégrées à l'analyse des systèmes, soit à celle de la séquence. Le premier choix a prévalu dans l'étude WASH-1400, le second dans l'EPS française de niveau 1 pour les REP 900 MWe.

Défaillances de cause commune

Si l'on pouvait assurer une indépendance effective entre dispositifs de sûreté, l'application répétée du principe de la redondance permettrait de réduire le risque accidentel autant qu'on le souhaiterait. Sous l'hypothèse d'indépendance, en effet, la probabilité d'une défaillance générale baisse de façon exponentielle à mesure que le nombre de niveaux de redondance augmente. Malheureusement, le simple fait qu'ils soient intégrés au même système suffit à engendrer de nombreuses sources de

dépendance entre les dispositifs : la localisation géographique, qui peut les soumettre aux mêmes conditions environnementales (température, humidité, radiations) et les exposer aux mêmes aléas (tels que les séismes) ; le contrôle et la maintenance par les mêmes opérateurs ; l'âge et une susceptibilité commune à l'usure et l'encrassement, etc. De ce fait, l'efficacité marginale des systèmes redondants a tendance à décroître rapidement et il est considéré que leur simple multiplication ne constitue pas une stratégie satisfaisante de sûreté. Les statistiques de fiabilité indiquent ainsi que si le taux de refus de fonctionnement est de 10^{-2} pour un certain équipement, il atteint 3.10^{-4} (et non 10^{-4}) pour un ensemble comprenant deux files identiques de l'équipement et 10^{-4} (et non 10^{-6}) pour un ensemble en comprenant trois (Libmann, 1996, p.63).

Bien que la possibilité de pannes corrélées d'équipements redondants ait été analysée aux débuts de la réflexion méthodologique sur la sûreté nucléaire (Siddal, 1957), il a fallu attendre les années 1980 pour qu'elle conduise à une définition claire de la notion de défaillance de cause commune. Les raisons de cette naissance difficile ne sont pas à chercher dans la notion de causalité, que l'analyse des systèmes comprend dans un sens contrefactuel et dont elle fait un usage courant. Dans la construction de l'arbre de défaillance de la figure 3.4.2, à titre d'exemple, l'événement ε'_5 est considéré comme cause de ε'_4 si l'on estime que ε'_4 n'aurait pas lieu si ε'_5 ne se produisait pas. Si le concept de cause commune a mis un certain temps à émerger, c'est parce qu'il a dû être distingué pas à pas des autres formes de dépendance entre composantes d'un système. L'analyse des systèmes consiste en effet à identifier et représenter un ensemble de relations de dépendance entre l'état d'un système, ceux de ses composantes et les interventions humaines en son sein. De ces dépendances découlent certains schémas de défaillance, tels que la défaillance jointe ou les défaillances en cascade de plusieurs composantes. L'analyse du système rend compte de ces schémas, qu'il s'agit donc de distinguer de défaillances relevant d'une cause commune extérieure à la représentation. Dans la figure 3.4.2, si ε_1 admettait une cause commune avec ε_6 , alors il se pourrait également que l'on constate une dépendance entre ε_1 et ε'_5 , ε'_4 , ε'_3 , ε'_2 et ε'_1 , mais ces relations découleraient toutes de la première étant donnée la structure de l'arbre. Ceci conduit à définir une défaillance de cause commune comme une dégradation, une perte de fonctionnalité ou une panne d'au moins deux composantes d'un système due à l'action d'une cause *unique, directe et extérieure au système*⁴⁰.

L'hypothèse d'absence de cause commune, que nous aborderons au cinquième chapitre sous le nom de condition de Markov, détermine la validité des arbres de défaillance et des estimations probabilistes qui en sont dérivées. Dans le cas extrême (mais pas du tout exceptionnel) où il existerait une cause commune à tous les événements d'une coupe minimale, sa réalisation suffirait à provoquer l'événement de tête. La probabilité de défaillance, qui serait donc supérieure ou égale à

40. Les définitions opérationnelles comprennent en outre des considérations de temporalité et de limites physiques qui permettent de caractériser la relation causale. Pour une définition de référence et une revue des travaux qui y ont mené, on pourra se reporter à (NRC, 1988).

celle de l'occurrence de la cause, pourrait alors être largement sous-estimée par la quantification d'un arbre incomplet. Outre l'attention que portent toutes les manuels d'analyse des systèmes à l'identification *a priori* des causes communes, une stratégie complémentaire s'est progressivement mise en place, qui consiste à déceler *a posteriori* la présence de causes communes latentes dans les défaillances observées et à estimer leurs effets sans nécessairement chercher à les identifier⁴¹. Un outillage empirique important a donc été constitué pour cela, qui comprend des bases de données internationales, différentes méthodes statistiques d'estimation paramétrique, ainsi que des logiciels de simulation. L'usage de ces outils est cependant resté circonscrit aux cas où des données importantes sont disponibles, c'est-à-dire aux défaillances les plus courantes.

Les limitations des méthodes statistiques employées expliquent d'ailleurs que l'identification empirique des causes communes ne soit à ce jour appliquée qu'à l'analyse des systèmes, alors qu'elle est tout aussi importante pour d'autres méthodes d'analyse de la sûreté. Les arbres d'événement, en particulier, reposent sur une multitude d'hypothèses d'indépendance mutuelle, d'un côté entre les initiateurs de différentes séquences accidentelles (sauf représentation explicite de la dépendance) et de l'autre entre l'initiateur et le succès ou l'échec des différentes missions impliquées. Comme nous l'avons vu, l'approche déterministe elle-même est exposée à l'existence de défaillances de cause commune puisqu'elle suppose l'indépendance des différents niveaux de la défense en profondeur. Dans ces cas, la prise en compte des causes communes continue de relever du jugement des analystes et d'être traitée de façon *ad hoc*. Dans le cas des événements les plus rares, elle pose des difficultés importantes et constitue l'un des principaux points de faiblesse des approches actuelles de la sûreté. Nous en offrirons quelques exemples dans le sixième chapitre.

Il se pourrait que des méthodes bayésiennes, qui se fondent sur des distributions de probabilité fixées *a priori* et révisées en fonction des observations, permettent l'évolution vers une démarche plus systématique dans ce domaine. Des travaux récents ont par exemple mis en avant l'utilité des copules pour représenter de façon économe une distribution multivariée représentant les dépendances entre variables d'un système complexe et pour la mettre en jour en fonction des observations (Bedford et Cooke, 2001 ; Kurowicka et Cooke, 2006). Mais une démarche générale reste, pour l'heure, à formuler et se heurte aux difficultés pratiques du bayésianisme dans un domaine tel que la sûreté nucléaire, que nous évoquerons dans la section suivante.

3.4.2.2 L'incertitude dans l'approche probabiliste : débats théoriques

Dans le champ de la sûreté nucléaire comme dans tous ceux où elle a été entreprise, l'application de la théorie des probabilités s'est trouvée confrontée à un choix cornélien entre les interprétations fréquentiste et personnaliste des probabilités et, au moment du passage à l'étape empirique, entre

41. Plus précisément, à les identifier statistiquement sans nécessairement chercher à les associer à un phénomène précis.

les méthodes statistiques classique et bayésienne. Avant d'analyser plus en détail le traitement de l'incertitude dans l'approche probabiliste, il est utile de rapporter quelques débats théoriques suscités par ce choix, qui résument la façon dont les questions épistémologiques soulevées par la représentation de l'incertitude ont été abordées.

Probabilités objectives et subjectives

La première difficulté a naturellement porté sur les conditions dans lesquelles des probabilités « subjectives », qui ne sont pas données par la fréquence observée d'un événement mais reflètent un degré de croyance dans ses chances de réalisation, pouvaient être utilisées dans les analyses de sûreté. Le principe même d'un tel usage était envisagé avec réticence par un grand nombre de spécialistes de la sûreté nucléaire imprégnés de la culture traditionnelle de l'ingénieur. Il est vrai, en outre, que l'ampleur des enjeux laissait peu de place à l'expression libre de croyances *a priori*, et qu'en général, le cas nucléaire se prêtait mal à l'argument habituel des bayésiens, selon lequel le poids des croyances *a priori* finit par s'effacer devant celui des observations, intégrées grâce à la règle de Bayes : l'information nouvelle risquait de ne pas arriver, ou d'arriver seulement à la faveur d'une catastrophe. Cependant, en dehors de toute considération théorique, l'absence de données fréquentielles rendait l'usage de probabilités subjectives incontournable dans un grand nombre de cas, au moins dans les premières évaluations probabilistes. Même lorsque de telles données existaient, il arrivait souvent qu'elles ne soient pas concordantes et laissent une marge d'interprétation importante à l'analyste. Enfin le recours aux croyances subjectives était chose courante dans l'approche déterministe de la sûreté sous les noms de jugement d'expert ou de pratique d'ingénierie⁴². La démarche probabiliste offrait la possibilité d'explicitier ces choix et de les contrôler.

Il fallait, au total, s'accommoder d'une version tempérée du bayésianisme, où l'usage de probabilités subjectives serait autant que possible encadré et fondé sur des éléments d'information objectifs, en attendant que de plus grandes quantités de données opérationnelles et expérimentales s'accumulent⁴³.

Sur le plan théorique, un effort important a été fait pour clarifier les fondements de l'approche probabiliste lors de la deuxième EPS complète réalisée aux États-Unis, celle de la centrale de Zion (CEC, 1981). Ce travail, qui a largement influencé le guide de procédure publié deux ans plus tard par la NRC (NRC, 1983), donnait une orientation résolument bayésienne à l'approche probabiliste de la sûreté. Dans l'un des articles scientifiques qui en ont résumé la philosophie, on pouvait ainsi lire :

"Probability as we shall use it is a numerical measure of a state of knowledge, a

42. Voir la sous-section 4.1.1, notamment à propos du concept d'Accident maximum crédible.

43. La rareté de certains événements et l'évolution des matériaux et de l'architecture des systèmes ont cependant contribué à limiter le domaine dans lequel l'hypothèse de stabilité des fréquences observées à long terme pouvait être acceptée avec un degré suffisant de confiance, de sorte que les EPS récentes continuent de faire une large part aux probabilités subjectives.

degree of belief, a state of confidence" (Kaplan et Garrick, 1981, p.17).

Dans le même temps, cependant, les auteurs de l'article cherchaient à opérer une synthèse qui intégrât la notion de fréquence. Reprenant un argument de De Finetti (1937), ils écrivaient la probabilité de survenue d'un événement ε comme l'espérance de sa fréquence de long terme inconnue :

$$P(\varepsilon) = \int_0^1 P(\varepsilon | f) \omega(f) df = \int_0^1 f \omega(f) df = E(f) \quad (3.4.1)$$

où la survenue de ε est représentée par une variable aléatoire suivant une loi de Bernoulli $B(f)$, et f est une variable aléatoire de densité ω sur $[0, 1]$. Les auteurs faisaient valoir que f est ici une grandeur objective, alors que $P(\varepsilon)$ est une probabilité subjective reposant sur la croyance ω relative à la distribution de f .

Pourtant, comme l'a relevé Dawid (2004, p.48), l'argument ne permet guère de réduire la fracture épistémologique entre bayésiens et fréquentistes. La fréquence est en effet ici une grandeur abstraite⁴⁴ sur laquelle on porte un jugement subjectif, et non un nombre prédéfini associé à l'événement comme l'estiment ces derniers (Abramson, 1981). La représentation de la fréquence comme variable aléatoire est d'ailleurs caractéristique de l'approche bayésienne, et n'a guère de sens dans une perspective fréquentiste, surtout si elle n'est pas liée à un échantillon d'observations. Evoquant une interprétation fréquentiste de son résultat, selon laquelle une évaluation subjective de la probabilité d'un événement E_{n+1} pourrait être corrigée une fois acquise l'information « objective » sur les précédents $E_1, \dots, E_n$, de Finetti lui-même précisait :

« Une telle interprétation nous semble inacceptable, non seulement parce qu'elle présuppose l'existence objective de probabilités inconnues, mais aussi parce qu'elle ne peut même pas être formulée correctement : en effet, la probabilité de E_{n+1} évaluée après la connaissance d'un certain résultat A des n épreuves antérieures, n'est plus $P(E_{n+1})$ mais $P(E_{n+1} | A)$ » de Finetti (1937, p.26).

Tout en adoptant clairement le point de vue de de Finetti, Kaplan et Garrick (1981) ont cependant été plus équivoques en intitulant leur argument « cadre de la probabilité de fréquence » (*probability of frequency framework*). Cette appellation a été reprise par un courant de la littérature sur la sûreté et le risque, qui semblait voir dans l'équation (3.4.1) le fondement d'une approche théorique permettant de concilier le fréquentisme et le bayésianisme⁴⁵ malgré la mise en garde de Howard :

"there is no reason to introduce frequency concepts when discussing either the probability of an event or the situation where we are uncertain about the underlying process"
Howard (1988, p.97).

44. Précisément, la valeur vers laquelle convergerait la fréquence de ε observée dans une séquence de tirages *interchangeables* de $B(f)$ à mesure que le nombre de ces tirages tendrait vers l'infini.

45. Aven (2003) est représentatif de ce courant.

L'ambiguïté a persisté jusqu'à ce jour, comme en témoigne un exemple récent tiré de cette littérature :

"In this approach there are two levels of probability introduced : (1) the relative frequency interpreted probabilities reflecting variation within populations (also referred to as stochastic or aleatory uncertainty) and (2) the subjective probabilities reflecting the analyst's uncertainty what the correct relative frequency probabilities are" (Aven et Renn, 2009, p.592).

La force de la contradiction entre ces termes et ceux de De Finetti est remarquable, et il est probable qu'une part de méprise a contribué au succès du « cadre de la probabilité de la fréquence » dans le domaine de la sûreté nucléaire.

S'il a donc créé une certaine confusion sur le plan épistémologique, ce cadre a cependant un avantage certain : en exprimant f sous forme d'une variable aléatoire, on y décompose l'incertitude relative à l'événement ε en une part relevant de la loi de probabilité qu'on lui associe (la densité de probabilités $\omega(f)$ dans (3.4.1)) et une autre liée à la vraisemblance de sa réalisation pour une loi donnée (la probabilité de ε conditionnellement à f).

Considérons le cas plus général d'un événement défini à partir des réalisations d'une variable T , pour lesquelles on dispose d'un modèle M dont $\theta \in \Theta$ est le vecteur des paramètres. On peut alors écrire (Apostolakis, 1990) :

$$P(T > t \mid M) = \int_{\Theta} P(T > t \mid \theta, M) \omega(\theta \mid M) d\theta = E_{\theta \mid M} P(T > t \mid \theta, M)$$

où la densité ω résume l'ensemble de l'information disponible concernant les paramètres du modèle (par exemple, un échantillon de valeurs relevées sur différents sites pour chaque paramètre). Si l'on dispose en outre d'un nombre fini de modèles $M_i, i = 1 \dots n$ ayant chacun une probabilité p_i d'être « la vraie » représentation du monde⁴⁶, on a :

$$P(T > t) = \sum_{i=1}^n \left(\int_{\Theta} P(T > t \mid \theta, M_i) \omega(\theta \mid M_i) d\theta \right) p_i = E_M E_{\theta \mid M} P(T > t \mid \theta, M) \quad (3.4.2)$$

Dans cette équation, nous avons décomposé l'incertitude relative à l'événement $(T > t)$ en trois parts : celle de sa survenue telle qu'elle peut être envisagée à l'aide d'un modèle du monde et d'un ensemble de paramètres donnés ; celle de l'ensemble de paramètres retenus pour un modèle donné ; et enfin celle du modèle lui-même.

46. Il s'agit bien sûr d'un monde restreint aux grandeurs pertinentes pour notre problème, dont nous donnerons une définition opérationnelle au chapitre 5.

Il faut insister sur le fait que, conformément à la tradition bayésienne, paramètres et modèles sont ici considérés comme des variables aléatoires dès lors qu'ils sont incertains. La distinction entre paramètre et modèle perd alors de sa pertinence, puisque l'on peut parfaitement imaginer passer d'un modèle à un autre par simple variation d'un paramètre (tel que le coefficient de corrélation entre deux variables), voire représenter au sein d'un même (méta-)modèle des représentations alternatives d'un phénomène, pondérées par des coefficients (et transformer ainsi l'incertitude portant sur le modèle en incertitude paramétrique). Cette question n'a d'ailleurs aucune importance réelle dans la démarche bayésienne, puisque, comme le montrent les équations (3.4.1) et (3.4.2), toutes les formes d'incertitude s'y résolvent dans la mesure de probabilité. La prise en compte de l'incertitude portant sur la détermination d'une probabilité, et plus encore la distinction entre les sources de cette incertitude, n'ont pas de justification dans la démarche bayésienne, dont l'un des principes fondamentaux est que l'incertitude est *entièrement* mesurée par la probabilité⁴⁷.

En distinguant différentes composantes de l'incertitude, le « cadre de la probabilité de fréquence » ne cherche pas à s'écarter de cette logique, comme il le ferait par exemple en s'interrogeant sur les motifs qui amèneraient un gestionnaire de la sûreté à adopter des *attitudes* différentes à leur égard⁴⁸. Son projet est, au contraire, de rappeler que la probabilité d'un événement n'est pas une évaluation obtenue à partir d'un modèle et d'un jeu de valeurs des paramètres concernés, mais une mesure de la croyance dans les chances de réalisation de l'événement par-delà ses schémas de représentation ; et de proposer une méthode pour la prise en compte des incertitudes de cette évaluation dans la mesure de probabilité.

Incertitude aléatoire et incertitude épistémique

Pour désigner les deux formes d'incertitude qui concernent, pour l'une, une réalisation particulière d'une variable et, pour l'autre, la représentation que l'on s'en donne, la citation d'Aven et Renn (2009) utilise les qualificatifs d'aléatoire et d'épistémique. L'incertitude aléatoire, également appelée variabilité, incertitude objective ou incertitude de type A, est généralement considérée comme une propriété du phénomène représenté, à l'image de l'issue d'une expérience menée de nombreuses fois dans le passé, le n -ième lancer d'une pièce ou le choix au hasard d'un individu au sein d'une population connue. Elle pourrait ainsi être représentée par les réalisations d'une variable aléatoire dont la distribution de probabilités serait connue. L'incertitude épistémique, ou encore ignorance, incertitude subjective ou incertitude de type B, serait une propriété de l'observateur, mesurant son degré de connaissance du phénomène représenté. Elle porterait donc sur la caractérisation d'un paramètre ou d'un modèle inconnus, mais que l'on pourrait supposer déterminés. La première serait ainsi celle prise en compte par les analyses probabilistes de risque, tandis que la seconde porterait sur

47. Pour un aperçu synthétique de la position bayésienne classique, on pourra se reporter à Lindley (2000).

48. Ce qui a constitué, comme nous le verrons aux chapitres 7 et 8, l'objet d'un courant de recherche important en théorie de la décision au cours des trois dernières décennies.

ces analyses elles-mêmes, c'est-à-dire sur les modèles et les valeurs de paramètres qu'elles utilisent.

Injectée dans le « cadre de la probabilité de fréquence », la distinction conduit à une interprétation courante de l'équation (3.4.1), selon laquelle le premier niveau décrirait une incertitude aléatoire à l'aide de probabilités « objectives », alors que le second refléterait une incertitude épistémique caractérisée par des probabilités subjectives. Naturellement, les deux distributions pourraient être dégénérées. Dans le premier cas, cela refléterait une situation d'ignorance relative à la valeur d'une constante, et dans le second une parfaite connaissance d'un processus stochastique.

Apostolakis et Kaplan (1981) semblent avoir été les premiers à proposer une telle séparation, pour des raisons que nous aborderons dans la suite. Leur proposition a été reprise dans le guide de procédures de la NRC (1983) puis, dans une certaine mesure, mise en œuvre dans l'analyse probabiliste de cinq centrales américaines élaborée dans la deuxième moitié des années 1980 (NRC, 1990). Elle fait désormais partie de l'état de l'art en matière d'évaluation de l'incertitude dans les EPS (Helton, 1994 ; Hofer, 1996 ; Paté-Cornell, 1996). Il s'est agi, en pratique, de répartir dans l'une ou l'autre catégorie toute grandeur incertaine au départ d'une analyse probabiliste. Vesely et Rasmuson (1984) ont suggéré pour cela de définir l'incertitude aléatoire comme celle qui est irréductible par l'apport d'informations nouvelles relatives au phénomène représenté, et l'incertitude épistémique comme celle qui pouvait être diminuée. Hoffman et Hammonds (1994) ont conjecturé qu'une incertitude était épistémique si l'objet de l'analyse était, en dernière instance, une grandeur fixe ; et qu'elle était aléatoire si cet objet était une distribution de probabilités (telle que la répartition d'un certain caractère dans une population).

Le jugement que l'on peut porter sur ces critères pratiques et sur la distinction elle-même change selon que celle-ci est comprise comme une hypothèse heuristique, dont la validité ne serait au mieux que locale, ou comme un principe universel.

Dans le premier cas, il s'agit pour l'analyste, dans le cadre limité d'une expérience, de considérer ses représentations de l'incertitude comme fixées pour certains des phénomènes étudiés et révisables pour d'autres (Winkler, 1996). En renversant l'observation de Vesely et Rasmuson, on pourrait alors dire que l'incertitude aléatoire est *définie par l'analyste* lorsqu'il choisit de ne pas réviser ses croyances à la lumière d'une nouvelle observation. L'intérêt principal d'une telle démarche est de lui permettre de tenir les différentes réalisations d'une variable pour indépendantes et identiquement distribuées. Imaginons, à titre d'exemple, que l'analyste s'intéresse au taux de panne d'un certain équipement, dont différentes unités sont mises en œuvre au sein d'une centrale, et que l'information dont il dispose consiste en un ensemble d'observations portant sur des équipements similaires dans d'autres centrales. Il peut estimer que les taux de panne observés ailleurs lui fournissent une distribution satisfaisante⁴⁹ du taux de panne aléatoire de chaque unité dans la centrale

49. Au sens où il peut l'adopter comme donnée.

considérée. Son incertitude pourra alors être qualifiée d'aléatoire. Il peut, au contraire, considérer le taux de panne de chaque unité comme une grandeur fixe et se donner comme distribution *a priori* de cette grandeur la distribution empirique établie dans d'autres installations. Cette distribution, qui mesure une incertitude épistémique, sera révisée suite à chaque observation. Le choix de l'une de ces deux options est, comme on le voit, affaire du jugement. Une différence importante entre elles, cependant, est que l'on peut postuler l'indépendance des taux de panne des différents équipements de la centrale sous la première option, mais pas sous la seconde (Apostolakis et Kaplan, 1981). En présence d'incertitude épistémique, il est donc nécessaire d'estimer la dépendance entre différentes réalisations d'une même variable, notamment lorsque les distributions de probabilité sont des croyances d'experts (van Oorschot et al., 2003).

Dans le second cas, on entend distinguer des événements intrinsèquement aléatoires d'autres dont le caractère incertain ne serait que le reflet de notre ignorance. Les termes « aléatoire », « incertitude objective » et « variabilité » invitent à une telle interprétation, en semblant indiquer la présence de phénomènes purement stochastiques qu'aucune connaissance nouvelle ne peut permettre de mieux cerner. Une telle hypothèse, à moins qu'elle repose sur une solide justification scientifique (mais elle ne le fait pas dans l'essentiel des cas considérés), semble très critiquable sur le plan méthodologique⁵⁰.

Considérons ainsi le postulat selon lequel l'incertitude aléatoire concernant un phénomène ne serait pas réductible avec l'apport d'informations nouvelles à son sujet.

L'information nouvelle est souvent comprise comme une observation supplémentaire (disons la n ième) de la variable en question. Poser, dans ce cas, qu'elle ne changerait pas notre incertitude, c'est-à-dire la probabilité que nous assignons à une $n+1$ ième réalisation, nous ramène à la traditionnelle ligne de partage entre fréquentistes et bayésiens : nous devons pour cela considérer entièrement connue la loi de la variable, ce qui permet alors de stipuler l'indépendance entre ses réalisations successives. Une telle hypothèse est locale par nature, puisqu'elle postule l'identité entre une distribution empirique et la loi sous-jacente. Mais de surcroît, la forme donnée à l'information nouvelle est indûment restrictive. L'information peut en effet porter sur des facteurs causaux influençant la variable et nous amener à modifier notre représentation de celle-ci, par exemple sous la forme d'une meilleure spécification. Le taux de panne d'un équipement peut s'avérer dépendre d'un aspect de sa fabrication que nous n'avions pas pris en compte, de sorte que l'échantillon d'observations dont nous disposons serait réduit à un sous-ensemble de l'échantillon initial, au sein duquel la variance observée de la variable serait moindre. En toute généralité, comme le montre l'équation (3.4.2), la distribution de probabilités *a posteriori* de la variable T diffère de sa distribution *a priori* dès lors que l'une de ses trois composantes change : la distribution de T à vecteur de paramètres θ et modèle

⁵⁰. Voir par exemple la charge de René Thom contre ce qu'il appelle la « fascination pour l'aléatoire » (Thom, 1980).

M donné, celle de θ à M donné et enfin celle de M . Dans une perspective ainsi élargie, l'hypothèse de fixité de la loi devient encore plus draconienne, puisqu'elle suppose celle de la représentation à l'intérieur de laquelle la variable est considérée⁵¹.

Il semble donc sage de s'en tenir à un usage local de la distinction entre incertitude aléatoire et incertitude épistémique, qui reflète la différence entre deux attitudes possibles d'un décideur faisant face à une situation incertaine : se satisfaire des connaissances existantes, en les estimant suffisantes pour décrire le comportement aléatoire du phénomène en jeu et rendre plausible l'hypothèse d'indépendance de ses réalisations ; ou se préparer à une meilleure compréhension des phénomènes en admettant la possibilité d'une corrélation entre les réalisations futures (ce qui, dans le domaine de la sûreté, constitue généralement une hypothèse pénalisante). Il est alors entendu que le schéma aléatoire / épistémique repose sur une approximation et que la ligne de démarcation qu'il établit entre les phénomènes étudiés peut se déplacer en fonction des circonstances. Dans l'exemple précédent, on peut imaginer que le choix de l'analyste sera influencé par le degré de similitude entre les conditions de la centrale étudiée et celles des centrales de l'échantillon ou par sa capacité à discerner un trait spécifique aux unités de la centrale étudiée. Le processus d'expérimentation et d'acquisition de connaissances peut ainsi engendrer un mouvement cyclique dans lequel une forme d'incertitude est transformée en l'autre (Faber, 2005 ; Field et Grigoriu, 2007 ; Der Kiureghian et Dietlevsen, 2009).

La prise en compte de la différence entre incertitude aléatoire et incertitude épistémique a conduit à une nouvelle vague de remise en cause de l'usage des probabilités subjectives. Ferson et Ginzburg (1996), par exemple, estiment que les probabilités, si elles conviennent à la description de la variabilité, ne peuvent être utilisées pour représenter l'ignorance car elles nécessitent de spécifier l'information plus qu'elle ne l'est objectivement. Ces auteurs plaident pour la substitution des probabilités subjectives par des intervalles de probabilités, plus conformes à la structure de l'information dans des situations d'ignorance. À leur suite, une littérature abondante a exploré des traitements alternatifs de l'incertitude épistémique fondés, par exemple, sur la théorie de l'évidence, la théorie des possibilités et celle des ensembles flous⁵². Des exercices comparatifs ont montré que les différences de méthodologie se traduisaient par des écarts importants dans les résultats, sans parvenir à proposer une synthèse (Ferson et al., 2004).

Par-delà les considérations théoriques, cependant, ces travaux expriment une réserve largement partagée à l'égard de l'approche bayésienne lorsqu'il s'agit d'assigner des probabilités à des représentations différentes du monde. Plutôt que l'incertitude épistémique en général, cette réserve concerne l'incertitude relative au vecteur de paramètres θ lorsqu'on ne dispose pas de données quantitatives

51. Nous développerons ce raisonnement sur le plan formel dans le chapitre 5.

52. Pour une vue d'ensemble, voir Helton, Johnson et Oberkampf (2004) et l'ensemble du numéro de l'hiver 2004 de la revue *Reliability Engineering and System Safety* consacré à cette question.

à son propos et, plus sérieusement encore, l'incertitude sur le modèle M . Le guide de procédures de la NRC considère ainsi que parmi les différents domaines d'incertitude dans la modélisation d'une centrale ou de ses systèmes, la plupart ne se soumet pas à une analyse quantifiée précise⁵³.

3.4.2.3 L'incertitude dans l'approche probabiliste : traitement pratique

Sources d'incertitude dans les EPS

Le guide de procédures de la NRC recommande, en pratique, la prise en compte de trois catégories d'incertitude épistémique : celles relatives aux données, aux hypothèses de modélisation et à la complétude de la représentation (NRC, 1983). Cette classification a été par la suite reprise par l'AIEA, dans des termes légèrement différents, dans la formulation de procédures de référence pour la conduite d'EPS : l'estimation des paramètres, la modélisation des scénarios et leur identification (AIEA, 1992, 1995 et 1996).

La distinction classique entre l'incertitude paramétrique et l'incertitude de modélisation s'est donc enrichie, dans la méthodologie des EPS, d'une troisième catégorie relative à la complétude de l'analyse. Il s'agit en fait d'une tentative initiale de faire une place particulière, dans l'incertitude de modélisation, au problème d'omission de certaines relations causales. Dans le cas de systèmes aussi complexes et de conception aussi récente que les centrales nucléaires, l'hypothèse d'identification exhaustive des modes de défaillance et des scénarios de développement des accidents est en effet particulièrement héroïque. On considère ainsi que l'ignorance d'un certain mode de défaillance d'un système correspondant à une combinaison particulière d'indisponibilités de ses composantes est une forme d'incertitude d'identification. On considère, par contraste, comme incertitude de modélisation la possibilité qu'une certaine combinaison identifiée comme coupe ne conduise pas toujours à une défaillance effective, par exemple parce que l'interaction entre composantes est plus complexe que la logique binaire de l'arbre de défaillance ne permet d'en rendre compte. En pratique, toutefois, les deux formes font l'objet du même type de traitement et sont rarement distinguées.

Le tableau 3.4 présente quelques exemples de facteurs d'incertitude dans les EPS de niveaux I, II et III.

Propagation de l'incertitude dans les EPS

Le comité scientifique chargé par la NRC en 1976 d'évaluer l'étude WASH-1400 estima que la méthodologie qui y était mise en œuvre constituait un nouvel état de l'art en matière d'identification des scénarios d'accident et de quantification du risque, mais porta une appréciation beaucoup plus sévère sur l'évaluation de l'incertitude affectant les résultats (Lewis et al., 1978). Plus récemment, une équipe d'experts mandatés par l'Agence Internationale de l'Énergie Atomique et l'Agence pour

53. "There are several areas within the plant and system modeling activity that give rise to uncertainty, but most are not amenable to accurate quantitative estimation or calculation" (NRC, 1983, p.3-112).

TABLE 3.4 – Exemples de facteurs d'incertitude dans les EPS

Élément affecté	Paramètres	Modélisation des scénarios	Identification des scénarios
Séquences d'accident (EPSI)	Fréquence d'une rupture de vanne (initiateur interne) Fréquence d'un séisme (initiateur externe) Durée d'une intervention	Défaillances de système Séquences accidentelles Propagation d'incendie	Événements initiateurs États critiques Dépendances entre initiateurs, erreurs humaines et défaillances de système
Conséquences internes (EPSII)	Conditions atmosphériques (température, pression) et physiques (densité, conductivité) initiales dans l'enceinte de confinement	Réponse du confinement Défaillance du confinement Transport de radionucléides	Processus physiques déclenchés par la fusion du cœur Déterminants du transport de nucléides
Conséquences externes (EPSIII)	Inventaire du cœur au moment de l'accident Conditions météorologiques Densité de population dans la région affectée	Dispersion atmosphérique Contamination des sols Impacts sanitaires Conséquences économiques	Types d'impacts sanitaires Types de conséquences économiques

l'Énergie Nucléaire pour examiner l'EPS du site d'enfouissement de déchets nucléaires à Yucca Mountain estima à la fois que l'incertitude y avait été traitée de façon incomplète et que les intervalles de variation obtenus pour les résultats finaux étaient excessifs (AEN et AIEA, 2002, pp. 52-3).

L'insuffisante prise en compte de l'incertitude et son ampleur ont ainsi constitué deux motifs constants de critique à l'encontre des EPS et de mise en garde contre une interprétation littérale de leurs résultats. La seconde forme de critique ne peut pas être réellement orientée vers la méthodologie des EPS. Le sens de la démarche probabiliste est précisément de quantifier l'incertitude ; l'indication d'une incertitude forte constitue une information précieuse pour la décision, qu'il s'agisse de collecter des données et d'entreprendre des recherches supplémentaires pour la réduire ou de s'assurer que les mesures de sûreté permettent de faire face dans la mesure souhaitable aux différentes issues possibles. On peut reprocher à une évaluation de ne pas intégrer l'ensemble des connaissances disponibles, mais pas de refléter les limites de ces connaissances. Nous nous concentrons donc sur la première forme de critique.

La méthodologie de référence en matière d'incertitude, telle que les développements rapportés ci-dessus l'ont façonnée, consiste à identifier de façon aussi complète que possible les sources d'incertitude épistémique dans les trois catégories, puis de mesurer leur impact sur les résultats de l'évaluation, en prenant bien garde à tenir compte de leurs corrélations éventuelles (Devictor et Bolado, 2005).

Pour l'incertitude paramétrique, la méthode la plus fréquente⁵⁴ est d'associer une distribution

54. Nous ne revenons pas ici sur la distinction entre méthodes bayésiennes et méthodes dites classiques, qui peut être répétée à chaque étape de notre description.

de probabilités à la valeur du paramètre, de la propager dans l'EPS par l'intermédiaire de méthodes classiques de simulation (Monte-Carlo) ou de calcul approché (méthode des moments), et enfin de représenter la marge d'incertitude sur le résultat par un indicateur de dispersion (tel que le ratio entre la valeur du 95e centile et la médiane).

Pour les incertitudes de modélisation et d'identification, la méthode de référence est l'analyse de sensibilité, qui consiste à estimer les résultats de l'EPS pour différentes « valeurs » d'entrée (c'est-à-dire sous différents modèles ou hypothèses de modélisation) sans considération de probabilités et à étudier l'effet des variations des valeurs d'entrée sur celles des résultats.

Étant donné le nombre de paramètres et d'hypothèses de modélisation en jeu dans une EPS, il est impossible de suivre une telle démarche de façon systématique. Il est donc nécessaire de traiter l'incertitude de façon différenciée, et la véritable cible des critiques de la première forme contre les EPS est souvent l'absence de méthode et de cohérence dans cette différenciation.

L'incertitude paramétrique a commencé à être prise en compte dès WASH-1400, puis dans les EPS de niveau 1 réalisées aux États-Unis au début des années 1980, selon une approche que l'on peut qualifier de bayésienne (Vesely et Rasmuson, 1984). Il semble que les paramètres retenus pour l'analyse d'incertitude étaient choisis de façon *ad hoc*, sans que ce choix fasse l'objet d'une formalisation quelconque. L'impact mesuré sur les probabilités d'accident était généralement d'autant plus élevé que celles-ci étaient faibles. Les « facteurs d'erreur » liés à l'incertitude sur les initiateurs internes allaient ainsi de quatre pour les accidents dont la probabilité était estimée à 10^{-4} par réacteur et par an à trente pour ceux dont la probabilité était estimée à 10^{-7} par réacteur et par an. Les facteurs d'erreur liés aux incertitudes sur les initiateurs externes (dans la minorité de cas où ceux-ci étaient considérés à la fois dans l'EPS et dans l'analyse d'incertitude) étaient légèrement supérieurs.

Bien que la pratique américaine consistât à intégrer les trois niveaux d'EPS, les incertitudes étaient rarement propagées au-delà du premier, jusqu'aux conséquences internes et externes de l'accident. Vesely et Rasmuson (1984) expliquaient ce fait par la complexité des modèles de propagation et le fait qu'ils étaient encore en pleine évolution. L'EPS de cinq centrales américaines conduite dans la deuxième moitié des années 1980 et connue sous le nom d'étude NUREG-1150 a représenté un important pas en avant dans ce domaine (NRC, 1990).

Deux récents exemples français permettent d'illustrer à la fois les progrès réalisés en matière de prise en compte des incertitudes et l'ampleur des difficultés qui demeurent.

Le premier concerne non pas une EPS, mais une évaluation du risque de leucémie chez les enfants et jeunes adultes vivant dans le canton de Beaumont-Hague lié aux rejets de l'usine de retraitement de La Hague. L'étude, conduite de 1997 à 1999 par un groupe d'experts intitulé Groupe Radioécologie Nord-Cotentin, avait conclu que le nombre de cas de leucémie attribuables aux rejets

entre 1978 et 1996 était proche de zéro (0,0014). Devant certaines réactions suscitées par ce résultat, les pouvoirs publics demandèrent au Groupe de procéder à une analyse de l'incertitude. L'évaluation initiale reposait sur une série de modèles décrivant les composantes du risque : expositions *ex utero* par transfert dans l'environnement (alimentation, baignade, etc.), expositions lors d'incidents ou d'accidents sur le site, expositions *in utero*, relations dose-effet. Afin de limiter l'ampleur de l'analyse d'incertitude, on choisit de la faire porter sur les seuls modèles de transfert environnemental et, parmi les milliers de paramètres que ceux-ci comportaient, sur les seuls 214 plus importants. Une analyse préliminaire de sensibilité permit de sélectionner les paramètres dont les variations individuelles avaient le plus grand impact sur le risque. Des distributions de probabilités furent estimées pour les paramètres, et mille tirages aléatoires réalisés selon la méthode de Monte-Carlo, sous l'hypothèse d'indépendance des paramètres (sauf dans un petit nombre de cas où la dépendance était évidente). Une évaluation alternative fut également menée à partir de mesures de possibilités, qui nécessitent une moindre information sur la distribution des variables. Les intervalles entre les fractiles à 5 % et à 95 % du risque final furent estimés à $[0,001; 0,0024]$ par la première méthode, et à $[0,0004; 0,0045]$ par la seconde (GRNC, 2002). L'évaluation renforça donc la principale conclusion de l'étude initiale, tout en montrant que celle-ci, en se basant sur des estimations ponctuelles des paramètres, était nettement biaisée (puisque la borne inférieure de l'intervalle de confiance était supérieure au risque calculé).

En dépit d'importantes hypothèses simplificatrices, cette analyse d'incertitude mobilisa plusieurs équipes d'experts durant plus d'une année.

Le second exemple provient de l'EPS de niveau 2 réalisée pour les réacteurs de 900 MWe en France par l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN, 2009). L'étude considère les séquences accidentelles à partir des scénarios de fusion établis par l'EPS de niveau 1, et regroupe ces séquences en 44 familles pour lesquelles elle évalue un terme source ainsi que la dose reçue en quinze jours par une personne se trouvant à deux kilomètres de la centrale. Elle va donc au-delà de la limite habituelle des EPS de niveau 2, c'est-à-dire le calcul des termes source, pour évaluer en outre des conséquences radiologiques extérieures sous une hypothèse conventionnelle. L'EPS a fait l'objet d'une évaluation de l'incertitude épistémique portant sur une vingtaine de paramètres du modèle de rejets, propagée selon la méthode de Monte-Carlo. L'incertitude relative aux résultats de l'EPS1, en revanche, n'a pas été propagée. L'effet de l'incertitude sur les résultats finaux de l'EPS 2 a été calculé pour chaque famille de séquences en termes à la fois de probabilité (à conséquence donnée) et de conséquence (à probabilité donnée), ce qui a permis d'en donner une représentation synthétique dans un diagramme (probabilités, conséquences). L'étude a en outre examiné un élément d'incertitude de modélisation : la dilution hétérogène, qui résulte de l'envoi dans le cœur du réacteur d'un bouchon d'eau à teneur insuffisante en bore. Les conséquences d'un

tel événement pour la physique du cœur sont insuffisamment comprises à ce jour, et l'hypothèse d'une rupture de la cuve et d'une perte de confinement n'a pas pu être exclue. Il s'agit d'un cas de ce que nous caractériserons dans la suite comme une incertitude causale, à laquelle il est difficile d'associer une distribution de probabilités. Le traitement retenu dans l'EPS 2 est une étude de sensibilité sous deux hypothèses causales extrêmes : dans un cas, la dilution hétérogène conduit à des rejets massifs et précoces de substances radioactives, dans l'autre elle n'a aucun effet sur le cœur. Les résultats correspondants ont été indiqués par deux points de même abscisse dans le plan (probabilités, conséquences) et la gamme des situations possibles y a été représentée par le segment de droite verticale joignant ces points. Le scénario de dilution hétérogène avec impact sur le cœur constitue de loin la première contribution au risque total.

Bien que de tels exemples existent et qu'un réel effort ait été fait pour répertorier les principales incertitudes de modélisation et d'identification, leur prise en compte effective dans les EPS reste rare. À cela, deux raisons nous semblent pouvoir être avancées : l'une est son coût, puisqu'elle suppose le développement de modèles scientifiques alternatifs ; l'autre est l'absence de réflexion et de cadre sur ce qu'il convient de faire face à une telle incertitude. Dans le cas de la dilution hétérogène, des dispositions ont été prises pour en écarter la possibilité dans l'EPR par conception ; dans les REP, cependant, l'incertitude demeure.

La question des probabilités subjectives

Alors que la remise en cause du recours aux probabilités subjectives dans les études probabilistes n'a, comme nous l'avons suggéré, jamais réellement cessé, les modalités de leur dérivation ont fait l'objet d'avancées méthodologiques importantes.

Jusque dans la deuxième moitié des années 1980, l'usage en la matière était de s'en remettre au jugement de l'analyste, plus ou moins étayé et documenté, selon les cas, par des résultats empiriques et théoriques. De rares EPS, dont celle de la centrale nucléaire de Zion (Commonwealth Edison Company, 1981), considéraient des distributions de probabilités de second ordre consistant pour chaque paramètre en un ensemble de distributions *a priori*, pondéré selon le degré de croyance des analystes. Dans ce cas, les deux niveaux de probabilité étaient révisés en fonction des données disponibles : d'abord pour les croyances de second ordre sur la base de données collectées dans l'ensemble des centrales de même type, puis pour chaque distribution de valeurs du paramètre sur la base de données spécifiques à la centrale.

Le recours aux probabilités subjectives fit l'objet d'une véritable formalisation pour la première fois dans l'étude NUREG-1150 (NRC, 1990, en particulier pp. 2-21 et suivantes). Étant relativement lourde, celle-ci fut cependant réservée aux distributions de probabilités qui paraissaient les plus incertaines et dont l'importance du point de vue du risque général était la plus grande. Les phénomènes concernés furent regroupés en catégories, et un panel d'experts sélectionné pour

chacune d'entre elles. Ces experts furent introduits aux résultats de l'économie comportementale, notamment ceux concernant la surpondération des événements de faible probabilité et l'excès de confiance. Ils disposèrent de temps et de financements pour étudier les questions qui leur étaient soumises. Pour finir, leurs croyances furent élicitées selon les méthodes de l'analyse de la décision, puis agrégées par simple sommation.

La méthodologie de NUREG-1150, qui plaçait l'élicitation des probabilités subjectives à l'intérieur du champ de l'analyse de l'incertitude, a été une référence pour les EPS réalisées depuis.

3.5 Caractérisation des deux approches

Après avoir décrit en détail les deux grandes approches de la sûreté, il nous reste à examiner dans quelle mesure elles peuvent être substituées l'une à l'autre ou, au contraire, se compléter. Nous commençons, pour cela, par comparer les deux approches dans le cas relativement simple de la fiabilité mécanique et des marges de sûreté, avant d'en proposer une caractérisation plus générale et de discuter les conditions de leur complémentarité.

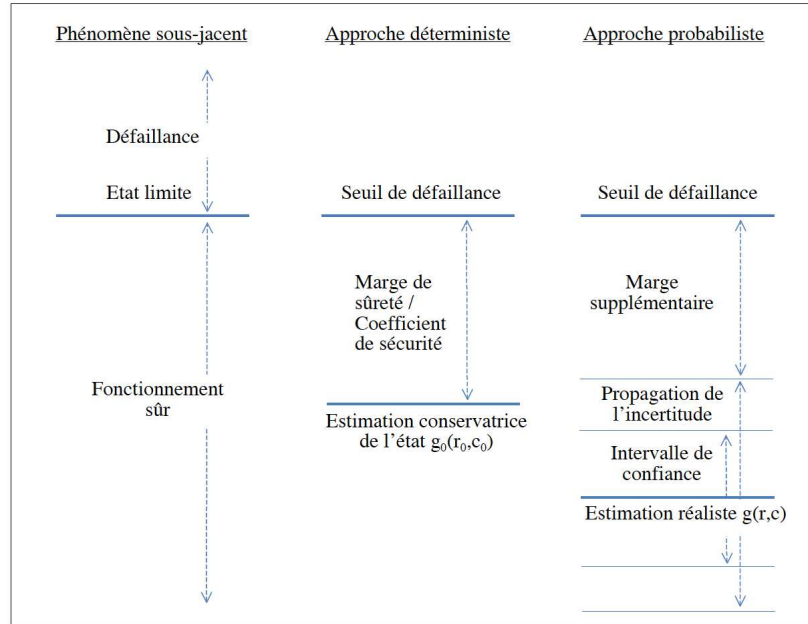
3.5.1 Une illustration de la différence entre les deux approches

L'analyse de la fiabilité mécanique cherche à représenter l'état limite à partir duquel une structure ou un composant peut être considéré comme défaillant (voir la première colonne dans la figure 3.5.1), et à définir le domaine dans lequel il doit être maintenu durant son fonctionnement afin d'éviter sa défaillance. L'état de l'élément est décrit par l'intermédiaire d'une fonction g de n variables $(x_1, \dots, x_n)$ tel que $g(x_1, \dots, x_n)$ est strictement positif, nul ou strictement négatif selon que l'élément est en fonctionnement sûr, à l'état limite ou défaillant. Il s'agit, par exemple, de l'écart entre la résistance R de l'élément et la contrainte maximale (de pression, température, poids, etc.) C qu'il subit durant son utilisation. La notion de marge de sûreté recouvre différentes mesures de l'écart minimal ainsi instauré entre la représentation de l'état effectif de l'élément et celle de son état limite.

Naturellement, des grandeurs telles que la résistance et la contrainte ne sont pas connues de façon exacte. Elles sont estimées à partir d'un ensemble de paramètres (relatifs aux propriétés du matériau, aux dimensions et à la forme de l'élément, etc.). Souvent, ces paramètres ne sont eux-mêmes pas mesurables, mais sont déterminés sur la base d'un échantillon d'observations ou par expertise. De surcroît, en raison des qualités propres du cas à l'étude, on est généralement dans l'incapacité de le situer précisément par rapport à cet échantillon.

Les méthodes déterministes d'évaluation de la fiabilité tiennent compte de cette incertitude en retenant des valeurs raisonnablement conservatrices pour les données d'entrée, et en adoptant

FIGURE 3.5.1 – Incertitude et marges de sûreté



également des hypothèses conservatrices dans le calcul de la contrainte et de la résistance. Elles aboutissent ainsi à une estimation pessimiste de la valeur de $g(R, C)$, soit $g_0(r_0, c_0)$, et elle s'assure que cette valeur est supérieure à une certaine marge strictement positive (ou, plus communément, que r_0/c_0 est au moins égal à un coefficient de sécurité supérieur à 1). Étant fondée sur une estimation volontairement pessimiste des grandeurs en jeu, la marge de sûreté ne donne pas une mesure précise de l'écart entre l'état effectif de l'élément et son état limite, soit $g(R, C) - 0$ (deuxième colonne de la figure).

Les méthodes probabilistes cherchent, en revanche, à représenter explicitement l'incertitude sur les deux grandeurs R et C en les considérant comme des variables aléatoires. Ces variables sont estimées à partir de grandeurs exogènes, dans des relations où l'incertitude est considérée comme aléatoire, et représentée par un terme d'erreur stochastique. En associant des distributions de probabilités à la résistance et à la contrainte, il devient possible de mesurer la marge entre la valeur la plus probable de leur écart $g(r, c)$ et une valeur pessimiste $g(r^-, c^+)$ que l'on retient par mesure de prudence⁵⁵. De surcroît, l'incertitude probabilisable sur la valeur des paramètres, de type épistémique, peut être propagée dans les calculs pour engendrer un intervalle de confiance sur le résultat. La somme des deux marges ainsi introduites dans le calcul constitue une mesure de l'écart entre l'état effectif et l'état limite tel qu'estimé par la méthode probabiliste (troisième colonne de la figure). Elle peut s'interpréter en termes de probabilité de défaillance (Ang et Tang,

55. La contrainte maximale c^+ peut par exemple être la borne supérieure de l'intervalle de confiance à 95% associé à la distribution de C , ou encore la valeur maximale atteinte par C dans un ensemble de scénarios décrits par un arbre d'événements.

1984).

La prudence mise en œuvre dans le cas déterministe ou l'intervalle de confiance associé à une estimation réaliste dans le cas probabiliste constituent-ils des marges de sûreté suffisantes ? Le passage suivant d'un manuel de référence en fiabilité en donne une idée :

"the traditions of craftsmanship were results of slowly gained experiences about in which way good and safe structures should be designed. This accumulation of experiences was the necessary basis for a formalized engineering practice that in steadily increasing extent by application of mathematical models could deviate from traditional structural solutions by cautious use of the insight that the development of the natural sciences brought about. It was clear to the code makers that mathematical models give idealized pictures of the reality and that it is difficult to guarantee that all phenomena of possible essential importance for the reliability are included or realistically represented in the models. To this the history showed too many cases of mistakes having more or less catastrophic consequences. Therefore the code was made such that its value assessments were belonging to a selected class of deterministic and suitably simple models of the behavior of structures, all well known from the engineering education. Within the defined model universe the values of one or more so-called safety factors could then be assessed by consequence computations. [...] These consequence computations served the purpose to show that the use of the given safety factor values in usual engineering computations resulted in the same structural dimensions as obtained from the craftsmanship rules. With the increased degree of problem structuring and understanding of the phenomena brought about by the mathematical model analysis it became possible slowly to change the value assessment in a controlled way such that more daring design of structures became possible" (Ditlevsen et Madsen, 2007, p. 9).

La caractérisation de la défaillance comme résultat d'un excès de contrainte par rapport à la résistance ($g(R, C) = 0$), ces deux notions faisant l'objet de définitions et de mesures précises, correspond ainsi à une représentation simplifiée, c'est-à-dire un modèle, de la réalité. Cette modélisation ne peut prétendre reproduire parfaitement la réalité en raison des limites de notre compréhension du monde⁵⁶. Son écart à la réalité est, par nature, incertain. Il résulte de l'action de facteurs non pris en compte dans la mesure de la contrainte et de la résistance, ou d'une action des facteurs pris en compte qui diffère de ce qui était attendu. La marge de sûreté est censée combler cet écart.

À l'intérieur même de sa branche déterministe, l'analyse de la fiabilité a fait l'objet de raffinements successifs permettant d'intégrer l'apport de différentes disciplines scientifiques par le biais du formalisme mathématique. Chaque nouvelle avancée a été intégrée par étapes, en sélectionnant

⁵⁶. Nous aurons l'occasion d'exposer plus amplement les arguments en faveur de la primauté d'une interprétation épistémique de l'incertitude au chapitre 5.

les modèles les plus appropriés, en leur appliquant des marges de sûreté telles que les principales caractéristiques des ouvrages restent inchangées dans un premier temps, et enfin en réduisant le degré de conservatisme une fois que la compréhension des phénomènes en jeu avait accru la confiance dans les prédictions des nouveaux modèles.

La démarche probabiliste apparaît ainsi comme une poursuite de ce processus d'amélioration de l'inférence par d'autres moyens. En remplaçant les valeurs constantes par une distribution de valeurs possibles, les méthodes probabilistes permettent de gagner en réalisme (estimation de la valeur la plus probable $g(r, c)$ - *best estimate*) et de mesurer en partie l'incertitude affectant le résultat (estimation d'intervalles de confiance $[r^-; r^+]$ et $[c^-; c^+]$, propagation de l'incertitude paramétrique). Elles restent cependant des représentations formelles, et supposent des régularités qui peuvent être violées dans la réalité. Parmi les hypothèses sur lesquelles les estimations reposent, le plus souvent de façon implicite, la plus importante est sans doute celle d'indépendance entre variables, qui suppose dans notre cas l'absence d'effet direct de C sur R et inversement, ainsi que l'absence d'une cause commune non spécifiée aux deux variables. Imaginons en effet que l'augmentation de la charge C , en laissant moins de temps à l'entretien, entraîne une diminution de la résistance R . Dans ce cas, l'écart $g(r^-, c^+) - g(r, c)$ calculé sous l'hypothèse d'indépendance surestime la marge d'incertitude effective, au sens où la combinaison (r^-, c^+) est plus probable qu'escompté.

L'existence d'incertitudes non probabilisables de ce type justifie le maintien d'une marge supplémentaire dans la colonne de droite de la figure 3.5.1, au-delà de la marge d'incertitude qui a été obtenue par l'analyse probabiliste (qu'elle soit d'origine aléatoire ou épistémique). Cette part d'incertitude peut être étudiée, dans certains cas, par des analyses de sensibilité, conduisant à une adaptation de la marge à la baisse ou à la hausse. Une annulation de celle-ci signifie simplement que l'on accorde une confiance totale au modèle probabiliste sur lequel reposent les estimations.

Ces observations restent valides lorsque l'on passe de la fiabilité physique d'une composante à la sûreté d'ensemble d'une installation nucléaire, même si les raisonnements et les pratiques gagnent naturellement en complexité. L'analyse de la sûreté des installations nucléaires fait ainsi appel à une multitude de modèles des phénomènes physiques et de la disponibilité de tous les systèmes importants pour la sûreté, où interviennent des milliers de paramètres. La pratique aujourd'hui la plus commune est d'avoir tour à tour recours aux méthodes déterministes et probabilistes dans ces différentes composantes (AIEA, 2008). Mais on constate une évolution progressive vers ces dernières, c'est-à-dire une substitution d'estimations réalistes augmentées d'une mesure de l'incertitude aux estimations conservatrices. Les seuils de défaillance réglementaires en vigueur à l'époque où la sûreté nucléaire reposait essentiellement sur l'approche déterministe ont cependant été conservés dans la plupart des cas. L'un des défis actuels de l'analyse de la sûreté est de savoir dans quelle mesure la compréhension des phénomènes à l'œuvre permet de réduire les marges de sûreté nécessitées

par le caractère incertain des modèles probabilistes (AEN, 2007).

3.5.2 Complémentarité des deux approches

En résumé, l'expérience d'exploitation des centrales nucléaires a montré que la défense en profondeur ne permettait pas d'annuler le risque, et que des méthodes probabilistes pouvaient être mises en œuvre à la fois pour la compléter et pour mesurer le risque résiduel ; dans le même temps, ces évaluations probabilistes sont le fruit d'une représentation nécessairement incomplète d'une centrale nucléaire et de son environnement, et de ce fait soumises à des incertitudes parfois difficiles à quantifier. Si les deux approches sont généralement incompatibles au niveau des mesures particulières, elles ont donc néanmoins une part de complémentarité du point de vue de l'architecture d'ensemble de la sûreté, qu'il convient d'exploiter au bénéfice de cette dernière.

Dans un article qui eut une certaine influence sur la politique de sûreté américaine (voir la sous-section 4.2.2.1), des membres de l'Advisory Committee on Reactor Safety, organe consultatif de la NRC, distinguaient en 1999 deux conceptions possibles de cette complémentarité (Sorensen et al., 1999), l'une basée sur une approche traditionnelle, qu'ils appelaient « structuraliste », de la défense en profondeur, et l'autre proposant une refonte « rationaliste » de cette défense.

Dans la première, la logique déterministe de la défense en profondeur continue d'inspirer les principes supérieurs de la sûreté tels que la succession de niveaux de défense et le rôle de chacun d'entre eux, comme elle l'a toujours fait. Les évaluations probabilistes servent à améliorer la sûreté à l'intérieur de ce cadre en identifiant les scénarios qui contribuent le plus au risque d'ensemble. Elles ne cherchent pas, en particulier, à modifier l'équilibre de la défense en profondeur, par exemple la répartition des efforts entre prévention de la fusion et confinement.

Dans la seconde, la démarche de la sûreté est fondée sur l'évaluation probabiliste, et la défense en profondeur constitue un « filet de sécurité » nécessaire pour en compenser les insuffisances. L'analyse des réacteurs par le biais d'EPS permet d'avoir une idée précise des risques à différents niveaux (fusion du cœur, rejets massifs, impacts sanitaires). Les incertitudes relatives aux EPS doivent être évaluées de façon systématique, particulièrement celles liées à l'incomplétude, et des mesures doivent être prises pour tenir compte de leurs effets. Ces mesures s'inspirent de façon ciblée de la logique de la défense en profondeur⁵⁷.

Le choix entre ces deux options relève en partie des préférences sociales et de la façon dont la politique de sûreté les prend en compte, ce qui constitue le sujet du chapitre suivant. Cependant,

57. Nous rapportons ici une version légèrement modifiée de l'analyse de Sorensen et al. (1999), qui comporte de nombreuses références à la réglementation américaine, notamment les objectifs probabilistes de sûreté et la prise en compte d'un niveau acceptable du risque. Les auteurs soutiennent que la « philosophie sous-jacente » du point de vue rationaliste est l'acceptabilité des probabilités d'accident. Sur ce point, notre opinion diffère de la leur : la primauté accordée à l'approche probabiliste nous semble parfaitement conciliable avec l'objectif de réduire autant que possible le risque ainsi mesuré, sans considérer qu'un niveau quelconque soit acceptable en soi. Il n'y a donc pas, selon nous, de correspondance entre le point de vue rationaliste et l'adhésion à un objectif d'acceptabilité du risque.

tout en disant leur préférence pour l'approche rationaliste, les auteurs reconnaissent qu'« aucune n'incorpore de moyen fiable pour déterminer si le degré de défense en profondeur atteint est suffisant » et qu'« en dernière analyse, toutes deux dépendent de la mise en discussion des risques et des incertitudes entre personnes qualifiées, qui s'accordent finalement sur les mesures à prendre au nom de la défense en profondeur ». La raison en est que les méthodes déterministes comme probabilistes sont limitées, de façon ultime, par notre compréhension des phénomènes à l'œuvre dans une dynamique accidentelle, et peuvent de ce fait échouer conjointement dans l'identification d'une séquence.

L'approche probabiliste est un complément indispensable de la stratégie de défense en profondeur. Elle seule peut proposer une hiérarchie des faiblesses et des améliorations possibles de cette stratégie, et tenter d'évaluer les risques résiduels engendrés par l'activité nucléaire, étape nécessaire aux décisions concernant l'acceptabilité de ces risques et leur gestion.

Or l'objectif d'identification et de quantification complète des risques d'accident pose un sérieux problème d'inférence dans le cas de systèmes d'une grande complexité, et dont le fonctionnement (et plus encore le comportement aux limites⁵⁸) n'a fait l'objet que de rares observations. Il suppose de choisir, dans un ensemble quasiment infini de relations de cause à effet possibles à l'intérieur du système, mais aussi dans son environnement lorsque celui-ci interagit avec le système, celles qui sont significatives du point de vue des risques d'accident, de les relier entre elles, d'en évaluer la vraisemblance sur la base d'informations parfois très parcellaires, et de parvenir à une combinaison de ces procédures qui soit cohérente, calculable et révisable.

Le risque d'omission d'interactions importantes pour la sûreté s'accroît à mesure que l'on suit les phases de progression des accidents dans ce qui constitue le deuxième niveau d'EPS, où il s'agit de représenter les événements possibles dans une situation rarement observée de fusion du cœur (Nourbakhsh et Kress, 2005). Mais il est présent même dans des conditions normales de fonctionnement, comme en a témoigné en 2002 l'incident de Besse Davies, à l'occasion duquel on a découvert le degré de susceptibilité de l'alliage utilisé dans le couvercle de la cuve à la corrosion sous l'effet de l'acide borique présent dans le fluide réfrigérant. Face à ce phénomène, les inspections habituelles s'avèrent insuffisantes pour assurer l'intégrité de la cuve (donc de l'enveloppe du circuit primaire), ce qui constitue une faille dans la défense en profondeur. En outre, de nouvelles séquences accidentelles, qui n'avaient pas été envisagées par les études probabilistes, doivent désormais être considérées possibles. Un exemple d'une telle séquence est une fuite d'eau sous pression par une éventuelle brèche ouverte dans la tête de cuve, qui entraînerait un endommagement des mécanismes de conduite des grappes de contrôle et conduisant à la neutralisation du système d'arrêt d'urgence

58. Ceci explique que la recherche sur les phénomènes attendus dans des situations de dégradation de l'installation (interactions corium-béton, percement de la cuve, spéciation des produits de fission) fasse l'objet d'une attention particulière.

du réacteur alors même qu'un accident de perte de réfrigérant primaire serait en cours.

3.6 Conclusion

La démarche déterministe visait, à l'origine, à garantir la sûreté des installations nucléaires dans des conditions où la connaissance de leur fonctionnement et des états dégradés était très limitée, et où la faiblesse des moyens de calcul et d'expérimentation interdisait de simuler de façon précise les phénomènes physiques en jeu. Elle se place sous l'hypothèse de survenue des accidents les plus graves pour concevoir et mettre en place une série de défenses successives contre leur progression. Elle recourt en outre, de façon systématique, à des hypothèses pessimistes dans le dimensionnement des différents systèmes et composantes. En se cumulant les unes aux autres, les marges ainsi introduites permettent, pense-t-on, de compenser toutes les incertitudes relatives à l'évolution du système. L'enceinte de confinement elle-même en fournit un bon exemple : alors qu'elle n'était pas conçue, initialement, pour contenir les conséquences d'une fusion du cœur que l'on tenait pour trop improbable, elle s'est avérée avoir cette capacité pour un bon nombre de séquences accidentelles, grâce à la somme d'hypothèses conservatrices intégrés à sa conception.

L'inconvénient de la démarche est qu'elle ne repose pas sur une représentation détaillée des installations et des dysfonctionnements qui peuvent s'y produire, mais sur des scénarios d'accidents tirés de la pratique et de l'imagination de leurs concepteurs⁵⁹. Elle est donc particulièrement exposée à des scénarios imprévus, notamment à des enchaînements de défaillances mineures dont les effets se cumuleraient, ainsi qu'à des événements qui affecteraient plusieurs lignes de défense à la fois. Ces faiblesses ont été mises en lumière par l'accident de Three Mile Island et par la myriade de défaillances de cause commune détectées au fil des ans dans des systèmes de sûreté réputés redondants.

Les évaluations probabilistes, à l'inverse, sont fondées sur une telle représentation. En cherchant à identifier et modéliser tous les événements significatifs du point de vue de la sûreté, elles parviennent à construire les séquences accidentelles à partir d'éléments simples et, de ce fait, n'ont pas à les sélectionner *a priori*. Le premier de leurs apports a ainsi été d'améliorer la compréhension des dynamiques accidentelles dans les centrales. Elles permettent, en la matière, d'aller plus loin que l'approche déterministe en complétant l'ensemble de scénarios-enveloppe considérés par celle-ci.

La démarche probabiliste remplace en outre l'usage systématique d'hypothèses conservatrices par une approche fondée sur la valeur la plus probable de chaque variable, augmentée d'une marge d'incertitude explicite. Elle permet ainsi de séparer de façon plus claire et rigoureuse la marge de

59. Il serait incorrect de dire que la démarche déterministe ne se fonde sur aucune représentation formelle. Comme nous l'avons vu, elle prend en effet pour point de départ l'ensemble des causes directes d'un accident grave : l'arrêt de l'une au moins des fonctions de sûreté. Nous chercherons dans le chapitre 5 à décrire cette représentation à partir du formalisme des réseaux bayésiens causaux.

sûreté disponible de ce qui relève de l'incertitude relative à la grandeur mesurée, et de s'assurer de la cohérence des marges de sûreté d'une variable à l'autre. De ce gain de cohérence découle un deuxième grand apport des EPS, qui est leur capacité à comparer des séquences d'accident, calculer les contributions d'événements-type au risque agrégé, et pouvoir ainsi définir des priorités dans l'amélioration de la sûreté.

En se fondant sur un modèle de l'installation, les évaluations probabilistes sont cependant sujettes à une incertitude sur l'estimation de ses grandeurs caractéristiques et sur la justesse de sa structure. Une évaluation insuffisante ou erronée de cette incertitude signifie la disparition des avantages de l'approche probabiliste, qui peut alors conduire à accepter des marges de sûreté insuffisantes et à accorder une importance exagérée à certaines séquences accidentelles.

Le second facteur plaçant pour une analyse rigoureuse de l'incertitude est l'encadrement des probabilités subjectives. Les EPS prennent en compte des événements de diverses natures : défaillance de composantes, interventions humaines, aléas naturels⁶⁰, etc. Les probabilités d'occurrence de ces événements sont évaluées par des méthodes physiques, statistiques et logiques. Dans de nombreux cas, ces événements sont trop rares et difficiles à reproduire expérimentalement pour que leur fréquence puisse être estimée de façon empirique. Les probabilités subjectives établies par des experts jouent donc un rôle important dans les EPS. À cet égard, la représentation explicite de l'incertitude constitue à la fois une obligation faite aux analystes de justifier leurs choix subjectifs, et un moyen d'encadrer et d'harmoniser des choix.

L'évaluation et la représentation de l'incertitude peuvent cependant avoir un coût considérable et rencontrer des difficultés méthodologiques importantes. Ceci explique qu'elle n'ait été conduite que de façon sélective, et qu'il n'y ait finalement qu'un nombre limité d'aspects du dimensionnement pour lesquels des méthodes probabilistes ont été substituées à l'approche déterministe⁶¹. Ceci est particulièrement vrai pour les incertitudes liées à la structure des modèles, qui ont rarement fait l'objet d'une évaluation complète.

Face à ce risque, l'unique parade possible est la remise en cause incessante, par les moyens de l'investigation scientifique, des approches existantes de la sûreté. Nous chercherons, dans le chapitre 6, à illustrer ce propos dans le cas particulier des agressions externes.

60. Comme nous l'avons dit, les EPS françaises n'intègrent pas les agressions internes ou externes.

61. Comme nous le verrons au chapitre suivant, certains régulateurs nationaux se sont montrés plus enclins à opérer une telle substitution que d'autres.

Chapitre 4

Les politiques de sûreté nucléaire

No one is in a position to demonstrate that a reactor accident with consequent escape of fission products to the environment will never happen.

Clifford K. Beck, 1967, cité par Walker (2000)

Prenez toutes les autres (techniques) depuis la Révolution Industrielle : elles ont toutes connu des accidents graves.

Massé (1981)

Après avoir étudié les méthodes d'analyse de la sûreté, nous nous intéressons à présent aux politiques publiques qui visent d'une part à établir des objectifs pour la sûreté des installations nucléaires au nom de la société, et de l'autre à s'assurer que ces objectifs de sûreté sont atteints. Ceci passe par la surveillance et le contrôle de la conception, de l'exploitation, des éventuelles modifications, de l'arrêt et du démantèlement des installations nucléaires, ainsi que la préparation à la gestion des accidents, un ensemble de domaines auquel on donne parfois le nom de contrôle de la sûreté nucléaire¹.

Il est usuel, dans les sciences sociales et en particulier en économie, d'étudier les politiques publiques sous deux angles : celui des mesures elles-mêmes, des objectifs qu'elles prétendent atteindre et des moyens qu'elles mettent en œuvre ; et celui des institutions chargées de les concevoir, de les adopter et de les implémenter. Nous adoptons tour à tour ces deux points de vue, en nous intéressant principalement aux cas des États-Unis et de la France.

Nous ne nous donnons pas pour objectif d'étudier l'ensemble de ce sujet en détail et avec profondeur, mais simplement d'établir quelques observations susceptibles d'apporter des éléments de réponse à notre interrogation principale : comment les décisions relatives à la sûreté nucléaire sont-elles prises à partir des approches analytiques décrites dans les parties précédentes ? Pour cela,

1. Quitte à anticiper sur le contenu de ce chapitre, notons ici qu'en France, l'ASN est une autorité de contrôle alors que son équivalent aux États-Unis, la NRC, est une agence de régulation.

il nous semble nécessaire de revenir sur la période historique qui a vu la naissance de l'industrie nucléaire civile et l'élaboration des premières politiques de sûreté. Bien entendu, il ne s'agit pas de proposer ici une lecture historique novatrice, mais de rappeler quelques éléments largement reconnus par les historiens, et qui nous semblent importants pour la compréhension des questions et des approches actuelles de sûreté nucléaire. Nos principales sources sont des témoignages d'acteurs de cette époque (notamment Goldschmidt (1980), Jeanneney (1981), Massé (1981), Tanguy (1988) et Boiteux (1993) en France, Okrent (1981) aux États-Unis) ainsi que des travaux de recherche plus récents (Balogh (1991), Walker (1992, 2000) et surtout Foasso (2003), auquel nous empruntons largement son récit historique de la naissance de la sûreté nucléaire).

La première partie du chapitre traite des années 1945-1967, au cours desquelles se sont posées les premiers problèmes de décision relatifs à l'existence d'une l'industrie nucléaire, et dont la sûreté ne représentait qu'un aspect. La deuxième et la troisième parties décrivent la constitution de la politique de sûreté respectivement aux États-Unis et en France, depuis la fin des années 1960 à nos jours. La dernière partie propose une discussion d'ensemble de la manière dont les risques liés aux activités nucléaires civiles sont pris en compte dans ces deux pays.

4.1 Les décisions initiales (1945-1967)

4.1.1 La difficile séparation entre atomes pour la guerre et atomes pour la paix (1945-1953)

4.1.1.1 Une technologie développée en marge de la course aux armements atomiques

Dans l'immédiat après-guerre, les principales puissances alliées, bien que conscientes des nombreuses potentialités de la technologie nucléaire naissante dans le domaine civil, avaient accordé une totale priorité à ses applications militaires.

La loi américaine de 1946 (*Atomic Energy Act* ou loi Mac Mahon), qui définissait le cadre du développement de l'énergie nucléaire aux États-Unis, mettait en place une politique du secret et du contrôle qui a été qualifié d'isolationnisme atomique (Goldschmidt, 1980). L'objectif était de maintenir l'exclusivité dont jouissaient les États-Unis en limitant l'accès des autres pays à la fois aux connaissances techniques et aux ressources en minerai d'uranium nécessaires à la réalisation de la fission. Sur le plan intérieur, la loi réservait à l'État fédéral le monopole de l'activité nucléaire, concentrée dans la production industrielle et l'amélioration de l'armement atomique. Les applications commerciales étaient interdites. Afin de contrôler et d'orienter le développement de la puissance atomique dans le pays, la loi créait une agence spécifique, l'*Atomic Energy Commission* (AEC), qu'elle plaçait sous le contrôle d'un comité mixte du Congrès.

L'AEC reprenait des mains de l'armée l'immense complexe industriel mis sur pied pendant la

guerre pour mener à bien le projet Manhattan et fabriquer en quantité suffisante les deux substances employées dans les bombes atomiques : l'uranium 235 et le plutonium². Ce complexe comprenait notamment les deux grandes centrales (piles dans le vocabulaire de l'époque) de Hanford, dans l'état de Washington, qui étaient destinées à produire du plutonium et dont la chaleur produite était évacuée par les eaux de la rivière Columbia, sans aucune tentative de la récupérer. L'Agence entreprenait également de financer et d'orienter les projets de recherche militaire dans le domaine atomique, notamment ceux s'intéressant à la combustion contrôlée à des fins de propulsion. C'est au sein de l'un de ces projets, celui de sous-marin nucléaire, que fut mis au point en 1953 un réacteur utilisant des éléments d'uranium métal gainés de zirconium, modéré et refroidi à l'eau ordinaire et ancêtre de la majorité des réacteurs en fonctionnement dans le monde.

L'orientation militariste américaine était partagée par le Royaume-Uni et l'URSS, pays les plus avancés après les États-Unis dans la maîtrise de la technologie nucléaire au sortir de la guerre (avec le Canada).

Au Royaume-Uni, le développement militaire et civil de l'énergie atomique était placé dès 1946 sous l'autorité d'un *Controller of Production* attaché au *Ministry of Supply*. Bien que handicapé par le retrait américain, le pays conservait suffisamment d'autonomie pour s'engager seul dans la fabrication de la bombe, notamment en construisant deux réacteurs destinés à la production de plutonium à Windscale. Ces réacteurs utilisaient l'uranium naturel, étaient modérés par le graphite et refroidis par air, ce qui permettait d'envisager une récupération aisée de la chaleur dégagée. Leur succès de la méthode de refroidissement par gaz conduisit à la décision, prise en deux temps en 1953 et 1955, de construire à Calder Hall huit centrales « à double objectif » : d'une part accroître la production militaire de plutonium, et de l'autre produire de l'électricité. En 1954, la décision était prise de placer le monopole de l'énergie nucléaire sous la supervision d'une autorité indépendante inspirée du modèle de l'AEC, la *United Kingdom Atomic Energy Authority* (UK AEA). En 1956, la première centrale à uranium naturel, graphite et gaz carbonique (UNGG) voyait le jour.

2. Deux éléments se prêtent facilement à la fission nucléaire : l'uranium 235 et le plutonium 239. Une masse de quelques kilos de l'une ou l'autre suffit à entretenir une réaction en chaîne, que celle-ci soit orientée vers l'explosion ou contrôlée, par freinage et absorption des neutrons émis, à des fins de récupération de la chaleur et de production d'électricité.

L'uranium 235 est présent à raison de sept parties pour mille dans l'uranium naturel, dont l'essentiel est constitué d'isotope 238. Pour entretenir une réaction en chaîne explosive, il doit être bien plus concentré, et donc au préalable isolé par un procédé de séparation isotopique pour produire de l'uranium enrichi de qualité militaire. Dans une réaction contrôlée, toutefois, la relative pauvreté de la matière fissile en uranium 235 peut être compensée dans une certaine mesure par le freinage des neutrons issus de la fission, qui augmente leur capacité à provoquer de nouvelles réactions. Pour cela, l'uranium doit être placé en présence d'un modérateur qui ralentisse les neutrons sans trop les absorber. Diverses substances ont cette propriété : l'eau lourde, composée d'oxygène et de deutérium (isotope de masse 2 de l'hydrogène) et le graphite (carbone pur) permettent d'obtenir une réaction en chaîne dans une masse de plusieurs tonnes d'uranium naturel ; l'eau ordinaire suffit si la teneur de l'uranium en isotope 235 est portée à 3,5% au minimum, et des quantités moindres de combustible sont alors nécessaires, ce qui permet de réduire significativement la taille des réacteurs.

Le plutonium n'existe pas à l'état naturel mais il est obtenu comme résidu de la combustion de l'uranium, par transmutation de l'isotope 238.

La matière fissile concentrée utilisée dans les bombes peut donc être obtenue de deux façons : par séparation de l'uranium 235 dans des usines d'enrichissement ou par production de plutonium dans des centrales.

La situation était différente en France. Cette dernière avait pris un certain retard par rapport aux États-Unis et au Royaume-Uni sur les plans scientifique et technique durant la guerre, qu'il s'agissait pour elle de compenser avant toute autre chose. Elle fut donc la seule à placer le développement nucléaire sous l'égide d'un organisme civil à vocation principalement scientifique. Créé dès 1945, le Commissariat à l'Énergie Atomique (CEA) se vit confier un mandat mixte : explorer les possibilités de la fission à la fois à des fins de Défense nationale et dans la production d'électricité. En raison des faibles quantités d'uranium naturel dont la France disposait dans le contexte international de l'époque, l'ambition initiale devait être limitée à la construction de réacteurs de recherche à faible puissance, notamment celle de la pile Zœ (puissance Zéro, Oxyde d'uranium, Eau lourde) entreprise dès 1946 dans le centre de Fontenay-aux-Roses. Ce n'est qu'au début des années 1950, grâce notamment à la découverte de minerai d'uranium dans le Limousin et sous l'impulsion du premier plan quinquennal, que la France s'est engagée dans la production à vaste échelle de plutonium destiné, en particulier, à mettre au point la bombe française. Le Plan prévoyait notamment la construction de deux grandes centrales modérées au graphite à Marcoule, initialement sans récupération de la chaleur. Ce schéma fut toutefois modifié en 1953 avec l'ajout d'une unité de production d'électricité, permise grâce à l'adoption d'une méthode de refroidissement par gaz carbonique comprimé mise en œuvre dans la deuxième pile expérimentale de Saclay. La France s'engageait ainsi, parallèlement au Royaume-Uni, dans la filière à double objectif UNGG³.

4.1.1.2 Un tournant politique

Le discours *Atoms for Peace*, prononcé en décembre 1953 par le président Eisenhower devant l'Assemblée générale des Nations unies, est souvent présenté comme l'acte de naissance politique des applications nucléaires civiles. Or *Atoms for Peace* constituait avant tout l'élément-clé d'une stratégie (en partie couronnée de succès) de détente et de repositionnement des États-Unis dans un monde encore traumatisé par l'usage de la bombe atomique à Hiroshima et Nagasaki. L'Union Soviétique avait créé une surprise et un émoi considérables parmi les pays occidentaux en réalisant sa première explosion nucléaire quelques années plus tôt, en août 1949. Les États-Unis avaient répliqué en 1950 en annonçant leur intention de fabriquer une bombe à hydrogène d'une puissance considérablement plus élevée que les armes existantes. En 1952, le Royaume-Uni était à son tour parvenu à faire exploser une bombe nucléaire. Enfin, en 1953, l'Union Soviétique avait de nouveau créé la surprise en provoquant une explosion thermonucléaire qui indiquait qu'elle était proche, elle aussi, de maîtriser la bombe H. L'escalade ainsi engagée entre les deux grandes puissances était coûteuse, périlleuse et produisait un effet déplorable pour l'image des États-Unis dans le monde.

3. Entre 1952 et 1966, le CEA construisit au total neuf réacteurs UNGG, notamment celui de Chinon qui était, au moment de sa mise en œuvre, l'unité de production d'électricité la plus puissante au monde. Ces réacteurs sont aujourd'hui tous déclassés.

Mais surtout, elle démontrait l'échec de la politique américaine d'isolationnisme, qui semblait handicaper le développement du pays dans un secteur prometteur sans avoir les effets escomptés sur le plan militaire.

La proposition d'Eisenhower consistait en une *réaffectation* des substances fissiles employées dans la production d'armements à des usages pacifiques, sous l'égide d'un organisme international qui veillerait à la coordination du programme dans les pays concernés. Favorablement accueillie par l'Union Soviétique, elle a finalement conduit à trois avancées importantes en matière de coopération internationale dans le domaine nucléaire : la levée du secret et la reprise de la collaboration scientifique, notamment à l'occasion de deux grandes conférences internationales sur l'utilisation de l'énergie atomique à des fins pacifiques, tenues à Genève en 1955 et 1958 ; la création, obtenue des soviétiques en contrepartie du point précédent, de l'Agence Internationale de l'Énergie Atomique, chargée du contrôle international de l'usage civil de l'énergie nucléaire ; enfin, l'institution de l'énergie nucléaire comme l'un des axes de reconstruction et de réunification de l'Europe de l'Ouest, incarné par la communauté européenne Euratom (1957) et par l'Agence Européenne de l'Énergie Nucléaire (1958), affiliée à l'Organisation Européenne de Coopération Économique⁴ qui supervisait la distribution des fonds du Plan Marshall.

Aux États-Unis, la révision de l'*Atomic Energy Act*, décidée en 1954 dans le prolongement du discours d'Eisenhower, modifiait radicalement la donne nucléaire intérieure : l'AEC était chargée de promouvoir l'utilisation civile de l'énergie nucléaire, notamment en encourageant l'arrivée d'acteurs privés, qui pouvaient désormais obtenir des licences pour la construction et l'exploitation de centrales nucléaires. Dès 1953, le promoteur et pilote du projet de sous-marin nucléaire, le commandant de marine Rickover, obtint l'accord de l'AEC pour la construction d'une première centrale civile reprenant le principe du réacteur sous-marin pour produire de l'électricité. La construction en fut confiée à la firme privée Westinghouse et en décembre 1957, le premier réacteur à eau pressurisée (REP)⁵ américain entra en fonctionnement à Shippingport (Pennsylvanie).

Le rôle de promotion joué par l'AEC et, d'une façon plus générale, par les autorités publiques américaines, alla en se renforçant au long des années 1950 en raison des réticences affichées par les entreprises privées de production d'électricité. Il apparaissait, d'une part, que la plupart de celles-ci avaient une taille insuffisante pour supporter les coûts fixes induits par la fabrication et l'exploitation d'une centrale. Lors de la construction de la centrale de Shippingport, une solution originale avait été trouvée à ce problème sous forme d'un montage public-privé : l'AEC avait financé la construction du réacteur proprement dit et en restait propriétaire ; elle avait également proposé une mise à disposition gratuite du combustible durant les premières années d'exploitation ; de son

4. Préfigurations respectives de L'Agence de l'Énergie Nucléaire (AEN) et de l'Organisation de Coopération et de Développement Économiques (OCDE).

5. *Pressurized water reactor* (PWR) en américain.

côté, l'opérateur privé en charge de l'exploitation avait couvert les coûts de construction de la partie non nucléaire et s'était engagé à acheter à prix élevé la chaleur produite. La technologie des REP mise en œuvre à Shippingport n'était cependant pas compétitive à cette époque et l'AEC y avait eu recours pour des motifs de prestige international. Ses adversaires l'estimaient trop grossière et sans avenir, au point où ils avait détourné son acronyme anglais en *Power Without Reason* (Balogh, 1991, p.107). L'AEC lança donc dès 1955 un programme de « démonstration » de puissance à travers lequel elle proposait un certain nombre d'avantages supplémentaires aux entreprises désireuses de se lancer dans l'aventure, notamment de financer une part importante des coûts de recherche et de développement de futures centrales. En 1956, une nouvelle loi fut adoptée qui exemptait de la législation sur la concurrence les opérateurs désireux de s'allier pour partager les coûts de construction et les bénéfices technologiques d'une centrale. Il restait, néanmoins, le problème des risques financiers liés à un possible accident nucléaire. L'industrie de l'assurance proposait alors une couverture de la responsabilité civile à hauteur de 60 millions de dollars, ce qui était estimé très inférieur aux dommages engendrés par un accident grave (voir ci-dessous). En 1957, le Congrès des États-Unis adopta le *Price-Anderson Act* qui garantissait aux opérateurs de centrales nucléaires une prise en charge intégrale par l'État des dommages au-delà de ce plafond, à hauteur de 500 millions de dollars. La loi fut adoptée initialement pour une durée de dix ans, qui devait suffire, aux yeux du législateur, à montrer la sûreté de fonctionnement des centrales et rendre le risque nucléaire totalement assurable⁶.

Il convient de noter que la politique américaine de soutien à l'émergence d'une industrie électronucléaire attire l'attention en raison à la fois de l'ampleur du revirement de 1953-54 et de l'organisation du secteur de l'électricité, principalement privé et faiblement concentré. Dans les autres pays engagés dans la construction de centrales civiles (ou à double objectif), cette politique était tout aussi marquée mais moins visible, car elle avait été mise en place de façon plus progressive au sein de structures verticalement intégrées et contrôlées par l'État. En particulier, ces États ont tous accepté de tenir le rôle d'assureurs en dernier ressort au bénéfice de l'industrie électronucléaire, généralement de façon implicite.

6. Ce schéma assurantiel a été depuis reconduit six fois, et il s'étend désormais jusqu'en 2025. La subvention qu'il procure aux opérateurs de centrales a été récemment estimée à 600 000 dollars par réacteur et par année par l'office du budget du Congrès (CBO, 2008). L'étude, qui se fonde sur les probabilités d'accident produites par l'EPS de niveau 2 de la centrale de Surry (NRC, 1990) *sans* prendre en compte les incertitudes afférentes (notamment celle liées à l'incomplétude de l'analyse), évalue à $3 \cdot 10^{-8}$ la probabilité d'un accident entraînant des dommages de 500 milliards de dollars.

4.1.2 Naissance des doctrines de sûreté (1954-1967)

4.1.2.1 Considérations de sûreté lors du développement des premières centrales

Les agences telles que l'AEC, le CEA et la UK AEA avaient, en marge de leur mission principale de promotion de la technologie nucléaire, également la charge d'en contrôler les dangers. Aux États-Unis, l'AEC mit ainsi en place dès 1947 le *Reactor Safeguards Committee* (RSC), constitué de scientifiques renommés, afin de la conseiller au sujet de la sûreté des centrales nucléaires.

À partir du début des années 1950, des doctrines de sûreté sont venues progressivement s'ajouter aux simples pratiques d'ingénierie mises en œuvre au cours de la décennie précédente dans la construction et le fonctionnement des réacteurs militaires ou de recherche. La tâche des premiers régulateurs de la sûreté était redoutable, puisqu'aucune technologie ne s'était jusqu'à lors développée aussi vite et n'avait représenté autant de promesses et de dangers à la fois. Les intéressés, qui étaient souvent des scientifiques engagés dans le développement des premiers réacteurs, étaient bien sûr conscients de ces dangers, comme en témoigne l'extrait suivant d'une lettre d'Edward Teller, père de la bombe H américaine, premier président du RCS et pionnier de la sûreté nucléaire, au président du comité mixte du Congrès :

"nuclear plants contain radioactive poisons. In a nuclear accident, these poisons may be liberated into the atmosphere or into the water supply. In fact, the radioactive poisons produced in a powerful nuclear reactor will retain a dangerous concentration even after they have been carried downwind to a distance of ten miles. Some danger might possibly persist to distances as great as 100 miles" (Teller, 1953).

Ces spécialistes semblaient cependant très confiants dans leur capacité à contrôler ces dangers. Dans la même lettre, Teller déclarait ainsi :

"The main factors which influence reactor safety are, in my opinion, reasonably well understood. There have been in the past a few minor incidents, all of which have been caused by neglect of clearly formulated safety rules. Such occasional accidents cannot be avoided. It is rather remarkable that they have occurred in such a small number of instances. I want to emphasize in particular that the operation of nuclear reactors is not mysterious and that the irregularities are no more unexpected than accidents which happen on account of disregard of traffic regulations"(Teller, 1953).

De même, on pouvait lire dans le premier rapport de la UK AEA en 1954 :

"The first important thing to recognise is that it is impossible for an 'atomic explosion' to take place in a power reactor. If nuclear power facilities are properly designed any accidents that may occur will be no more dangerous than accidents in other industries. [...] The reactors that will be built for the commercial production of electricity will

present no more danger for people living nearby than many existing industrial works that are sited within built-up areas. Nevertheless the first stations, even though they will be of inherently safe design, will not be built in heavily built-up areas" (cité par Foasso (2003, p. 103)).

En France, quelques années plus tard, le premier Cours de Génie Atomique élaboré au CEA professait :

« Les installations atomiques posent des problèmes de sûreté qui sont dans leur ensemble du même ordre de grandeur que ceux posés par certaines installations classiques. C'est ainsi que l'explosion dans un grand port d'un navire chargé de munitions ou d'essence, ou l'incendie d'un dépôt de gaz asphyxiants au milieu d'une ville, peuvent entraîner autant de morts, de maladies incurables ou de dégâts matériels qu'un grave incident nucléaire sur un navire ou une centrale atomique. Cependant, certains facteurs conduisent à attacher à la sûreté en matière atomique une importance encore plus grande qu'en matière classique : [...] les répercussions psychologiques d'accidents nucléaires peuvent être plus grandes étant donné la sensibilité des opinions publiques en ce domaine » (De Valthaire, 1960).

D'un point de vue actuel, l'analogie entre une technologie aussi complexe que celle des réacteurs nucléaires, pour laquelle on ne disposait alors d'aucun recul historique et d'autres issues de la révolution industrielle, totalement maîtrisées et étudiées depuis des décennies sur le plan statistique⁷ est surprenante. Elle semble témoigner d'une confiance excessive dans la compréhension de la sûreté des réacteurs et d'une représentation assez naïve des interactions entre homme et machine. En revanche, l'importance accordée aux aspects psychologiques et aux réactions de l'opinion publique peuvent paraître surprenantes d'actualité.

L'éclairage historique privilégie cependant une tout autre interprétation de ces déclarations (Foasso, 2003, pp. 78 et suivantes). La préoccupation des opinions publiques occidentales en matière nucléaire ne se portait alors pas vers les centrales, mais vers l'arme atomique. Or, comme nous l'avons vu, la technologie nucléaire civile avait été dérivée des usages militaires, et lui était encore étroitement associée. Les scientifiques appelés à conseiller les autorités publiques en matière de sûreté étaient alors partagés entre deux objectifs : d'une part, de contribuer au développement de l'énergie nucléaire civile en démystifiant son usage, et notamment en luttant contre son association, dans l'esprit du public, à la bombe atomique ; d'autre part d'imposer une approche rigoureuse de la sûreté nucléaire, quitte à retarder pour cela le développement industriel. La thèse selon laquelle il était à la fois possible et nécessaire de garantir la sûreté des centrales nucléaires (à quelques

7. À titre d'exemple, les premières statistiques fédérales d'accidents de la route ont été collectées aux États-Unis au cours des années 1910 (Eastman, 1984).

incidents mineurs près, tous dus à la négligence des hommes) permettait de ménager ces deux objectifs.

Pour protéger les populations, les premières préférences des scientifiques américains s'inscrivaient dans la droite ligne du projet Manhattan, dont le principe central de sûreté était d'éloigner les sites nucléaires, qui n'étaient alors pas confinés, des zones densément habitées. En 1950, le RSC édicta une première règle de sûreté en ce sens concernant le choix des sites devant accueillir des centrales nucléaires de puissance. La règle définissait le rayon d'une zone d'exclusion autour d'une centrale en fonction de la puissance thermique de celle-ci.

La conception restrictive des scientifiques ne convenait cependant pas aux responsables industriels car elle limitait le choix de sites adaptés, imposait l'achat de surfaces considérables de terrain et supposait aussi la construction d'importantes lignes électriques dans l'hypothèse de construction de centrales civiles. En 1951, l'AEC se dota d'un second comité d'experts, cette fois provenant en large part de l'industrie et traitant spécifiquement du problème des sites, l'*Industrial Committee on Reactor Location Problems*. Deux ans plus tard, les deux comités fusionnèrent au sein de l'*Advisory Committee on Reactor Safeguards* (ACRS), placé sous la direction C. Rogers McCullough. McCullough était un ingénieur ayant mené une grande partie de sa carrière au sein de la firme Monsanto. Tout en assurant la présidence de l'ACRS, il fut successivement employé par Monsanto, puis par l'AEC elle-même et de nouveau par Monsanto entre 1953 et 1957, fournissant ainsi un exemple précoce des relations délicates entre agences réglementaires, entreprises privées et experts⁸ (Balogh, 1991, p.132).

Il faut noter que les responsables industriels étaient également nombreux à considérer qu'en raison de l'analogie avec la bombe atomique, les premières phases du développement de l'industrie nucléaire seraient particulièrement sensibles ; c'est d'ailleurs probablement l'un des facteurs qui expliquaient leur réserve. Dans des conditions où un accident important pouvait constituer un faux pas fatal, un niveau élevé de sûreté devait non seulement être conciliable avec leurs objectifs économiques, mais même en faire partie. Dans les termes de la théorie économique, les coûts sociaux d'un accident étaient donc, au moins en partie, internalisés durant cette phase sensible de décollage de la technologie. La divergence des industriels avec les scientifiques portait sur les moyens à mettre en œuvre pour la sûreté des centrales. L'alternative qu'ils préconisaient à l'éloignement était le confinement : entourer les réacteurs d'un bâtiment étanche et résistant à la pression qui permettrait de retenir les gaz radioactifs émis par un éventuel accident. Selon la communication de deux ingénieurs de la General Electric Company à la première conférence de Genève, la présence d'une telle enceinte et de dispositifs supplémentaires de protection permettrait de rendre les centrales « intrinsèquement sûres » (Parker et Healey, 1955).

8. Désignées en anglais par l'expression des « portes tournantes » (*revolving doors*).

La nouvelle loi de 1954 clarifia les responsabilités de l'AEC en matière de sûreté, qui continuaient à coexister avec les objectifs de développement des usages à la fois civils et militaires de l'énergie nucléaire : la Commission devait veiller à avoir une *assurance raisonnable* que les applications industrielles n'engendreraient pas de *risque indu* pour la santé et la sécurité de la population.

Dans les deux années qui suivirent, le premier réel débat sur la sûreté des centrales divisa la communauté nucléaire américaine. Il portait sur l'arbitrage souhaitable entre l'importance de la zone d'exclusion autour d'une centrale et les dispositifs de sûreté dont elle était équipée (notamment l'enceinte de confinement). La bataille fut gagnée par les partisans des dispositifs de sûreté et en 1956, l'AEC décida d'un assouplissement de ses règles de localisation alors même que la puissance des centrales en conception s'accroissait de façon marquée.

La loi de 1954 avait également défini les étapes de la procédure d'obtention d'une licence de construction et d'exploitation de centrale nucléaire, qui comprenaient notamment la réalisation d'un rapport préliminaire des dangers (*preliminary hazards summary report*). Pour les opérateurs privés, il s'agissait désormais de *démontrer* qu'ils avaient pris des dispositions suffisantes pour garantir la sûreté de leur centrale, même si le régulateur ne disposait pas encore pour cela d'une doctrine formelle transcrite sous forme de principes, de normes et de mesures.

4.1.2.2 La lente marche vers la quantification de l'incertitude

Dans un article présenté à la première conférence de Genève en 1955, trois membres de l'ACRS faisaient une synthèse de la doctrine du Comité en matière de sûreté. Cet article évoquait déjà les trois fonctions de sûreté décrits au chapitre précédent (la sous-criticité, le refroidissement et le confinement), faisait la preuve d'une parfaite compréhension des conséquences catastrophiques d'un accident majeur et proposait une définition de la sûreté où pointait déjà la notion de risque résiduel acceptable : « ce que l'on entend par sûreté du réacteur, c'est le fait de réduire ces dangers à un risque calculé et acceptable aussi minime que possible » (McCullough, Mills et Teller (1955) cités par Foasso (2003, p.81)).

Mais quel était ce risque ? L'étude WASH-740 publiée par l'AEC en 1957 tentait pour la première fois d'évaluer les conséquences maximales d'un accident nucléaire en se plaçant sous l'hypothèse que les dispositifs de sûreté avaient failli. En considérant un réacteur d'environ 150 MWe situé à 30 kilomètres d'une ville d'un million d'habitants, et un accident où la moitié des produits de fission présents dans le cœur en fin de cycle du combustible seraient dispersés, elle évaluait les dommages à 3 400 morts, 40 000 personnes contaminées, 400 000 km^2 de terres agricoles contaminées et un coût total de 7 milliards de dollars de l'époque. L'étude s'ouvrait par un chapitre sur la probabilité d'un tel accident, pour exclure en termes définitifs la possibilité de la connaître, tout en énumérant avec une remarquable prescience les trois méthodes qui permettraient au moins, vingt ans plus tard, son

évaluation :

"The probability of occurrence of publicly hazardous accidents in nuclear power reactor plants is exceedingly low. [...] One fact must be stated at the outset : no-one knows or will ever know the exact magnitude of this low probability of a publicly hazardous reactor accident. In trying to establish some estimation of this quantity, three possible approaches might be used :

1. Operate enough reactors for a sufficient length of time to obtain an indication of the accident probability.
2. Give careful consideration and approximate numerical values to all separate factors which would either prevent or cause such an accident, then try to calculate, or guess, the composite result of these factors and hence the likelihood of occurrence of accidents.
3. Obtain a weighted average of the best judgements and judicious opinions of the most experienced and knowledgeable experts in the field" (AEC, 1957, p.3).

L'étude estimait ces pistes impraticables car trop incertaines, mais les explorait malgré tout, certes pas de façon approfondie mais assez pour citer, à titre d'exemple, des probabilités subjectives attribuées par des experts à différents scénarios accidentels. Concernant un accident majeur avec rejets massifs de substances radioactives, ces probabilités étaient comprises entre 10^{-5} et 10^{-9} par réacteur et par an, et l'étude proposait même un calcul en espérance sur la base de la borne supérieure de cet intervalle : le résultat était de trois personnes tuées en moyenne par an par de tels accidents dans le pays, que l'étude comparait aux 40 000 victimes des accidents de la circulation. L'étude estimait ensuite que l'objectif ultime des évaluations de sûreté n'était pas de calculer les probabilités d'accident, mais de s'assurer que ces probabilités étaient extrêmement faibles. Le chapitre se terminait sur ces mots :

"Thus, since there is protection against 'credible' accidents, no damages to the public will occur unless 'incredible' accidents take place. It must be recognized, of course, that errors in judgement can be committed, with resulting occurrence of what was believed to be an "incredible" accident. Nevertheless, the consistent and rigorous execution of these procedures for every reactor warrants *a considerable degree of confidence* that safeguards against serious accidents have been incorporated, and that the probabilities of such occurrences are small" (AEC, 1957, p.6, nous soulignons).

L'étude WASH-740 constituait une analyse remarquable par sa relative objectivité et par l'étendue des phénomènes examinés. Toutefois, comme le montrent ces extraits, elle annonçait déjà l'inconfort de la position de analystes de sûreté face à la notion de probabilité et préfigurait le « cadre de la probabilité de fréquence » (voir la sous-section 3.4.2.2) : la probabilité d'un accident est ici considérée comme une grandeur objective qu'il est impossible de connaître, mais éventuellement possible

d'estimer ; pour cela, on peut s'en remettre à des croyances d'experts, c'est-à-dire des probabilités subjectives ; mais l'incertitude du résultat (c'est-à-dire, en l'espèce, les écarts entre probabilités subjectives) rend son utilisation délicate ; on attribue finalement une probabilité subjective élevée (un degré de confiance considérable, sans plus de précision) au fait que la probabilité objective d'accident est petite.

L'argument de WASH-740 à propos de la faiblesse très probable de la probabilité d'accident ne permit pas de compenser l'impact des conséquences catastrophiques que décrivait l'étude, et sa publication conduisit à une série de réactions très négatives. L'AEC tenta de les circonscrire rapidement, tout en prolongeant le raisonnement amorcé dans WASH-740, à travers le concept d'Accident maximum crédible (AMC). L'AMC constituait l'événement le moins vraisemblable et le plus grave auquel on estimait devoir se préparer. Pour l'obtenir, on devait commencer par passer en revue tous les scénarios d'accident que l'on parvenait à identifier pour une installation, puis écarter ceux qui semblaient tout à fait improbables selon le jugement subjectif d'experts. Il n'était plus nécessaire de considérer l'éventualité de ces accidents possibles mais « incroyables » et de chercher à prévoir leurs conséquences.

Le concept d'AMC a joué un rôle fondamental dans la constitution de l'approche déterministe de la sûreté en fournissant un ancrage à l'activité naissante de contrôle et de régulation. Les industriels étaient désormais vigoureusement engagés dans le développement de l'énergie nucléaire, et exigeaient de disposer de normes sur lesquelles s'appuyer⁹. L'AEC, pour sa part, avait besoin de concepts opérationnels qui puissent être appliqués de façon uniforme aux différents types de centrales à l'étude. En 1958, un article de représentants de l'AEC à la deuxième conférence internationale de Genève décrivait les premiers éléments d'une doctrine où étaient énoncés certains des principes futurs de l'approche déterministe, et où l'AMC jouait un rôle fondamental (Beck, Mann et Morris, 1958) :

- la nécessité de vérifier la qualité non seulement de conception, mais aussi de réalisation et de conduite de l'installation ;
- l'intégration de systèmes de protection mutuellement indépendants (notamment différents confinements de la matière radioactive) ;
- enfin, en se plaçant sous l'hypothèse de l'accident maximal, la garantie que ses conséquences (en termes de doses reçues par une personne située à l'extérieur de la centrale) ne dépasseraient pas une certaine limite considérée acceptable.

Dans une autre contribution présentée à la même conférence, Farmer et deux de ses collègues de la UK AEA présentèrent cependant les prémisses d'une approche alternative (Farmer, Fletcher et Fry,

9. Selon les mots de François Cogné, spécialiste français de la sûreté et directeur de l'Institut de Protection et de Sûreté Nucléaire : « Devant la montée des demandes des spécialistes de la sûreté, les concepteurs et les réalisateurs, telles des grenouilles qui demandent un roi, ont été les premiers à demander de figer les règles, de "baliser le champ de mines" ; ils ont été servis, tout particulièrement aux États-Unis » (Cogné, 1984, p.21).

1958). L'objectif, dans leur esprit, était de « réduire les risques jusqu'à un niveau de probabilité assez bas pour être socialement et économiquement acceptable ». Il fallait donc s'intéresser au risque, ce qui excluait de sélectionner les événements sur la base de leur seule probabilité comme le supposait la méthode de l'AMC. Pour cela, il fallait se donner un modèle théorique de l'installation, fondée sur l'analyse conceptuelle des défaillances et sur les probabilités. Cette représentation devrait être validée ou amendée sur la base de l'expérience d'exploitation des centrales, et l'on pourrait notamment extrapoler la probabilité des défaillances importantes et rares à partir de l'observation de dysfonctionnements et d'erreurs mineures et plus fréquentes. Ceci supposait que soit créé un système de classification des événements importants pour la sûreté et que les données correspondantes soient recueillies dans toutes les centrales en fonctionnement. Les britanniques ont été les premiers à mettre effectivement sur pied un tel système en 1959. Outre qu'elle posait ainsi les fondements de l'approche probabiliste, la contribution de Farmer, Fletcher et Fry comportait un certain nombre de propositions qui firent date :

- définir un domaine de fonctionnement sûr et équiper le réacteur de systèmes de contrôle et d'instrumentation qui le ramènent à l'intérieur de ce domaine à chaque fois qu'il en sort ;
- agir à la fois sur la probabilité des accidents, définis par une excursion du réacteur hors de son domaine de fonctionnement sûr, et sur leurs conséquences pour le réacteur et pour les hommes ;
- concevoir l'enveloppe du combustible et celle du fluide réfrigérant comme deux barrières étanches successives empêchant la dispersion des radionucléides ;
- préférer au principe de sûreté intrinsèque, qui suppose une très grande fiabilité des systèmes de contrôle et de protection, celui de redondance.

On peut distinguer là des aspects importants de ce qui sera progressivement formalisé comme la méthode de la défense en profondeur : le premier niveau de la défense, les principes de prévention et de maîtrise, la méthode des barrières formalisée par la suite par les analystes français, et enfin le principe de défaillance unique.

Malgré cette critique en creux du concept d'AMC, celui-ci s'imposa durant les années 1958-1967 comme le fondement des normes de sûreté, particulièrement aux États-Unis. Farmer reprit l'attaque en 1964, lors de la troisième conférence internationale de Genève, puis lors d'un colloque sur la sûreté des réacteurs organisé en 1967 par l'Agence Internationale de l'Énergie Atomique, cette fois de façon plus frontale :

'There is no logical way of differentiating between "credible" and "incredible" accidents; the "incredible" is often made up of a combination of very ordinary events - for example, the breakdown or deterioration that occurs in normal plant and their measuring instruments; and the "credible" may actually be exceedingly improbable. The

logical way of dealing with this situation is to seek to assess the whole spectrum of risks in a quantity-related manner' (Farmer, 1967, p.304).

Le cœur de la critique de Farmer était donc l'absence d'une démarche systématique d'identification des scénarios accidentels fondés sur une représentation théorique des centrales. Les tenants de l'AMC justifiaient leur démarche en arguant que la prise en compte de l'accident maximal, que l'on considérait être la perte de réfrigérant, permettait de bénéficier par la même occasion d'une protection contre les accidents plus fréquents et moins graves. Farmer leur répondait que des événements plus fréquents et moins graves pourraient bien se combiner pour aboutir à un accident « incroyable ».

Dans le même article, Farmer représentait pour la première fois son célèbre diagramme, dans lequel il définissait la limite d'acceptabilité des risques sous forme d'une droite décroissante dans le plan (rejets, fréquence). Il fixait arbitrairement la pente de cette droite à $-1,5$, de sorte qu'un accroissement des rejets de deux ordres de grandeur devait être compensé par une réduction de la fréquence de trois ordres de grandeur. Il justifiait ce coefficient d'une part par la non linéarité de la relation entre rejets et conséquences sanitaires finales, et de l'autre par une attitude d'aversion pour le risque.

4.1.2.3 Le rattrapage français

À partir de la fin des années 1950, sous l'effet des deux conférences de Genève et de l'accident de Windscale¹⁰, le CEA décida de renforcer l'organisation de la sûreté de ses installations nucléaires. En 1960, il se dota d'une Commission de sûreté des installations atomiques (CSIA), qui rassemblait ses directeurs et s'appuyait sur des sous-commissions spécialisées. La plus importante de celles-ci était la sous-commission à la sûreté des piles (SCSP), placée sous la direction de Jean Bourgeois, pionnier de la sûreté nucléaire en France et également chef du Département d'étude des piles à partir de 1962. Bourgeois constitua progressivement un groupe d'experts en sûreté, le Groupe Technique à la Sûreté des Piles, qui allait forger la doctrine française de sûreté nucléaire.

Initialement très influencés par les conceptions américaines, les experts français prirent peu à peu leurs distances dans le sillage de leurs confrères britanniques, avant de se démarquer également de ces derniers pour élaborer leur propre doctrine. Celle-ci était caractérisée par un mélange de prudence et de pragmatisme.

Les experts français partageaient les réserves de Farmer à l'égard de la notion d'AMC, sans toutefois avoir de concept de substitution sur lequel s'appuyer. Ils se montraient également plus

10. Le 7 et le 8 octobre 1957, au cours d'une opération de routine dite « recuit Winger » dans la pile n°1 de la centrale de Windscale, une série d'erreurs d'interprétation et de manœuvre des opérateurs conduisirent à une augmentation incontrôlée de la température du cœur, qui prit feu. Il fallut presque quatre jours aux équipes de secours pour détecter le feu et le maîtriser en noyant le cœur à l'aide de lances à incendie. Un nuage radioactif se forma par évaporation de l'eau et, dans les jours suivants, des traces de retombées furent observées au Royaume-Uni, en Belgique, en Allemagne et en Norvège.

réservés que les britanniques à l'égard de la possibilité de construire des centrales nucléaires à proximité de zones fortement peuplées. Ils contribuèrent notamment au développement de la démarche déterministe de sûreté en formalisant la méthode dite des barrières (voir la sous-section 3.3.1), que Farmer, Fletch et Fry avaient évoquée dans leur article de 1958.

L'organisation des instances françaises en charge de la sûreté était caractéristique d'une méthode qualifiée plus tard de *French cooking*. L'objectif n'était pas d'édicter des règles, mais d'instaurer un « dialogue technique » afin de trouver des solutions pratiques aux problèmes de sûreté. Organe interne au CEA, le CSIA était certes constitué sur le mode d'un tribunal : tout projet de construction, de mise en route ou de modification d'une installation était soumis à son autorisation préalable. Cependant, son mode de fonctionnement était fondé sur la coopération entre responsables de la sûreté et du développement¹¹ et la recherche d'un compromis technique permettant « l'obtention dans le domaine des radiations d'une sûreté aussi parfaite que possible » (CEA (1960), cité par Foasso (2003, p.131) ; nous soulignons).

4.2 La politique de la sûreté nucléaire aux États-Unis

4.2.1 Crise et sortie de crise (1967-1986)

4.2.1.1 Un essoufflement institutionnel et doctrinal

Avec le développement de l'industrie nucléaire civile et la montée des préoccupations de sûreté, il est devenu patent que le double mandat des agences nucléaires les exposait à un conflit d'intérêts permanent. Cette contradiction interne était particulièrement forte dans le cas de l'AEC. À la fin des années 1960, la crédibilité de la Commission fut sérieusement entamée par une série de faux pas dans lesquels ses organes de décision choisirent délibérément de sous-estimer le risque qu'ils étaient censés contrôler.

Deux controverses

À la fin des années 1960, la branche « réglementaire » de la Commission était consciente que l'hypothèse d'accidents plus graves que l'AMC ne pouvait plus être ignorée, comme en témoigne la déclaration de Beck, directeur adjoint à la réglementation, placée en exergue à ce chapitre. Les préoccupations étaient d'autant plus fortes que les unités entrant alors en exploitation étaient d'une puissance considérablement plus élevée que celles qui avaient permis la mise au point de la défense en profondeur, et qu'elles se situaient plus près de zones densément peuplées. En raison de l'augmentation de la masse de combustible présente dans le cœur, on craignait alors qu'une fusion du cœur n'entraînât inévitablement le percement de la cuve, la défaillance de l'enceinte et des rejets

11. Ces deux figures allant jusqu'à se confondre dans une même personne dans le cas de Jean Bourgeois.

massifs. Pour la branche « industrielle » de la Commission, toutefois, l'hypothèse de fusion ne devait pas être prise en compte dans la réglementation.

En 1967, le rapport d'un comité d'experts réuni par l'AEC confirmait que dans un scénario d'accident de perte de réfrigérant primaire (APRP, équivalent de l'anglais *Loss of Coolant Accident* ou LOCA) doublé d'une défaillance du système de refroidissement de secours, la matière fissile ne pourrait être retenue à l'intérieur de l'enceinte (Walker, 2000)¹². En 1970-71, des tests réalisés sur le réacteur d'essais d'Idaho jetèrent un doute sérieux sur la capacité du système de refroidissement de secours (*Emergency Core Cooling System*, ECCS) à compenser les effets d'une perte de réfrigérant primaire et, plus fondamentalement encore, sur la validité de l'approche suivie dans l'analyse de la sûreté. En pleine période de développement de l'énergie nucléaire, le trouble de l'AEC fut tel qu'elle choisit de ne pas divulguer les résultats des tests pendant quelques temps. En 1971, après cinq années de travail, la Commission adopta un ensemble de critères destinés à couvrir tous les aspects de la conception des centrales nucléaires, les *General design criteria* portés en annexe du titre 10 partie 50 du code de réglementation fédérale (10CFR50). Parmi ceux-ci, l'un portait sur le système de refroidissement de secours. Il exigeait que l'ECCS puisse compenser les effets d'un accident de perte de réfrigérant primaire, aussi grave soit-il, et qu'il réponde au critère de défaillance unique.

Alors que le débat sur l'efficacité de l'ECCS, et donc sur la probabilité d'accidents graves, battait son plein, la Commission se trouva engagée dans une deuxième controverse majeure. La réglementation imposait alors aux installations nucléaires des seuils de rejets (en fonctionnement normal) sur la base de la limite d'exposition du public recommandée par la Commission Internationale de Protection Radiologique (0,5 rem par an) et de l'hypothèse extrêmement conservatrice d'une présence permanente à proximité de la centrale (24 heures par jour, 365 jours par an). De surcroît, les rejets effectifs des centrales ne dépassaient guère quelques pour-cent des seuils autorisés. Cela ne suffit pas à protéger la réglementation des interrogations sur l'effet de l'exposition à de faibles doses de substances cancérogènes, qui devint à cette époque l'une des questions les plus incertaines et les plus ouvertes à la controverse dans le domaine de l'analyse des risques sanitaires (et le reste à ce jour). En analysant les données épidémiologiques de populations exposées aux explosions nucléaires de Hiroshima et de Nagasaki d'une part, et à l'imagerie médicale par rayons X d'autres part, Gofman et Templin, deux scientifiques renommés¹³, montrèrent en 1969 qu'il existait une réponse positive même à de très faibles doses de rayonnements, et que ceux-ci avaient de ce fait un effet mutagène

12. C'est à cet époque qu'a été imaginée le scénario du « syndrome chinois » : un cœur de réacteur en fusion qui traverserait non seulement la cuve et le radier de l'enceinte de confinement, mais... la planète entière, pour refaire surface en Chine !

13. John W. Gofman était un chimiste et bio-physicien qui avait effectué sa thèse au début des années 1940 sous la direction de Glenn Seaborg, responsable de la découverte du plutonium, puis travaillé avec lui au sein du projet Manhattan. À la demande de Seaborg, devenu entre-temps directeur de l'AEC, il avait créé en 1963 un département de recherche biomédicale au Lawrence Livermore National Laboratory pour étudier les effets sanitaires des radiations ionisantes.

et cancérigène nettement plus important qu'on ne le pensait précédemment. S'appuyant sur ces résultats, Gofman et Templin estimaient que si l'ensemble de la population américaine était exposée aux seuils de rejets de l'AEC, il en résulterait 32 000 cas supplémentaires de cancers fatals par an. En dépit du caractère totalement irréaliste de cette hypothèse, le travail de Gofman et Templin eut un grand retentissement. L'AEC commença par défendre ses positions pendant deux ans, avant de les abandonner brusquement en juin 1971 en proposant des objectifs de conception des centrales qui conduisaient à une division par cent des niveaux de rejets admissibles. Or ce revirement, loin de convaincre les critiques qui lui reprochaient de se montrer trop conciliante avec les points de vue des industriels, lui aliéna en outre ces derniers, ainsi qu'un bon nombre d'experts de la radioprotection.

La fin d'une ambivalence ?

Le sentiment général à l'égard de l'énergie nucléaire, encore positif au milieu des années 1960 ¹⁴, l'était désormais beaucoup moins. Avec le débat sur les faibles doses, les centrales devenaient suspectes non plus seulement en raison des risques d'accident, mais même dans leurs conditions normales de fonctionnement. Individus et associations utilisaient de plus en plus fréquemment l'ensemble des dispositions que la législation américaine mettait à leur disposition pour contester les décisions publiques et arrêter les projets les projets industriels.

En 1970, la Cour d'appel du District de Columbia ordonna l'arrêt de la construction d'une centrale à Clavert Cliffs (Maryland) et condamna l'AEC pour non respect de la législation sur l'environnement. La révision de la procédure d'octroi de licences suite à cette décision entraîna un gel des permis de construction de centrales pendant dix-huit mois.

En 1972, en vue de préparer la révision du Price-Anderson Act, le sénateur Pastore, président du comité mixte du Congrès, demanda à l'AEC de lancer une analyse générale sur la sûreté des centrales nucléaires. Consciente de la caducité de son cadre méthodologique et de la nécessité d'examiner de plus près les scénarios d'accident de perte de réfrigérant primaire (LOCA), l'AEC saisit cette occasion pour confier à Norman Rasmussen, professeur au *Massachusetts Institute of Technology*, une étude d'un genre nouveau dont l'objectif était de parvenir à une quantification des risques d'accident.

Mais la faiblesse de la Commission constituait désormais un handicap pour l'énergie nucléaire civile. À l'occasion de la survenue du premier choc pétrolier, le Président Nixon demanda au Congrès la création d'une nouvelle agence qui pourrait se concentrer sur les aspects réglementaires et accélérer la délivrance de licences. En 1974, le Congrès adopta l'Energy Reorganization Act, qui démantelait l'AEC et instituait la Nuclear Regulatory Commission (NRC), avec pour charge les seules fonctions réglementaires de l'ancienne Commission.

14. Les premières manifestations publiques hostiles à l'énergie nucléaire semblent avoir eu lieu en 1965 en Californie, où il avait été décidé de construire des centrales dans des zones sismiques (Foasso, 2003, p.138).

Ces turbulences n'étaient pas spécifiques aux États-Unis, même si elles y étaient particulièrement visibles. Au Royaume-Uni, le gouvernement décida également de séparer ses activités de développement de l'énergie nucléaire et de celles de surveillance de la sûreté, qui furent placées sous la responsabilité du *Nuclear Installations Inspectorate* au sein du *Health and Safety Executive* en 1975. Des évolutions similaires avaient lieu en France à la même époque.

4.2.1.2 La naissance difficile de la méthode probabiliste

Le travail de Rasmussen

Après trois ans de travail de l'équipe rassemblée autour de Norman Rasmussen, l'étude WASH-1400 fut publiée en octobre 1975 par la nouvelle Commission (une version préliminaire en avait été mise en circulation un an plus tôt). Outre la méthodologie et les évaluations que nous avons décrites au chapitre précédent, elle comportait une comparaison de risques exprimés en termes d'espérance de perte de vie d'un individu moyen aux États-Unis, à la manière de l'article publié quelques années plus tôt par Chauncey Starr (Starr, 1969). Les accidents des centrales nucléaires y étaient mis en regard d'événements tels que les cyclones, les tremblements de terre, les chutes de météorite, les accidents de la circulation, les chutes d'avions, les incendies, les explosions, etc. Les estimations indiquaient que le risque lié au fonctionnement de cent centrales nucléaires était largement inférieur à ceux de l'automobile, du transport aérien ou des cyclones. Parmi les aléas considérés, seul la chute de météorites engendrait des risques comparables : une chance sur dix mille de faire dix morts, et une sur dix millions d'en faire dix mille. Ces conclusions étaient délibérément mises en avant, notamment dans la note de synthèse (*executive summary*) du document.

Ces conclusions valurent à l'étude de fortes critiques. Après s'être penché sur la version préliminaire du rapport, un groupe de scientifiques réunis par l'*American Physical Society* (APS) souligna notamment le peu de sens qu'il y avait à extrapoler ainsi des calculs où la prise en compte des conséquences sanitaires avait été limitée aux 24 premières heures d'exposition. En juin 1976, la Commission des affaires intérieures de la Chambre des représentants organisa des auditions au sujet de l'étude, au cours desquelles il fut estimé que ses conclusions, par l'impression de certitude et de complétude qu'elles laissaient, étaient abusives. Visiblement embarrassée par cette tournure, la NRC demanda une évaluation officielle à un groupe de scientifiques dirigés par Harold Lewis, qui avait aussi été le premier signataire du rapport de l'APS. En septembre 1978, le comité Lewis rendit son rapport. Sur le fond, celui-ci portait, comme nous l'avons vu, un regard nuancé sur WASH-1400 : s'il critiquait certains de ses parti pris, notamment dans le traitement des incertitudes ou l'inclusion d'hypothèses conservatrices dans des calculs réalistes, il saluait également une avancée très importante dans les techniques d'analyse des risques d'accident. Son désaccord portait surtout sur la présentation des résultats et les interprétations auxquelles celle-ci avait conduit :

"There have been instances in which WASH-1400 has been misused as a vehicle to judge the acceptability of reactor risks. In other cases it may have been used prematurely as an estimate of the absolute risk of reactor accidents without full realization of the wide band of uncertainties involved. Such use should be discouraged" (Lewis et al., 1978).

La critique portait également sur un autre aspect de l'exploitation des résultats de WASH-1400 : la comparaison de risques sans prise en compte des autres coûts et des bénéfices qui leur étaient associés. En notant que l'opinion publique formait son jugement sur l'énergie nucléaire en considérant non seulement les risques de pertes humaines liées aux accidents graves des réacteurs, mais aussi l'ensemble du cycle du combustible, les aspects économiques ou environnementaux, ainsi que les questions de prolifération, le rapport Lewis prenait pour cible ultime la notion d'acceptabilité du risque telle qu'elle avait été façonnée par les travaux de Farmer et Starr.

Une révolution méthodologique à entreprendre, des appréciations négatives (même si elles ne portaient pas sur le cœur du rapport), un contexte institutionnel et politique instable : il n'en fallait pas plus pour que la jeune NRC choisisse de se désolidariser des conclusions et même des estimations avancées par WASH-1400. Cette infortune aurait probablement condamné l'ensemble de l'approche probabiliste pour quelque temps si n'était survenu, à peine deux mois après la déclaration de la NRC, l'accident de Three Mile Island (TMI).

Le tournant de TMI

Bien qu'il n'ait pas donné lieu à des rejets significatifs, l'accident de TMI constitua un choc considérable aux États-Unis en montrant que les promesses de sûreté totale étaient effectivement intenables. La Commission d'enquête nommée par le Président des États-Unis critiqua sévèrement la NRC, en dénonçant dans sa politique une forme de présomption de sûreté :

"if the country wishes, for larger reasons, to confront the risks that are inherently associated with nuclear power, fundamental changes are necessary if those risks are to be kept within tolerable limits. [...] After many years of operation of nuclear power plants, with no evidence that any member of the general public has been hurt, the belief that nuclear power plants are sufficiently safe grew into a conviction. [...] The Commission is convinced that this attitude must be changed to one that says nuclear power is by its very nature potentially dangerous, and, therefore, one must continually question whether the safeguards already in place are sufficient to prevent major accidents" (Kemeny et al., 1979).

La Commission chercha également, dans ses recommandations à caractère institutionnel, à impliquer davantage la branche exécutive dans la supervision de l'agence de sûreté et dans les choix stratégiques en matière de sûreté, en demandant notamment que l'agence fasse un rapport annuel

au Président et que ses commissaires soient nommés par celui-ci, et en notant que la question plus vaste des choix énergétiques du pays et de son engagement dans la filière nucléaire devait sans doute être reposée à la lumière de l'accident :

"The Commission's findings with respect to the accident and the regulation of the nuclear industry – particularly the current and potential state of public safety in the presence of nuclear power – have, we believe, implications that bear on the broad question of energy. But the ultimate resolution of the question involves the kind of economic, environmental, and foreign policy considerations that can only be evaluated through the political process" (Kemeny et al., 1979).

De fait, TMI mit un terme au développement de l'industrie nucléaire civile aux États-Unis, qui n'a pas repris à ce jour.

Sur le plan de la méthode, TMI confirma à la fois les doutes déjà sérieux relatifs à la démarche déterministe et son absence de prise en compte de l'hypothèse d'accidents graves, et l'intérêt de l'approche systématique adoptée par Rasmussen et son équipe. À la NRC, où une équipe d'employés venait de passer deux mois à effacer toute mention de WASH-1400 dans les documents officiels, il fallu en consacrer deux de plus à faire machine arrière (Keller et Modarres, 2005, p.280). Les deux principales commissions d'enquête sur les causes de l'accident parvinrent à la conclusion qu'il serait désormais souhaitable de compléter la démarche de sûreté traditionnelle par des analyses probabilistes semblables à WASH-1400 (Kemeny et al., 1979; Rogovin et al., 1980). La seconde recommanda en outre que les accidents dus à des défaillances multiples soient prises en compte dans le dimensionnement des centrales et que des objectifs probabilistes inscrits dans la réglementation donnent une indication claire du niveau de sûreté qui était estimé « assez sûr » (*safe enough*) du point de vue de la puissance publique.

À la suite de l'accident, plusieurs opérateurs de centrales nucléaires entreprirent de leur propre chef des EPS afin, d'une part, de décider d'améliorations techniques de leurs installations et, de l'autre, de disposer d'une mesure du risque pour les populations avoisinantes. Les premières EPS post-TMI furent celles de la centrale de Zion (publiée en 1981) et des tranches 2 et 3 de la centrale d'Indian Point (publiée en 1982)¹⁵. Ces études furent menées pour une large partie par un cabinet de consultants, *Pickard, Lowe and Garrick*, qui joua un rôle important dans la formalisation de la méthodologie des EPS.

La NRC, de son côté, entama en 1979 une conversion progressive, mais jamais démentie, aux méthodes probabilistes. Elle passa commande à l'*American Nuclear Society* et l'*Institute of Electrical and Electronic Engineers* d'un guide de procédure sur les EPS qu'elle publia en 1983 (NRC, 1983). En 1986, elle lança une EPS de cinq centrales américaines qui visait à mettre WASH-1400

15. Il s'agissait des deux centrales les plus proches de grandes métropoles : celle de Zion, fermée depuis 1998, est située à 65 km de la ville de Chicago, et celle d'Indian Point à 50 km de New York.

à jour sur la base des dix années d'expérience opérationnelle et de données supplémentaires et de méthodes affinées. Le résultat, publié en 1990, constitua comme nous l'avons vu un nouvel état de l'art dans le domaine nucléaire, notamment en matière d'élicitation des opinions d'experts (NRC, 1990). En 1986, la NRC répondit aussi à la demande qui lui avait été faite par la commission Rogovin, et assigna aux opérateurs de centrales un objectif probabiliste qui définissait ce qu'était, de son point de vue, un niveau de risque radiologique acceptable (NRC, 1986). Rétrospectivement, cette réglementation peut être considérée comme le point de départ d'une nouvelle doctrine de la sûreté.

4.2.2 L'émergence d'une nouvelle doctrine

4.2.2.1 La conversion à la « décision informée au sujet du risque »

Les *Safety goals for the operation of nuclear power plants* consistaient en deux critères qualitatifs de santé publique, l'un collectif et l'autre individuel, et deux autres les traduisant en termes quantitatifs :

- le fonctionnement des centrales nucléaires ne devait pas engendrer de risque supplémentaire significatif pour la santé et la vie des individus ;
- les risques engendrés par le fonctionnement des centrales nucléaires pour la société dans son ensemble devaient être au plus comparables à ceux créés par toute technologie de génération d'électricité concurrente et viable ;
- le risque de mort rapide causée par un accident de réacteur pour un individu moyen vivant à proximité d'une centrale devait être au plus égal à 0,1 % de la somme des autres risques de mort rapide auxquels les individus américains sont généralement exposés ;
- le risque de cancers mortels causés par le fonctionnement d'une centrale pour la population vivant à sa proximité devait être au plus égal à 0,1 % de la somme des risques de cancers mortels liés à toute autre cause.

Ainsi spécifié, le critère individuel concernait les risques liés aux accidents, alors que le critère collectif portait sur l'ensemble du fonctionnement des centrales, en situation normale comme accidentelle. Le texte du règlement précisait bien que ces objectifs devaient être compris comme une *indication* du niveau de protection de la population que les concepteurs et opérateurs de centrales devaient chercher à atteindre. Comme l'expliquait le président de la NRC en 2001, l'intention n'avait été en aucune manière d'appliquer ces critères dans des décisions spécifiques (par exemple de licence d'exploitation), mais de signaler l'orientation générale de la politique de sûreté aux acteurs de l'industrie et au personnel de la Commission lui-même (Meserve, 2001).

En 1988, avec la *Generic Letter 88-20*, la NRC fit un pas supplémentaire dans la promotion des méthodes probabilistes en demandant aux opérateurs de conduire des examens individuels de

leurs centrales afin d'évaluer les risques d'accidents graves liés à des événements internes (*individual plant examination* ou IPE) ou externes (*individual plant examination of external events* ou IPEEE) (NRC, 1988b et 1988c). La lettre exigeait particulièrement des opérateurs qu'ils améliorent leur compréhension des probabilités d'endommagement du cœur et de relâchement de matière radioactive, et qu'ils agissent si nécessaire afin de réduire ces probabilités. Pour cela, elle leur recommandait de conduire une EPS de niveau 1 complétée par une analyse plus sommaire du comportement de l'enceinte de confinement conformes aux guides établis par la NRC, mais déclarait que sous réserve d'approbation, des méthodes alternatives pouvaient également être utilisées. Elle soulignait cependant que la réalisation d'une EPS pouvait par la suite simplifier certaines démarches réglementaires, et encourageait les opérateurs à y impliquer leurs propres personnels, de façon à les former aux raisonnements et outils probabilistes et à intégrer au mieux la connaissance spécifique qu'ils possédaient de leur centrale.

Le secteur électronucléaire américain avait largement anticipé cette appel. S'ils s'étaient montrés initialement très réservés à l'égard de l'approche probabiliste, en raison notamment du caractère problématique de la notion de risque en droit¹⁶, les industriels s'étaient rapidement convaincus des possibilités qu'elle offrait pour analyser la sûreté de leurs installations, mais aussi tenter d'en convaincre le public et négocier avec la puissance publique un allègement de ses exigences réglementaires¹⁷. Au début des années 1990, une grande majorité de centrales américaines avait fait l'objet d'une évaluation probabiliste spécifique, bien que selon des méthodes et avec des degrés d'approfondissement et une qualité d'ensemble variables.

Cependant, même destinés à un usage très général, les objectifs de sûreté s'avérèrent inapplicables en raison du degré d'incertitude affectant les mesures de risque. En 1990, la NRC fut amenée à adopter, pour son usage interne, des critères de substitution se référant directement aux grandeurs mesurées par les EPS : une probabilité de fusion du cœur (*core damage frequency*, CDF) inférieure à 10^{-4} et une probabilité de rejets rapides et importants (*large early release frequency*,

16. L'étude WASH-1400 et ses prolongements politiques venaient en effet perturber le cadre juridique dans lequel opérait l'industrie nucléaire civile, construit autour de l'association entre la notion de risque indu (*undue risk*) et le concept déterministe d'accident maximum crédible : les analyses de sûreté réglementaires attestaient que toutes les mesures nécessaires avaient été prises dès lors que la population était protégée en cas d'accident de gravité inférieure à celle de l'AMC ; ceux de gravité supérieure étaient exclus.

17. À titre d'exemple, le programme de recherches privées IDCOR (*Industry Degraded Core Rulemaking*) a été lancé dès 1981 pour étudier les questions techniques et réglementaires posées par les scénarios d'accident grave. En 1984, IDCOR était parvenu aux conclusions suivantes : la probabilité de survenue d'un accident grave dans un réacteur à eau légère est extrêmement faible ; dans une telle éventualité, le terme source serait vraisemblablement bien moindre que les études précédentes ne l'ont estimé ; les risques pour la population liés à des accidents graves sont très inférieurs aux niveaux calculés dans différentes études, et également inférieurs aux objectifs de sûreté (à l'époque préliminaires) de la NRC. Sur cette base, il était estimé que « des changements majeurs de conception ou de conduite des réacteurs ne sont pas nécessaires » (Buhl et al., 1987). Dans les années suivantes, IDCOR s'était donné pour objectif de défendre le point de vue de l'industrie auprès de la NRC, notamment en matière de tenue de l'enceinte de confinement et de termes-source, et de développer une approche intégrée pour l'analyse des risques d'accident grave dans chaque centrale. Cette approche, qui a fait l'objet de nombreuses discussions entre les responsables du programme et la NRC, faisait partie des méthodologies admises par la NRC pour la réalisation d'un examen individuel IPE.

LERF) inférieure à 10^{-5} .¹⁸

Dans les années suivantes, ce furent les industriels qui firent pression sur la Commission pour qu'elle adopte la démarche probabiliste comme fondement général de sa politique. En 1995, celle-ci publia une déclaration de politique en ce sens, où l'on pouvait notamment lire :

'The use of Probabilistic Risk Assessment technology should be increased in all regulatory matters to the extent supported by the state-of-the-art in PRA methods and data and in a manner that complements the NRC's deterministic approach and supports the NRC's traditional defense-in-depth philosophy' (NRC, 1995).

Dans la terminologie de la Commission, cette nouvelle politique n'était pas basée sur le risque (*risk-based*) mais informée à son sujet (*risk-informed*). Pour la mettre en pratique, la Commission adopta en 1998 une série de guides réglementaires expliquant comment l'information sur le risque devait être prise en compte dans certaines de ses activités de surveillance et de contrôle. Le guide 1.174 définissait ainsi les principes de la « décision informée au sujet du risque » dans le cas d'une modification de l'installation proposée par l'opérateur (NRC, 1998) :

1. le changement proposé doit se conformer aux réglementations existantes, sauf pour celles dont il demanderait explicitement à être exempté ;
2. le changement proposé doit être conforme à la philosophie de la défense en profondeur ;
3. le changement proposé doit maintenir des marges de sûreté suffisantes ;
4. si le changement proposé accroît le risque, cet accroissement doit être faible ;
5. l'impact du changement proposé devra faire l'objet d'un suivi par l'intermédiaire de stratégies de mesure de la performance.

Le quatrième principe constituait la véritable innovation du guide en matière de doctrine, puisque les demandes d'autorisation ne devaient pas être accompagnées, jusqu'à lors, d'une information sur le risque. Dans le guide, le risque y était compris, sauf cas exceptionnel¹⁹, comme le couple formé par les probabilités de fusion du cœur (CDF) et de rejets rapides et importants (LERF), et le quatrième principe était précisé sous la forme de critères quantitatifs :

- une variation de CDF (respectivement LERF) inférieure à 10^{-6} (resp. 10^{-7}) est acceptable et ne nécessite pas de disposer d'une estimation globale de cette probabilité elle-même ;

18. Dans les réacteurs de conception récente à cette époque, la probabilité conditionnelle de perte de confinement de l'enveloppe dans un scénario de fusion du cœur était estimée inférieure à 0,1. Le critère de rejets retenu est bien plus tolérant puisqu'il correspond, si la probabilité de l'ensemble des scénarios de fusion est inférieure à 10^{-4} , à une probabilité conditionnelle de perte de confinement qui puisse aller jusqu'à 0,5 pour les scénarios les plus défavorables (voir à ce propos NRC (1990)). En effet, dans de nombreux scénarios de fusion, la perte de confinement est peu probable.

19. Par exemple, une modification qui affecterait la tenue à long terme de l'enveloppe du confinement ne pourrait pas être évaluée à partir des métriques usuelles et devrait faire l'objet d'une analyse spécifique.

- une variation de CDF (resp. LERF) comprise entre 10^{-6} et 10^{-5} (resp. 10^{-7} et 10^{-6}) n'est acceptable que s'il peut être établi que la probabilité est elle-même inférieure à 10^{-4} (resp. 10^{-5}) ;
- une variation de CDF (resp. LERF) supérieure à 10^{-5} (resp. 10^{-6}) n'est pas acceptable.

Le guide précisait en outre que toute évaluation probabiliste devait faire l'objet d'une analyse d'incertitude d'autant plus complète que le couple (risque, variation proposée) était proche des frontières d'acceptabilité. Dans le cas d'une analyse complète, l'incertitude paramétrique devait être propagée en tenant compte des corrélations possibles entre paramètres, et les niveaux de risque retenus devaient correspondre aux valeurs moyennes des distributions engendrées par ces propagations ; concernant les incertitudes de modélisation et de complétude, il fallait montrer que la décision (au sens du principe 4) n'était pas sensible à des choix alternatifs de modèles et de périmètre de l'analyse, dans la mesure où de telles alternatives pouvaient être raisonnablement envisagées²⁰.

4.2.2.2 Des défis importants

La position de principe affichée en 1995, suivie d'un début de mise en pratique à travers les guides réglementaires de 1998, marquaient clairement un changement de doctrine de la NRC que l'on peut caractériser par trois positions : premièrement, même si elle continuait d'affirmer l'importance de la défense en profondeur et du maintien de marges de sûreté, elle indiquait que c'était désormais l'évaluation du risque qui orientait les considérations de sûreté ; deuxièmement, elle reposait sur une position plus claire que jamais vis-à-vis de la notion d'acceptabilité, en admettant que des modifications structurelles puissent délibérément entraîner une augmentation du risque, tant que celui-ci, majoré d'une marge d'incertitude suffisante, restait en-deçà du niveau considéré acceptable ; troisièmement, elle requérait que cette marge d'incertitude soit évaluée de façon aussi précise et complète que possible à chaque fois que cela serait nécessaire. La règle permettant de juger de cette nécessité reposait sur un principe normatif que l'on pourrait énoncer ainsi : *il y a lieu de rechercher un degré de confiance d'autant plus élevé vis-à-vis d'une évaluation de probabilités que les enjeux de la décision correspondante sont importants*²¹.

Suite à cette déclaration générale, la NRC entreprit de réviser l'ensemble de ses réglementations techniques, règles d'autorisation et procédures de surveillance afin de les fonder sur la prise en compte du risque. Ce chantier, d'une ampleur considérable, nécessita notamment de compléter et d'affiner les objectifs de sûreté qui portaient, comme nous l'avons vu, sur la prévention des accidents de fusion et des rejets massifs. Un nouveau programme de surveillance de la sûreté des réacteurs

²⁰. Ceci correspond aux bonnes pratiques en matière d'analyse de l'incertitude décrites dans la sous-section 3.4.2.3.

²¹. Hill (2009) a proposé un traitement formel et une axiomatisation d'un comportement de ce type face à une incertitude de second ordre.

(*Reactor oversight process*) établit ainsi une nouvelle partition de la sûreté en sept « pierres angulaires » (*cornerstones*) : les événements initiateurs, les systèmes de protection, l'intégrité des barrières, les plans d'urgence, la radioprotection des travailleurs, la radioprotection de la population et la sécurité contre les actes de malveillance. Dans chacun de ces domaines, un ou plusieurs indicateurs de performance (éventuellement composites), assorti d'objectifs détaillés et de règles d'inspection, fut été choisi. À titre d'exemple, l'un des indicateurs retenus dans le domaine des événements initiateurs était le nombre d'arrêts non planifiés au cours des quatre derniers trimestres, sur la base de 7 000 heures de fonctionnement en état critique, avec les critères suivants : contribution négligeable au risque ne nécessitant pas d'action supplémentaire pour moins de trois arrêts (zone verte), contribution limitée justifiant des investigations supplémentaires pour plus de trois et moins de six arrêts (zone blanche), contribution importante nécessitant une intervention pour plus de six et moins de 25 arrêts (zone jaune) et contribution inacceptable pour plus de 25 arrêts (zone rouge).

Un exemple des changements apportés aux réglementations techniques est celui de règles applicables au système d'injection de secours (règlement 10CFR50.46). La règle déterministe précédente exigeait de dimensionner le système pour une grosse brèche du circuit primaire cumulée à une perte des alimentations électriques, tout en le soumettant au critère de défaillance unique. La règle « informée au sujet du risque » consista à demander les cumuls conventionnels pour les brèches inférieures à une taille limite, ainsi qu'un dimensionnement selon des hypothèses réalistes pour les brèches de taille supérieure. Elle découlait des analyses des scénarios de perte de réfrigérant primaire, qui ont constamment indiqué, depuis WASH-1400, une contribution plus importante des brèches de petite taille, due à leur probabilité plus élevée. La réforme offrit en outre aux opérateurs la possibilité de choisir entre l'ancienne et la nouvelle règle.

Le changement de paradigme opéré par la NRC ne se déroula pas sans friction. Aux personnels de la Commission, il imposait une modification radicale de méthodes de travail. Plus fondamentalement, peut-être, il leur inspirait des réserves de nature technique, liées notamment à la portée limitée et l'incomplétude des EPS fournies par les opérateurs, la prise en compte très insuffisante des incertitudes et les problèmes de compatibilité entre les conceptions déterministes et probabilistes (Fleming, 2003).

En réponse aux deux premières préoccupations, la NRC adopta un guide réglementaire concernant la méthodologie des EPS (NRC, 2007) et élaborait un document de référence concernant le traitement des incertitudes (NRC, 2009). L'industrie, de son côté, développa un standard conforme aux prescriptions de la Commission (ASME/ANS, 2009).

Le dernier point soulevait la question délicate de la prééminence nécessaire de l'une des deux

approches, que Sorensen et al. (1999) avaient posée dès 1999²². Si la position officielle de la Commission consistait à affirmer son attachement à la défense en profondeur tout en plaidant pour une meilleure prise en compte des probabilités, sa conception de l'équilibre entre les deux approches semblait s'être déplacée en faveur de la seconde. Le passage suivant du guide 1.174 assignait à la défense en profondeur le rôle de palliatif aux insuffisances des analyses du risque, et envisageait la possibilité que son équilibre soit réévalué à la lumière de celles-ci :

'If a comprehensive risk analysis is done, it can provide insights into whether the extent of defense-in-depth (e.g., balance among core damage prevention, containment failure, and consequence mitigation) is appropriate to ensure protection of public health and safety. However, to address the unknown and unforeseen failure mechanisms or phenomena, traditional defense-in-depth considerations should be used or maintained' (NRC, 1998, p.11).

Cette position n'était toutefois pas assez clairement établie pour répondre à certaines questions pratiques. À titre d'exemple, le principe de préservation des barrières de confinement devait-il, malgré tout, l'emporter sur les considérations de contribution au risque dans un cas comme les fuites des tubes de générateurs de vapeur ? En 2011, la NRC s'engagea dans une clarification du sens qu'il fallait désormais accorder au concept de défense en profondeur, dans le guide 1.174 et par suite dans l'ensemble de sa réglementation. Cette réflexion se poursuivait à l'heure où ces lignes étaient écrites.

4.3 La politique de la sûreté nucléaire en France

4.3.1 Quelques aspects du tournant nucléaire de la France (1967-1983)

4.3.1.1 Un revirement radical

Les choix du gouvernement français ne témoignaient pas, jusqu'à la veille du premier choc pétrolier, d'une volonté stratégique de développement de l'énergie nucléaire.

Parallèlement à l'édification d'une filiale « française » de centrales UNGG par le CEA, le gouvernement avait confié à Électricité de France le soin d'entreprendre la construction de réacteurs à eau sous pression (REP) utilisant de l'uranium enrichi, en collaboration avec la société américaine Westinghouse. Cette filière alternative bénéficiait à la fois de l'expérience américaine et de l'augmentation des capacités françaises d'enrichissement de l'uranium. La première tranche de ce type, construite à Chooz, divergea en 1967²³. L'apparition d'une nouvelle technologie et d'un acteur de poids au sein de l'industrie nucléaire française n'entraîna pas seulement une concurrence

22. Voir la section 3.5.2 ci-dessus.

23. Le réacteur de Chooz A a été arrêté en 1991.

acharnée entre les deux filières et leurs promoteurs. Elle déstabilisa aussi le processus d'évaluation de la sûreté, en engendrant un conflit d'intérêts pour le CEA, qui concentrait en son sein les compétences françaises dans ce domaine. En 1967, le ministre de l'Industrie décida que l'examen de la centrale UNGG de Saint-Laurent-des-Eaux serait soumise non pas à la Commission de Sûreté des Installations Atomiques du CEA, mais à un groupe d'experts *ad hoc* pour la sûreté des réacteurs, constitué de représentants du CEA, d'ÉdF et des constructeurs. Cette instance fut rebaptisée Groupe permanent (GP) d'experts pour la sûreté des réacteurs en 1972²⁴.

Le gouvernement avait par ailleurs installé en 1955 une Commission de production d'électricité d'origine nucléaire (PEON) afin d'évaluer le coût de l'énergie nucléaire. Cette commission estima en 1967 que la filière à eau légère avait un avantage de coût de 20 % sur sa rivale. En 1969, le gouvernement décida de l'abandon de la filière technologique développée en France dans l'après guerre au profit des réacteurs à eau légère²⁵. Mais dans le contexte de l'époque, cette décision semblait surtout préfigurer la disparition de l'industrie nucléaire civile française, tant celle-ci paraissait condamnée par l'abondance du pétrole.

Le développement économique de l'après-guerre avait été caractérisé par une croissance démographique soutenue, une augmentation rapide du revenu par tête et l'accession d'une part importante de la population aux produits de grande consommation, en particulier l'automobile. Les grandes orientations économiques étaient dictées par l'administration centrale, dans le cadre établi par les Plans quinquennaux. Pour satisfaire les besoins énergétiques considérables du pays, le gouvernement français avait décidé, au début des années 1960, de privilégier les importations de pétrole au détriment de la production nationale de charbon. Le pétrole apparaissait alors comme une ressource surabondante, et la création de l'OPEP en 1960, un an après la mise en place de mesures protectionnistes dans ce secteur par les États-Unis, semblait surtout destinée à inonder les marchés européens et japonais par l'intermédiaire de prix toujours plus bas. Les gouvernements successifs ayant confirmé et approfondi ce choix stratégique, la France avait basculé, à la veille du choc pétrolier de 1973, dans le « tout-pétrole »²⁶ (Jeanneney, 1981).

Le désarroi provoqué par le quadruplement des prix du pétrole et le nouveau contexte géopolitique mondial expliquent la soudaineté du retournement opéré par le gouvernement français en 1973-74. En 1973, quelques mois avant la guerre du Kippour et le choc pétrolier, celui-ci avait déjà décidé de porter de 8 à 13 GWe le programme de construction de centrales prévu au VI^e Plan.

24. Outre le « GP réacteurs », deux autres GP furent alors créés, le premier pour les accélérateurs de particules et le second pour les autres installations nucléaires de base. Ces instances continuent d'exister et sont aujourd'hui au nombre de sept (réacteurs, déchets, laboratoires et usines, transports, équipements sous pression, radioprotection industrie et recherche, radioprotection médicale et médico-légale).

25. À la suite de cette décision, Électricité de France passa commande pour les deux types de réacteurs à eau légère exploités aux États-Unis : les réacteurs à eau sous pression, sous licence de Westinghouse, et les réacteurs à eau bouillante, sous licence de General Electric. En 1975, elle opta définitivement pour le premier.

26. Plus précisément, la consommation énergétique nationale était alors satisfaite à 78% par les importations, et aux deux tiers par celles de produits pétroliers.

En 1974, en l'espace de quelques semaines, un plan (dit « plan Messmer », du nom du Premier Ministre) fut mis sur pied et adopté, consistant à démarrer la construction de toutes les tranches planifiées avant la fin de 1975, et à engager 50 GWe supplémentaires sur la période 1974-1980.

Au total, en trois ans allant septembre 1971 à juillet 1974, six réacteurs d'une puissance de 900 MWe avaient été mis en construction dans les centrales de Fessenheim et du Bugey²⁷. Les années suivantes ont vu ce rythme s'intensifier de façon effrénée. Entre novembre 1974 et avril 1976, en l'espace d'une année et demi, douze nouveaux réacteurs furent mis en chantier : quatre au Tricastin, quatre à Dampierre et quatre à Gravelines ; le contrat pluriannuel 1 (CP1) fut complété entre 1977 et 1979 par les quatre réacteurs du Blayais et deux supplémentaires à Gravelines. En 1976, le contrat pluriannuel 2 (CP2) démarra avec la construction de deux réacteurs à Saint-Laurent, suivis de quatre à Cruas (1978-79) et quatre à Chinon (deux en 1977, un en 1980, un en 1981). En 1977, une nouvelle génération de REP d'une puissance de 1300 MWe fut introduite, dont huit tranches furent mises en chantier de 1977 à 1980 (quatre à Paluel, deux à Flamanville et deux à Saint-Alban, dites de type P4) et cinq autres de 1979 à 1981 (deux à Cattenom, deux à Belleville et une à Nogent-sur-Seine, dites de type P'4). Au total, en une décennie, la France avait procédé à la construction de 47 réacteurs pour une puissance totale de 47,5 GWe.²⁸

Dans cette réorientation stratégique, un grand nombre de décisions ont été prises par un groupe d'acteurs et d'experts de l'industrie nucléaire, comme le disait sans ambages un rapport parlementaire en 1977 :

« On sait que toute la politique nucléaire française est élaborée et proposée par la commission de production d'électricité d'origine nucléaire (commission dite PEON). Or, cette commission est composée pour une large part par les représentants d'ÉdF et du CEA ainsi que par les représentants des industriels intéressés à la réalisation du programme. Cette composition en elle-même fait problème. On n'imagine pas que la politique des constructions scolaires soit pour l'essentiel élaborée par les entreprises du bâtiment. On peut être assuré que les personnalités de grande capacité et de haute qualité qui composent la commission PEON savent, autant que d'autres, faire prévaloir ce qu'elles considèrent comme étant l'intérêt national. Mais leur formation comme leurs choix professionnels donnent à penser qu'ils examinent davantage les possibilités du développement du nucléaire que les orientations à donner à notre politique énergétique. Il en résulte que la place des différentes formes d'énergie se trouve prédéterminée, en quelque sorte sans débat au fond, dès lors que la commission a arrêté ses choix sur

27. Ces réacteurs du groupe dit CP0 (contrat pluriannuel 0) ont été mis en service entre janvier 1978 et janvier 1980.

28. Les orientations de la politique énergétique ayant été confirmées par le gouvernement issu des élections de 1981, sept nouvelles tranches P'4 furent construites entre 1982 et 1984 (une à Nogent, deux à Penly, deux à Golfech, deux à Cattenom), suivies de quatre tranches de 1400 MWe entre 1984 et 1991 (deux à Chooz et deux à Civaux, dites de type N4).

le déroulement du programme nucléaire » (Commission des Finances, de l'Économie Générale et du Plan, 1977).

4.3.1.2 Une politique de sûreté largement « déléguée »

Ce qui était vrai du processus de décision concernant la politique énergétique en général l'était *a fortiori* des décisions relatives à la sûreté.

Certes, des décisions importantes avaient été prises afin d'adapter l'organisation institutionnelle de la sûreté nucléaire à la montée en charge du secteur. Il en était découlé une architecture et des principes d'ensemble qui ont été conservés depuis, même au cours des réformes importantes des années 2000. Les fabricants (Framatome et Alsthom) et l'exploitant (ÉdF) devaient répondre de la sûreté de leurs installations et montrer, en plaçant leurs dossiers, que les exigences réglementaires étaient satisfaites. Une autorité de sûreté tranchait : il s'agissait du Service central de sûreté des installations nucléaires (SCSIN) du Ministère de l'Industrie, créé en 1973. Le processus de séparation des responsabilités entamé en 1967 avait ainsi été poursuivi : désormais officiellement séparé du CEA, le SCSIN pouvait revendiquer une position d'arbitre impartial. Il s'appuyait sur un pôle unique d'expertise publique, l'Institut de protection et de sûreté nucléaire (IPSN), créé en 1976 par le regroupement des services du CEA en charge de la sûreté et de la radioprotection. L'IPSN couvrait ainsi l'ensemble des risques liés aux activités nucléaires. Il menait à la fois des activités d'évaluation et d'essais et, en demeurant rattaché au CEA, restait en contact avec la recherche et développement en matière nucléaire. Enfin, pour les décisions importantes, le SCSIN consultait les Groupes permanents (GP), où se prolongeait le « dialogue technique » à la française.

Dans le même temps, toutefois, cette organisation institutionnelle privilégiait le traitement des questions de sûreté sous un angle exclusivement technique. En témoignait l'ampleur des responsabilités du SCSIN, qui se trouvait en pratique en charge de l'ensemble de la politique de sûreté nucléaire :

« Ce service, principalement responsable de l'étude, de la définition et de la mise en œuvre de la politique en matière de sûreté nucléaire ainsi que des problèmes qui s'y rattachent, regroupe les moyens du ministère de l'industrie, du commerce et de l'artisanat en matière de sûreté nucléaire.

Il est notamment chargé :

De préparer et de mettre en œuvre toutes actions techniques du département relatives à la sûreté des installations nucléaires et en particulier : élaborer la réglementation technique concernant la sûreté des installations nucléaires et suivre son application ; organiser et animer l'inspection de ces installations, sans préjudice d'application des dispositions de la loi du 30 décembre 1996 dans ce domaine ;

D'examiner pour avis les programmes du commissariat à l'énergie atomique qui s'y rapportent ainsi que les propositions budgétaires correspondantes et suivre l'exécution des programmes ;

De suivre, le cas échéant, les travaux de recherche et développement des autres établissements publics relevant du département dans le domaine de la sûreté nucléaire ;

De recueillir toutes informations utiles sur les problèmes de sûreté nucléaire et les mesures prises en ce domaine en France et à l'étranger ;

De proposer et d'organiser l'information du public sur les problèmes se rapportant à la sûreté ;

Et d'une façon générale, d'examiner les mesures propres à assurer la sûreté des installations nucléaires, notamment les mesures proposées dans ce domaine par le commissariat à l'énergie atomique en application de l'article 2 du décret du 29 septembre 1970 susvisé, et promouvoir leur mise en œuvre.

Le service central de sûreté des installations nucléaires doit être à tout moment en mesure de faire la synthèse des problèmes se rapportant à la sûreté des installations nucléaires. Il prépare et propose, en ce domaine, les positions françaises dans les discussions avec les gouvernements ou les administrations des pays étrangers et organise sur le plan français les modalités pratiques de ces discussions »²⁹.

Le SCSIN était certes placé sous l'autorité du ministre de l'Industrie, appuyé par un Conseil supérieur de la sûreté nucléaire (CSSN), puis, à partir de 1975, sous celle du Comité interministériel de la sécurité nucléaire (CISN). La mission officielle du CSSN était d'analyser et d'éclairer l'action gouvernementale en matière de sûreté nucléaire, et celle du CISN était de coordonner les actions du gouvernement dans le champ plus large de la sécurité nucléaire³⁰. En pratique, toutefois, l'objectif des deux instances, comme celle du Ministère (par ailleurs tutelle de l'exploitant ÉdF) était en grande partie de ne pas entraver ou retarder le développement du secteur.

Le traitement de la question de la localisation des nouvelles centrales au cours des années 1976-78 est symptomatique de cette orientation (Foasso, 2003, pp.437-44). À partir du lancement du plan Messmer, la question avait progressivement gagné en acuité, et la position affirmée dans les années 1960, qui consistait à éloigner systématiquement les centrales des zones à forte densité de population, devenait de plus en plus problématique. En 1976, une grande partie des instances en charge de la sûreté, en particulier le SCSIN et certains membres du CISN, souhaitaient l'adoption de règles claires d'acceptabilité des sites. La Direction de la sûreté nucléaire du CEA a alors mis au point un critère de sélection qui s'inspirait de l'approche de Farmer et préfigurait un troisième

29. Décret n° 73-278 du 13 mars 1973 portant création d'un conseil supérieur de la sûreté nucléaire et d'un service central de sûreté des installations nucléaires au ministère du développement industriel et scientifique, article 5.

30. Outre la sûreté, la sécurité nucléaire recouvrait la protection contre les radiations ionisantes, la prévention des actes de malveillance et le contrôle des matières fissiles depuis la production jusqu'à l'état de déchets.

niveau d'évaluation probabiliste de la sûreté, dont la méthodologie venait tout juste d'être proposée par l'équipe de Rasmussen aux États-Unis. Il s'agissait d'un découpage de la région délimitée par un périmètre de 50 kms autour d'un site en secteurs géographiques, au sein desquels le nombre d'habitants était pondéré par un coefficient de transfert atmosphérique calculé à partir de données météorologiques. Un classement de 18 sites existants et potentiels avait été réalisé à partir de cette méthode. Cependant, son adoption officielle comme critère de sélection des sites aurait nécessité de reconnaître la possibilité d'accidents entraînant des rejets importants, ce qui aurait risqué de provoquer des réactions d'hostilité et de retarder la construction des centrales. Il fut donc décidé de ne pas donner suite à cette demande des services en charge de la sûreté, et de laisser à Édf une plus grande latitude dans les choix de localisation.

D'avantage qu'une délégation de l'autorité publique, le transfert des décisions en matière de sûreté à l'administration résultait donc de leur subordination à un principe supérieur : la réalisation du programme électronucléaire français. Nous allons voir à présent que ce fractionnement de la décision publique était favorisé par une doctrine de sûreté centrée sur l'approche déterministe, qu'il renforça à son tour.

4.3.1.3 Réticences à l'égard de la notion de risque

La première génération de REP ayant été construite sous licence américaine, les principes de sûreté de ces centrales ont naturellement été inspirés de la norme ANSI 18.2 alors en vigueur aux États-Unis (ANSI, 1973). Celle-ci établissait notamment quatre catégories d'états de la centrale et une correspondance entre la fréquence estimée de ces états et leurs conséquences maximales admissibles (voir la table 3.2 en page 72). L'opérateur Électricité de France proposa en 1975 aux autorités de sûreté de reprendre cette classification, et reçut leur assentiment. La règle s'appliqua initialement aux réacteurs de 900 et 1300 MWe, et fut par la suite étendue au « palier » 1400 MWe, sans toutefois jamais acquérir de valeur réglementaire.

Il faut rappeler ici que les limites ainsi posées à la fréquence d'événements initiateurs faisaient partie des dispositifs de la défense en profondeur, puisqu'elles permettaient de mesurer et de contrôler la qualité de conception et de conduite de l'installation. Pour les catégories initiales de conditions de fonctionnement, les conséquences considérées étaient des notions formelles, et ne pouvaient être mises en rapport avec des conséquences effectives qui eussent été inacceptables indépendamment de la probabilité associée. Lorsque la possibilité d'accidents graves fut reconnue et que celles-ci durent être considérées, l'appréciation de l'acceptabilité se reporta sur le risque résiduel engendré par l'installation, conduisant, comme nous l'avons vu, à l'adoption en 1977-78 d'un objectif probabiliste global de survenue de « conséquences inacceptables ». Cet objectif fut fixé par le Service Central de Sûreté des Installations Nucléaires à travers deux lettres adressées à Électricité de France. La

première de ces lettres énonçait les intentions du régulateur :

« D'une façon générale, le dimensionnement des installations d'une tranche comportant un réacteur à eau pressurisée devrait être tel que la probabilité globale que cette tranche puisse être à l'origine de conséquences inacceptables ne dépasse pas 10^{-6} par an.

Dès lors, lorsqu'une approche probabiliste sera utilisée pour apprécier si une famille d'événements doit être prise en compte pour le dimensionnement d'une telle tranche, il conviendra de considérer que cette famille d'événements doit effectivement être prise en compte si la probabilité qu'elle puisse conduire à des conséquences inacceptables est supérieure à 10^{-7} par an, cette valeur ne pouvant être dépassée, pour la famille d'événements examinée, que s'il est possible de démontrer que les calculs de probabilités effectués sont suffisamment pessimistes. [...]

Il apparaît par ailleurs nécessaire qu'Électricité de France poursuive des efforts appropriés visant à permettre dans les meilleurs délais l'utilisation d'approches probabilistes pour le plus grand nombre possible de familles d'événements » (SCSIN, 1977).

D'après des acteurs avec lesquels s'est entretenu Foasso (2003, p.379), l'objectif général de probabilité dérivait d'un raisonnement à l'échelle de l'ensemble du programme français : maintenir à moins de 0,01 la probabilité d'un accident grave sur un parc de cinquante réacteurs d'une durée de vie de quarante ans chacun (soit un total de 2 000 années réacteur). Cependant, sachant que ni les autorités publiques ni l'opérateur n'envisageaient à cette époque d'engager d'EPS générale sur le modèle de l'étude WASH-1400, l'affichage d'un objectif probabiliste global ne s'apparentait guère qu'à une déclaration de principes. L'objectif par famille ne semblait guère plus contraignant, neutralisé comme il l'était par l'ajout d'une possibilité de dépassement exprimée en termes peu précis. En outre, les « conséquences inacceptables » qu'il s'agissait d'éviter ne furent jamais définies par un texte officiel³¹.

La véritable avancée sur le plan de la doctrine était d'ordre méthodologique : les autorités en charge de la sûreté invitaient l'opérateur à adopter les méthodes probabilistes, et affirmaient que certains accidents (dont une première liste était annexée à la lettre) ne pouvaient plus être considérés impossibles et devaient devenir des objets d'étude. Par ce biais, elles prenaient leurs distances avec la notion d'accident de dimensionnement, version française de l'AMC, et semblaient s'orienter vers la définition d'un niveau général de risque que la société accepterait en contrepartie des avantages liés à la technologie nucléaire, comme les y avait invités Farmer une décennie plus tôt et avec une dizaine d'années d'avance par rapport à leurs homologues américains.

31. Libmann (1996, p.163) explique que cela permit par la suite de donner une interprétation spécifique de cette notion à chaque fois qu'une famille d'accidents était étudiée par des méthodes probabilistes, et de faire évoluer cette interprétation dans le temps. Il n'en demeure pas moins que la portée initiale du texte ne pouvait être que limitée par une telle ambiguïté.

Si l'opérateur ÉdF répondit favorablement à la demande de prise en compte d'un ensemble de « conditions de fonctionnement complémentaires » par le biais de méthodes fiabilistes, il se montra en revanche fermement opposé à l'introduction d'objectifs probabilistes dans la réglementation. La réaction d'ÉdF amena le Service Central de Sûreté des Installations Nucléaires à préciser ses intentions, et sans doute aussi à les revoir à la baisse, dans une seconde lettre d'orientation :

« Les termes de ma lettre SIN n° 1076/77 n'impliquent pas [...] que la démonstration de la sûreté d'une tranche comportant un réacteur nucléaire à eau pressurisée doive être aujourd'hui apportée par une analyse probabiliste exhaustive. Par contre, l'utilisation d'approches probabilistes devrait permettre de mieux justifier voire d'améliorer la définition et le classement des situations prises en compte dans le dimensionnement d'une telle tranche » (SCSIN, 1978).

Le *modus vivendi* trouvé à l'issue de cet échange épistolaire pouvait être ainsi résumé : l'approche probabiliste serait cantonnée à un rôle d'appui à une démarche générale déterministe. Les autorités de sûreté renoncèrent à inscrire dans la réglementation un objectif probabiliste de sûreté, fut-il indicatif, et définir par ce biais un niveau de risque socialement acceptable. Deux raisons de fond peuvent être avancées pour expliquer ce choix.

En premier lieu, les méthodes probabilistes n'étaient sans doute pas suffisamment robustes pour pouvoir jouer le rôle que leurs partisans voulaient leur voir tenir. Comme nous l'avons vu, l'étude WASH-1400 avait été fortement critiquée à la fois pour son recours aux probabilités subjectives et pour l'ampleur des incertitudes qui affectaient ses estimations. Une idée commune, notamment parmi les industriels, était alors que les seuils de probabilité affichés étaient largement arbitraires et ne pouvaient constituer une base valide pour la conception des centrales nucléaires ou l'évaluation de leur niveau général de sûreté. L'ampleur des progrès méthodologiques réalisés dans ces deux directions (l'intégration des croyances d'experts et la prise en compte des incertitudes) depuis les premières EPS montre *a posteriori* que cette critique avait une certaine validité.

En second lieu, le concept de risque était lui-même considéré avec un certain scepticisme par les décideurs, en particulier dans le cas d'événements tels qu'un accident nucléaire. Si l'école française avait pris ses distances vis-à-vis d'une approche exclusivement déterministe de la sûreté, celle-ci gardait, aux yeux des décideurs et d'une partie du public, des avantages estimables. Comme l'expliquaient deux experts de la sûreté nucléaire et promoteurs de l'approche probabiliste à propos de l'attitude des pouvoirs publics :

« on préfère s'en remettre au jugement des organismes chargés d'examiner la sûreté et qui doivent répondre par oui ou par non à la question "l'installation proposée est-elle sûre". Toute tentative de définir un seuil d'acceptabilité de risques fait craindre des réactions, éventuellement brutales, du public concerné. [...]

Nous avons sérieusement sous-estimé un aspect spécifique de la sûreté des réacteurs nucléaires, lorsque les autorités réglementaires et le public sont concernés : c'est le désir permanent de définir une limite supérieure pour les pires conséquences de tout accident "possible". En d'autres termes, nous ne parvenons pas à nous débarrasser du vieux concept de l'accident maximal vraisemblable, pour lequel des dispositions sont prévues afin de contenir la majeure partie des rejets de radioactivité » (Tanguy et Cogné (1978) cités par Foasso (2003, p.384)).

Il est à noter, en outre, que le programme nucléaire français battait son plein en 1977 avec la construction d'une vingtaine de REP de 900 MWe des contrats pluriannuels 0, 1 et 2 et la mise en chantier de plusieurs réacteurs de 1300 MWe. Dans un tel contexte d'effort financier et de prise de risque, l'appétit des décideurs, tant parmi les autorités publiques que chez l'opérateur, pour un changement de paradigme réglementaire devait être aussi limité que les ressources qu'ils désiraient consacrer aux nouvelles méthodes d'analyse de la sûreté.

Il existait donc une synergie entre le mode de raisonnement déterministe et la priorité accordée au développement industriel. De ce fait, et malgré l'enthousiasme des experts français pour les nouvelles méthodes probabilistes, la première EPS française ne fut entreprise que six ans plus tard, en 1983, avec un retard important par rapport aux États-Unis, à l'Allemagne ou au Royaume-Uni.

Si la décision de 1978 peut donc être comprise dans son contexte historique, il est en revanche plus surprenant qu'elle soit devenue, et reste, la pierre angulaire de la doctrine de sûreté nucléaire en France.

4.3.2 Changements institutionnels, continuité doctrinale

Le cadre légal et institutionnel hérité des années 1970 a été largement rénové au cours des années récentes ; la doctrine de la sûreté, elle, n'a pas connu de véritable bouleversement.

4.3.2.1 Un cadre institutionnel rénové

Le contexte politique et, par suite, institutionnel, de la sûreté nucléaire en France connut d'importantes évolutions après la catastrophe de Tchernobyl.

Sur le plan de la sûreté proprement dite, l'un des principaux enseignements tirés de l'accident, au-delà des spécificités du réacteur et du contexte dans lequel il était exploité, était l'« importance de questions organisationnelles telles que la nécessité d'un régime réglementaire efficace et d'une culture de sûreté » (AIEA, 1996b). Un plus grand accent fut mis sur l'importance, du point de vue de l'efficacité de la défense en profondeur, du respect scrupuleux des procédures par le personnel, de la régularité et la qualité des inspections et de la priorité accordée au maintien de la sûreté par les opérateurs et les régulateurs, tous éléments jusque là en marge des approches formelles, que l'on

convint de désigner par le concept général de « culture de sûreté ».

En France, de surcroît, la gestion calamiteuse de la question des retombées radioactives par les autorités sanitaires mit en lumière les nécessités et périls de la communication en temps de crise, ainsi que les vertus d'une plus grande indépendance des autorités en charge de l'évaluation des risques. La pression en faveur d'une réforme des institutions s'accrut les années suivantes et déborda le champ du nucléaire, avec la survenue d'une succession de crises dans différents domaines de la santé publique : scandale du sang contaminé, crise de la vache folle, controverses au sujet des organismes génétiquement modifiés, épidémies.

À partir de 1987, le Parlement prit une part plus active à l'élaboration de la politique de sûreté nucléaire, en particulier en termes d'organisation institutionnelle. Suite à un premier rapport sur les conséquences du désastre de Tchernobyl, l'Office parlementaire des choix scientifiques et technologiques se saisit de la question de la sûreté nucléaire en 1990, en demandant au député Claude Birraux d'établir un rapport annuel à ce sujet. Au total, entre 1987 et 2001, l'Office rendit seize rapports ayant trait au secteur nucléaire, dont dix s'intéressaient spécifiquement à la sûreté. Dans l'ensemble, ces travaux exercèrent une pression constante en faveur d'une transparence accrue de l'information et d'une plus grande indépendance des instances publiques en charge de l'expertise et du contrôle de la sûreté nucléaire.

L'évolution institutionnelle eut lieu par étapes à partir du début des années 1990, dans le prolongement des tendances qui s'étaient dessinées au cours des années 1973-76.

En 1991, le SCSIN fut remplacé par la Direction de la sûreté des installations nucléaires (DSIN), placée sous la double tutelle du Ministère de l'Industrie et celui de l'Environnement, ce qui, en pratique, lui permettait de jouir d'une plus grande indépendance. L'autorité de la DSIN se renforça encore en 2002, avec sa transformation en Direction générale et l'élargissement de son champ de compétence à la radioprotection. La nouvelle Direction générale de la sûreté nucléaire et de la radioprotection (DGSNR) fut placée, outre ses tutelles précédentes, sous celle du Ministère de la Santé. Enfin, la loi 13 juin 2006 relative à la transparence et à la sécurité en matière nucléaire (dite loi TSN) créa, en place de la DGSNR, une Autorité de sûreté nucléaire, dotée du statut (informel en droit) d'autorité administrative indépendante. L'ASN est dirigée par un collège de cinq Commissaires, nommés pour un mandat non reconductible de six ans et irrévocables (sauf manquement grave). Elle est en charge de la réglementation, du contrôle et de l'information du public dans l'ensemble des champs de la sûreté nucléaire et de la radioprotection.

L'IPSN, de son côté, connut des réformes internes importantes dès 1990 afin de renforcer son autonomie vis-à-vis du CEA, avant d'être définitivement séparé de celui-ci par la loi du 1er juillet 1998. Issu de la fusion de l'IPSN avec l'Office de protection contre les rayonnements ionisants (OPRI) et doté d'un statut d'établissement public industriel et commercial, l'Institut de radioprotection et

de sûreté nucléaire (IRSN) est désormais l'unique organisme public en charge de l'expertise sur les risques nucléaires et radiologiques, soutien scientifique et technique de l'Autorité de sûreté nucléaire.

Même si la loi TSN réaffirme la primauté de l'État dans l'élaboration de la politique de sûreté (« L'État définit la réglementation en matière de sécurité nucléaire et met en œuvre les contrôles visant à l'application de cette réglementation »), il apparaît clairement qu'avec la création de l'ASN et de l'IRSN et l'évolution du statut des entreprises publiques Édf et Areva, les trois pôles qui ont contribué à la formulation de cette politique depuis l'origine ont gagné une large autonomie à son égard.

4.3.2.2 Principaux éléments de la doctrine française

Le fondement de la doctrine de sûreté française est resté, depuis l'origine, l'approche déterministe qui consiste principalement à appliquer à toutes les grandeurs d'intérêt des marges de sûreté estimées suffisantes et à organiser une défense en profondeur à l'aide de dispositifs techniques et humains. L'amélioration continue de cette défense par l'analyse, la recherche et le retour d'expérience fait l'objet d'une attention particulière, et il est estimé qu'aucun niveau de risque n'est acceptable en soi. Il est ainsi considéré que les réacteurs existants doivent faire l'objet de réévaluations de sûreté, qui permettent d'intégrer à leurs défenses des connaissances nouvellement acquises et de rapprocher leurs niveaux de sûreté de celles de centrales de conception plus récente. Ces réévaluations sont le plus souvent mises en œuvre à l'occasion des visites décennales de sûreté, qui entraînent un arrêt prolongé des réacteurs.

Afin d'alimenter le retour d'expérience, un effort important de collecte de données est réalisé dans l'ensemble des installations françaises (comme d'ailleurs dans la plupart des réacteurs au monde). Tout événement significatif du point de vue de la sûreté doit ainsi être identifié, répertorié et signalé à l'autorité de sûreté. Ceci permet à la fois d'assurer un suivi permanent des événements et des mesures correctives engagées par les exploitants, et de reconsidérer les hypothèses sur lesquels les défenses sont basées.

Le traitement des scénarios de rupture de tubes des générateurs de vapeur illustre les révisions qui peuvent ainsi découler du retour d'expérience. Cet accident, qui peut conduire à la fois au relâchement direct de vapeur primaire par les soupapes de sûreté du circuit secondaire³² et au dénoyage du cœur, était classé jusqu'à la fin des années 1980 en condition de fonctionnement de quatrième catégorie, c'est-à-dire que sa probabilité était estimée être dans une fourchette allant de 10^{-6} à 10^{-4} . Or l'observation de six ruptures dans le monde, pour 1600 années-réacteur de fonctionnement, amena en 1989 à réévaluer cette probabilité et à requalifier l'accident en condition de fonctionnement de troisième catégorie (intervalle de probabilité de 10^{-4} à 10^{-2}). Ce changement

32. C'est en cela que les tubes des générateurs de vapeur constituent une exception au principe d'existence de trois barrières distinctes de confinement.

conduit à diminuer les conséquences estimées acceptables pour cet événement, et donc à prendre des dispositions afin de limiter ses conséquences effectives.

Un second exemple important est celui du retour d'expérience de l'inondation du Blayais, que nous évoquerons dans le chapitre 6.

Il est intéressant d'essayer de cerner précisément la place des méthodes probabilistes dans cette architecture, comme nous l'avons fait dans le cas des États-Unis.

La première EPS française de niveau 1 fut donc conduite de 1983 à 1990 par l'IPSN, en collaboration avec la société Framatome, pour les réacteurs de la tranche 900 MWe. En application du principe de double-expertise, l'étude préliminaire fit l'objet d'une revue par Édf, dont les remarques furent prises en compte dans la phase suivante. Une seconde EPS de niveau 1 fut menée de 1986 à 1990 par Édf pour les réacteurs de la tranche 1300 MWe, et c'est l'IPSN qui en fit la revue externe. En 1996, l'IPSN lança une EPS de niveau 2 pour les réacteurs 900 MWe, dont les premiers résultats furent publiés en 2003. L'ensemble de ces travaux a fait l'objet de mises à jour régulières, à la fois pour intégrer des améliorations de méthodes ou de données et pour tenir compte de certains changements intervenus dans la structure ou la conduite des centrales.

L'EPS publiée par l'IPSN en 1990 se donnait pour objectif « d'évaluer l'importance des problèmes de sûreté en exploitation et de juger l'intérêt d'éventuelles modifications », et plus spécifiquement :

- « la mise en évidence des points faibles de conception,
- l'analyse des spécifications techniques d'exploitation et des procédures de conduite,
- l'analyse des essais périodiques et de la maintenance en exploitation,
- l'identification des domaines de recherche » (IPSN, 1990a, p. 6).

La quantification du risque, qui est le principe même de l'EPS et en laquelle les experts de sûreté plaçaient de grands espoirs dans les années 1970, n'était pas même évoquée dans cette liste.

Cette absence était révélatrice d'une position de fond. Bien qu'elle ait été adoptée dans un nombre croissant d'applications, l'approche probabiliste a continué à jouer en France un rôle restreint d'aide à l'analyse de la sûreté. L'Autorité de sûreté nucléaire décrit ainsi sa place dans la doctrine française de sûreté nucléaire :

« La démonstration de la sûreté des réacteurs nucléaires français repose pour l'essentiel sur une approche déterministe, c'est-à-dire que les dispositions de conception retenues par l'exploitant sont justifiées par l'étude d'une série d'accidents de dimensionnement et par l'application de règles et critères qui incluent des marges et des conservatismes.

Cette approche est complétée par la réalisation des études probabilistes de sûreté (EPS). Les EPS sont une méthode d'évaluation des risques fondée sur une investiga-

tion systématique des scénarios accidentels. Elles se composent d'un ensemble d'analyses techniques permettant d'apprécier les fréquences d'événements redoutés et leurs conséquences. Elles permettent d'obtenir une appréciation globale du niveau de sûreté, intégrant aussi bien la fiabilité des équipements que le comportement des opérateurs.

À ce titre, elles apportent une aide dans la définition et la hiérarchisation des actions à mener en vue de maintenir un niveau de sûreté satisfaisant ou de l'améliorer » (ASN, 2002b).

L'enchaînement des deux dernières phrases de cet énoncé peut surprendre : si ces méthodes « permettent d'obtenir une appréciation globale du niveau de sûreté », pourquoi faut-il simplement les utiliser comme « aide dans la définition et la hiérarchisation des actions à mener en vue de maintenir un niveau de sûreté satisfaisant » ? Ne semble-t-il pas naturel de se demander si cette appréciation *indique elle-même* un niveau de sûreté satisfaisant ? dans le cas contraire, d'identifier, parmi les facteurs de risque pris en compte dans l'EPS, ceux qui paraissent les plus insatisfaisants ? à la vue de cette indication, s'interroger sur l'adéquation des moyens mis en œuvre pour assurer la sûreté des installations nucléaires ? etc.

L'explication de ce hiatus réside dans un argument ainsi résumé par un expert de l'IRSN en 1996 :

« [L']intérêt [de la méthodologie probabiliste] ne se limite évidemment pas à l'estimation de valeurs absolues de probabilités de fusion de cœur, de rejets ou de morts. Ces valeurs sont, en effet, entachées d'une très large incertitude ce qui conduit à des difficultés pour définir des objectifs probabilistes globaux. Elle donne surtout de nombreuses informations sur les composantes de ces probabilités et leurs poids relatifs. Il est ainsi possible de dégager des priorités et de porter l'effort sur les points sensibles » (Libmann, 1996, p.271).

L'incertitude des évaluations probabilistes du risque constitue donc une faiblesse qui en limite et oriente l'usage : si elles permettent d'identifier les principaux contributeurs au risque et d'apporter des améliorations ponctuelles aux dispositifs de sûreté existants, elles ne peuvent en revanche servir de base à des objectifs réglementaires. Cette appréciation semble se référer à deux précédents : les intentions affichées par le SCSIN dans ses deux lettres à ÉdF en 1977-78, dont elle confirme la désuétude, et les *safety goals* définis par la NRC en 1986, qui se sont avérés inapplicables en l'état.

Cette position laisse cependant certaines questions sans réponse, notamment lorsqu'il s'agit de hiérarchiser les actions à mener et de dégager des priorités. Le rapport de l'EPS 2 pour les REP 900 MWe permet d'illustrer ce point (IRSN, 2009).

Le rapport présente des évaluations de probabilités et de conséquences pour l'ensemble des familles accidentelles associées à une fusion du cœur. L'évaluation des conséquences va au-delà des

termes-source, point d'achèvement habituel d'une EPS2, pour considérer la dose de radiations reçue par une personne située à 2 km de la centrale durant quinze jours. Deux des familles accidentelles évaluées sont, à titre d'exemple, la rupture induite des tubes de générateurs de vapeur, à laquelle on a attribué une probabilité de $4 \cdot 10^{-6}$ et des conséquences équivalentes à 200 mSv, et la perte de réfrigérant primaire avec bipasse de l'enclaustrage de confinement, pour laquelle ces valeurs sont respectivement de $2 \cdot 10^{-8}$ et de 50 Sv.

Le rapport commence par représenter l'ensemble des familles sur un diagramme de Farmer avec une frontière d'acceptabilité de pente $-1,5$. Le diagramme ne répartit pas seulement les scénarios accidentels en deux groupes, situés de part et d'autre de la droite ; il fournit aussi, à travers la distance à cette droite³³, un critère de hiérarchisation des scénarios. L'application de ce critère conduirait, en dehors de toute considération d'acceptabilité intrinsèque, à donner la priorité à la seconde famille.

Après avoir précisé que « cette représentation est indicative, de tels critères n'étant pas utilisés en France », le rapport procède toutefois à une hiérarchisation des scénarios sur la base de ce qu'il appelle une « définition du risque » : le produit de la probabilité par les conséquences (la dose à 2 km et à quinze jours). Cette définition constitue naturellement un critère alternatif de sélection des familles accidentelles, que l'on peut représenter par une droite de pente -1 dans le diagramme de Farmer ; en d'autres termes, elle équivaut à une indifférence au risque. Son application conduit à inverser l'ordre de traitement des deux scénarios cités.

4.4 Conclusion

Nous avons vu, au début du chapitre précédent, que les définitions opérationnelles de la sûreté (telle que donnée par l'AIEA) se référaient à une appréciation externe de ce qu'étaient des « conditions de fonctionnement correctes » ou des « risques indus ». Cette appréciation est portée par un décideur public au nom de la société. De son point de vue, la sûreté constitue donc une question plus large, que l'on peut formuler comme un problème de décision face à l'incertitude. Il s'agit en premier lieu de l'incertitude inhérente aux dangers d'une installation nucléaire, que l'on s'efforce de limiter par des interventions destinées à prévenir les accidents et limiter leurs retombées. Pour concevoir ces interventions et prévoir leurs effets, il faut améliorer la compréhension des phénomènes à l'œuvre et des développements possibles au sein de l'installation, ce qui constitue une deuxième forme de réduction de l'incertitude. De façon synthétique, on pourrait ainsi énoncer l'objectif d'un décideur public en charge de la sûreté d'un système tel qu'une installation nucléaire : *s'assurer, par l'étude d'un système et la modification de son comportement, que l'incertitude liée aux dangers de son fonctionnement est suffisamment limitée pour répondre aux attentes de la société.*

33. ou à toute droite parallèle, le facteur important étant la pente, qui définit le degré d'aversion pour le risque.

Il est naturel de penser que les attentes de la société, dans le cas d'espèce, dépendent également des bénéfices liés aux activités nucléaires civiles, considérés sur l'ensemble du cycle de combustible, ou encore des avantages relatifs d'autres formes d'énergie. Plus encore que pour d'autres choix publics en situation d'incertitude, il ne saurait être question pour une instance quelconque d'internaliser l'ensemble des préférences sociales et de rendre une décision au nom de la société face à une question d'une telle complexité. Il ne saurait pas plus être question de rassembler l'ensemble des éléments de la décision dans une analyse unique destinée à indiquer un choix optimal, à l'image de l'analyse coûts-bénéfices. Pour cette raison, il semble nécessaire que tous les éléments d'information pertinents soient rendus publics avec une grande clarté.

Si l'on adopte un tel point de vue, l'objectif de sûreté exige donc de l'autorité publique qu'elle s'attache à prendre en considération l'incertitude dans son intégrité et qu'elle verse l'ensemble de cette information au débat public de façon claire.

Dans le contexte historique qui les a vus naître, les politiques de sûreté nucléaire s'étaient donné pour but d'une part de réaliser un effort exceptionnel de prévention des accidents graves, et de l'autre de rassurer les populations au sujet des risques qu'elles couraient. Ce double objectif a conduit à un déni évident de l'incertitude dans laquelle se trouvaient ces autorités, que favorisaient également des organisations institutionnelles qui soit les plaçaient en situation de conflit d'intérêt, soit les soumettaient aux arbitrages d'instances supérieures elles-mêmes sujettes à de tels conflits. Deux éléments des doctrines originelles de la sûreté permettent d'illustrer cette attitude de déni : la façon dont les méthodes déterministes ont été mises en œuvre et les choix sémantiques à travers lesquels la politique de la sûreté a été formulée et présentée.

L'approche déterministe correspondait certes à une attitude de grande précaution, où l'on cherchait à assurer que toutes les mesures raisonnables étaient prises pour éviter la survenue d'événements défavorables, indépendamment des chances de cette survenue. Cependant, l'absence de recul vis-à-vis de ces méthodes, particulièrement de la part des autorités gouvernementales chargées de contrôler la sûreté, a entraîné l'éviction de la question de l'incertitude au moins jusqu'à la fin des années 1960. À cette époque, Farmer, l'un des plus farouches opposants du rôle exclusif accordé aux concepts déterministes tels que l'Accident Maximum Crédible dans la politique de sûreté, expliquait que ceux-ci ne permettaient pas aux experts de sûreté d'établir un dialogue avec les concepteurs et les exploitants (cité par Foasso (2003, p.95)). L'hypothèse selon laquelle l'approche de sûreté serait insuffisante était en effet d'emblée considérée incroyable. Ce cloisonnement à l'intérieur d'une approche qui comportait de lourdes incertitudes a été conforté par le fait que, dès lors qu'elles ne pouvaient être pondérées par une quelconque probabilité (celle-ci étant impossible à évaluer du fait de l'empilement de mesures conservatrices), les conséquences d'un accident plus important que l'AMC paraissaient inacceptables. C'est finalement la controverse publique qui a fini par faire une

place à l'incertitude dans la démarche des autorités de sûreté, particulièrement aux États-Unis.

Le choix de la terminologie des politiques de sûreté témoigne également d'une absence d'attention pour la question de l'incertitude. L'expression « démontrer la sûreté » nous semble, à ce titre particulièrement révélatrice, d'autant qu'elle reste d'usage courant, notamment en France. Pourtant, ayant admis que la sûreté ne pouvait être démontrée, l'Agence pour l'Énergie Nucléaire se trouvait bien embarrassée en 1991 pour expliquer ce que pouvait bien désigner l'expression :

"Absolute proof of continuing safe behaviour is impossible for all technical systems, including radioactive waste disposal systems. What must be achieved is *a convincing and indirect demonstration* that the proposed disposal system *provides a sufficient level of safety* to both current and future generations" (AEN, 1991, nous soulignons).

Or même s'il ne s'agit que d'un niveau suffisant de sûreté, dès lors que celui-ci est mesuré par des objectifs réglementaires en termes de santé publique ou même de probabilités de rejets, la qualification de démonstration est totalement fallacieuse et conduit à nier le caractère incertain des représentations théoriques qui permettent d'affirmer la réalisation de ces objectifs. Le terme ne pourrait tout au plus convenir que pour indiquer que certaines mesures techniques ont bien été prises. L'Agence pour l'Énergie Nucléaire semblait partager cette opinion en 1999 :

"The term "demonstration of safety", as used in the present report, is not intended to imply a rigorous proof of safety, in a mathematical sense, but rather *a convincing set of arguments that support a case for safety*" (AEN, 1999, nous soulignons).

« Un ensemble convaincant d'arguments » semble effectivement être une terminologie plus appropriée qu'une « démonstration » pour décrire les garanties que l'on peut apporter dans une situation où les connaissances restent lacunaires. Ceci nous semble particulièrement vrai pour des autorités publiques dont le rôle est précisément de souligner ces lacunes et de pousser à leur réduction.

Les choix opérés quant à l'usage des concepts déterministes et celui des mots illustrent le fait que la politique de sûreté, outre être subordonnée à la politique industrielle dans le domaine nucléaire, relevait d'une vision technocratique : le rôle des premiers régulateurs qu'étaient les « experts de la sûreté » dans la prise en compte des incertitudes liées aux installations nucléaires se résumait à la mise en place d'un ensemble de dispositifs techniques. Il était certes question de remettre en cause ces derniers sur un plan technique, mais pas d'en interroger la suffisance du point de vue de la société.

Après quatre à cinq décennies d'expérience d'exploitation des centrales électronucléaires et aussi d'évolution importante des sociétés, les politiques de sûreté nucléaire ont connu des bouleversements majeurs tant dans leur contenu que dans leur organisation institutionnelle. L'autorité de sûreté a désormais des responsabilités clairement établies, aussi bien dans la justification de la démarche de sûreté auprès de la société que dans la vérification de la mise en place de cette démarche

pas les acteurs industriels. Il n'est pas certain, cependant, que l'ensemble des conséquences de ce changement aient été tirées s'agissant de la place de l'incertitude dans les doctrines de sûreté³⁴.

Comme nous l'avons décrit, des deux formes de doctrine rationaliste et structuraliste distinguées par Sorensen et al. (1999), il semble que les États-Unis s'orientent vers la première, qui accorde une primauté aux évaluations probabilistes, alors que la France reste fidèle à la seconde, fondée sur la défense en profondeur. Notre propos n'est pas de discuter ces choix d'une grande complexité technique. Nous avons simplement tenté d'indiquer que le processus de transition en cours aux États-Unis comportait des risques, notamment celui d'une insuffisance des évaluations probabilistes. En France, se pose un problème de cohérence des choix lorsqu'ils sont orientés par une approche probabiliste qui reste partielle, en l'absence de critères clairs de pondération des probabilités et des conséquences.

Cependant, que l'on retienne une doctrine ou l'autre, il semble aujourd'hui demeurer une lacune au sujet des incertitudes qui affectent les deux approches. Ces incertitudes, qui reflètent les limites de nos connaissances des phénomènes à l'œuvre, portent sur les hypothèses fondamentales de modélisation et sont difficilement probabilisables. Leur prise en compte reste, pour l'heure, fragmentaire et fréquemment dictée par la survenue d'événements particuliers. Les modalités d'une recherche systématique des failles dans les dispositifs de sûreté liés à ces formes d'incertitude restent à définir.

34. Ni, d'ailleurs, s'agissant de la terminologie, mais nous ne revenons pas sur ce point.

Troisième partie

Une approche causale du risque

Chapitre 5

Une représentation par les réseaux bayésiens causaux

Qu'est-ce en effet que l'aléatoire ?

*On ne peut en donner une définition que purement négative :
est aléatoire un processus qui ne peut être simulé par aucun mécanisme [...].*

Thom (1980)

5.1 Introduction

Dans les chapitres précédents, nous avons défini la sûreté dans ces termes : *s'assurer, par l'étude d'un système et la modification de son comportement, que l'incertitude liée aux dangers de son fonctionnement est suffisamment limitée pour répondre aux attentes de la société*. Nous avons vu qu'elle s'appuyait, dans le domaine nucléaire, sur deux grandes démarches, fondées principalement sur des raisonnements déterministes pour l'une et probabilistes pour l'autre. Nous avons également souligné les limites de chacune de ces démarches, et indiqué l'utilité d'un cadre analytique permettant de mener une analyse comparative de ces deux démarches. Nous nous donnons pour objectif, dans ce chapitre, de bâtir un tel cadre à partir de la théorie des réseaux bayésiens causaux (RBC).

Nous consacrons la première partie de ce chapitre à une brève exposition de la théorie des réseaux bayésiens et de leur version causale en nous contentant, pour l'essentiel, de proposer notre synthèse des principaux résultats établis par les fondateurs de cette théorie. Nous développons notre propre approche dans la seconde partie du chapitre, notamment en proposant une généralisation du concept d'intervention. Nous montrons comment le formalisme des réseaux bayésiens peut être utilisé pour rendre compte des démarches d'inférence en analyse de la sûreté. Nous caractérisons

alors les approches déterministe et probabiliste de la sûreté comme deux stratégies d'étude et de modification d'un RBC partiellement identifié : le réseau *idéal* décrivant l'ensemble des dangers d'un *système*. Sous cet éclairage, il devient plus aisé d'examiner les limites de chacune des deux approches ainsi que l'étendue de leur complémentarité et, pour finir, de proposer une définition précise du risque et du risque acceptable.

Il peut être utile, avant d'entrer dans le vif du sujet, de préciser que le système sur lequel nous portons toute notre attention dans ce chapitre n'est pas nécessairement une entité physique. Il s'agit de toute *description* d'une situation à partir d'un ensemble de grandeurs et de relations entre ces grandeurs, ces dernières vérifiant certaines propriétés de stabilité et d'autonomie que nous expliciterons. Les modèles scientifiques, à titre d'exemple, constituent fréquemment de telles descriptions.

5.2 Les réseaux bayésiens causaux

5.2.1 Les réseaux bayésiens

Soit $X = (X_1, \dots, X_n)$ une variable aléatoire vectorielle à valeurs dans $\chi = \chi_1 \times \dots \times \chi_n$, que l'on peut se représenter comme une description de n aspects des états d'un système. Nous supposons que l'ensemble χ des états du système est fini¹. Soit $\Delta(\chi)$ l'ensemble des distributions de probabilités sur χ .

Considérons une représentation graphique D des composantes de X , où chacune d'entre elles est associée à un *nœud* relié à d'autres par des *arcs dirigés*. Dans la suite, nous appellerons D indifféremment *graphe* ou *structure*. Lorsque D ne comporte pas de cycle (c'est-à-dire de succession d'arcs qui permette, en suivant le sens des flèches, de partir d'un nœud et de revenir à lui), nous l'appelons *graphe dirigé acyclique*. Nous appelons en outre *sous-graphe* de D tout graphe dirigé acyclique qui comporte les mêmes nœuds que D et un sous-ensemble (au sens strict) de ses arcs. À chaque fois que deux nœuds sont reliés par un arc, nous appelons *parent* celui au départ de l'arc et *enfant* celui à l'arrivée. Nous parlons de même d'*ascendants* et de *descendants*. Nous noterons les ensembles d'indices des parents, des enfants, des ascendants et des descendants de X_i par $Pa(i)$, $En(i)$, $As(i)$ et $De(i)$. Nous appelons *racine* tout nœud sans parent et *feuille* tout nœud sans enfant. Nous appelons enfin *sentier* toute succession d'arcs, sans considération de direction.

L'ordre dans lequel les composantes de X sont classées joue un rôle important dans les développements qui suivent. Nous dirons alors d'un ordre $(1, \dots, n)$ qu'il est *compatible avec d* lorsque chaque variable y précède l'ensemble de ses descendants dans D .

1. La prise en compte d'un ensemble d'états infini et de variables continues pose des difficultés méthodologiques non négligeables pour certains des développements que nous rapportons ici, même si elles semblent plus facilement envisageables pour la version particulière des RBC dont nous faisons usage dans la partie suivante de ce chapitre (et dans les chapitres suivants).

Dans la suite du chapitre, nous noterons les variables par des lettres majuscules (X), leurs valeurs génériques par des lettres minuscules italiques (x) et leurs valeurs spécifiques par des lettres minuscules droites (x). Les variables (ou valeurs) indexés par un ensemble d'indices désigneront le vecteur des variables (ou valeurs) correspondantes, et les ensembles indexés par un ensemble d'indices désigneront le produit des ensembles correspondants. Ainsi $\chi_{Pa(i)} = \bigotimes_{j \in Pa(i)} \chi_j$. Dans tous les cas où un ensemble d'indices sera vide, la variable vectorielle sera définie, par convention, comme une constante convenablement choisie. Enfin, nous adopterons l'écriture simplifiée $P(x | y)$ pour $P(X = x | Y = y)$.

5.2.1.1 Condition de Markov et réseaux bayésiens

Remarque 5.1. Tout graphe dirigé acyclique admet au moins une racine et une feuille.

Démonstration. Supposons qu'un graphe dirigé acyclique n'admette pas de feuille. Alors il devrait être possible de le parcourir en suivant, depuis chaque nœud, un arc dirigé menant à un nœud qui n'a pas encore été visité (faute de quoi on aurait suivi un cycle). Or, une fois que les n nœuds auraient été ainsi visités, cette hypothèse deviendrait absurde. Pour établir l'existence d'au moins une racine, il suffit de reproduire le raisonnement en remontant les arcs. $\square$

Considérons par ailleurs une distribution de probabilités $P \in \Delta(\chi)$ sur les réalisations de X . Pour un ordre $(1, \dots, n)$ donné sur les composantes de X , P peut être écrit comme le produit de n distributions conditionnelles :

$$P(X_1, \dots, X_n) = P(X_n | X_{n-1}, \dots, X_1) \dots P(X_2 | X_1). P(X_1) \quad (5.2.1)$$

Nous nous intéressons aux relations d'indépendance conditionnelle entre les X_i . Pour trois groupes disjoints d'indices I, J et K inclus dans $\{1, \dots, n\}$, la notation $(X_I \perp\!\!\!\perp X_J | X_K)_P$ signifie « X_I est indépendant de X_J conditionnellement à X_K selon la distribution P », c'est-à-dire : $P(X_I | X_J, X_K) = P(X_I | X_K)$.²

Définition 5.2. On dit qu'un graphe dirigé acyclique D est un réseau bayésien pour une distribution de probabilités P s'il existe un ordre compatible avec D tel que conditionnellement à ses parents dans D , toute variable soit indépendante de tous ses autres prédécesseurs selon P , en d'autres termes tel que pour tout i :

$$P(X_i | X_{i-1}, \dots, X_1) = P(X_i | X_{Pa(i)})$$

où $Pa(i)$ désigne l'ensemble des indices des parents de X_i dans D .

². Cette notation, ainsi que la première étude des propriétés de l'indépendance conditionnelle exploitées par la théorie des réseaux bayésiens, sont dues à Dawid (1979).

On dira également que D représente P ou que P vérifie la condition de Markov (ou encore est Markov) relativement à D .

La décomposition (5.2.1) devient alors :

$$P(X_1, \dots, X_n) = \prod_{i=1}^n P(X_i | X_{pa(i)}) \quad (5.2.2)$$

On démontre facilement que la factorisation de P sous la forme (5.2.2) est équivalente à la condition de Markov.

5.2.1.2 Indépendance conditionnelle et d-séparation

Après avoir établi un lien entre graphes dirigés acycliques et distributions de probabilités relatifs à un ensemble de variables, nous cherchons à présent un équivalent graphique à la propriété d'indépendance conditionnelle. Notons $\Delta^D(\chi)$ l'ensemble des distributions de probabilités sur χ qui sont Markov relativement à D .

Définition 5.3. Pour trois groupes disjoints d'indices I, J et K inclus dans $\{1, \dots, n\}$, on dit que X_I est d-séparée de X_J par X_K dans le graphe D , que l'on note $(X_I \perp\!\!\!\perp X_J \mid X_K)_D$, si tout sentier allant d'un élément de X_I à un élément de X_J est bloqué par X_K , c'est-à-dire qu'il contient soit une chaîne $X_h \rightarrow X_k \rightarrow X_l$ telle que X_k est dans X_K , soit une fourche $X_h \leftarrow X_k \rightarrow X_l$ telle que X_k est dans X_K , soit une fourche inversée $X_h \rightarrow X_k \leftarrow X_l$ telle que ni X_k ni aucun de ses descendants ne sont dans X_K . Lorsque deux groupes de nœuds ne sont pas d-séparés par un troisième, on dit qu'ils sont d-connectés par lui.

Théorème 5.4. Si deux variables vectorielles X_I et X_J sont d-séparées par une troisième X_K dans le graphe D , alors elles sont indépendantes conditionnellement à X_K pour toute distribution de probabilités P représentée par D . Inversement, si X_I et X_J sont d-connectées par X_K dans le graphe D , alors elles sont dépendantes conditionnellement à X_K pour au moins une distribution de probabilités P représentée par D .

$$(X_I \perp\!\!\!\perp X_J \mid X_K)_D \Leftrightarrow \forall P \in \Delta^D(\chi) \quad (X_I \perp\!\!\!\perp X_J \mid X_K)_P$$

Démonstration. Suffisance : Pour démontrer que la d-séparation de X_I et X_J par X_K dans D entraîne l'indépendance de X_I et X_J conditionnellement à X_K pour toute distribution de probabilité Markov relativement à D , nous pouvons raisonner par récurrence sur le nombre de variables. Le résultat est trivial pour $n = 1$; supposons-le vérifié à l'ordre $n - 1$, $n \geq 2$. Sans perte de généralité, nous pouvons supposer que les nœuds sont ordonnés selon un ordre compatible avec D , et que le dernier d'entre eux est une feuille (d'après la remarque 5.1, nous savons que l'un d'entre eux au

moins est une feuille). Appelons D' le graphe obtenu en amputant D de X_n et des arcs qui mènent à elle. Quatre cas sont possibles :

1. $n \notin I \cup J \cup K$. Alors, pour tout $P \in \Delta^D(\chi)$:

$$\begin{aligned} P(X_1, \dots, X_{n-1}) &= \sum_{x_n \in \chi_n} \prod_{i=1}^n P(X_i | X_{pa(i)}) \\ &= \prod_{i=1}^{n-1} P(X_i | X_{pa(i)}) \sum_{x_n \in \chi_n} P(X_n | X_{pa(n)}) \\ &= \prod_{i=1}^{n-1} P(X_i | X_{pa(i)}) \end{aligned}$$

La distribution marginale de $(X_1, \dots, X_{n-1})$ engendrée par P est donc Markov relativement à D' . Comme X_I et X_J sont d-séparées par X_K dans D' , X_I est indépendante de X_J conditionnellement à X_K pour P .

2. $n \in I$. Posons $I = I' \cup \{n\}$. Alors :

$$P(X_I | X_J, X_K) = P(X_n | X_{I'}, X_J, X_K) \cdot P(X_{I'} | X_J, X_K)$$

et :

$$P(X_n | X_{I'}, X_J, X_K) = \sum_{x_{Pa(n)} \in \chi_{Pa(n)}} P(X_n | X_{Pa(n)}) \cdot P(X_{Pa(n)} | X_{I'}, X_J, X_K)$$

Pour que X_I soit d-séparée de X_J par X_K , il faut que chaque parent de X_n soit d-séparé de X_J par X_K ou qu'il soit lui-même une composante de X_K , de sorte que :

$$P(X_{Pa(n)} | X_{I'}, X_J, X_K) = P(X_{Pa(n)} | X_{I'}, X_K)$$

et donc :

$$P(X_n | X_{I'}, X_J, X_K) = P(X_n | X_{I'}, X_K)$$

Comme par ailleurs $X_{I'}$ est d-séparée de X_J par X_K dans D' , nous savons d'après le cas 1 que $X_{I'}$ est indépendante de X_J conditionnellement à X_K pour P . Par suite :

$$\begin{aligned} P(X_I | X_J, X_K) &= P(X_n | X_{I'}, X_K) \cdot P(X_{I'} | X_K) \\ &= P(X_I | X_K) \end{aligned}$$

3. $n \in J$. Ce cas est symétrique au précédent.

4. $n \in K$. Posons $K = K' \cup \{n\}$. Étant une feuille, X_n ne peut d-séparer aucune paire de nœuds et doit en revanche être d-séparée de X_I ou de X_J (sinon, X_n serait au centre d'une fourche inversée

qui d-connecterait X_I et X_J). Il s'ensuit donc d'une part que X_I et X_J sont d-séparées par $X_{K'}$ dans D , et de l'autre que X_n est d-séparée de X_I ou de X_J par $X_{K'}$ dans D . Dès lors, soit X_I et (X_J, X_n) sont donc d-séparées par $X_{K'}$ dans D , ce qui nous ramène au cas 3, soit (X_I, X_n) et X_J sont donc d-séparées par $X_{K'}$ dans d , ce qui nous ramène au cas 2. Dans un cas comme dans l'autre, nous est alors facile de montrer que X_I est indépendante de X_J conditionnellement à $(X_{K'}, X_n)$ pour P .

Nécessité : Considérons trois groupes disjoints d'indices I, J et K tels que X_I soit d-connectée à X_J par X_K dans le graphe D . Alors il existe $i \in I, j \in J$ tels que X_i est d-connectée à X_j par X_K , c'est-à-dire qu'il existe au moins un sentier reliant X_i et X_j tel qu'aucune chaîne ou fourche n'y a pour nœud central une composante de X_K , et toute fourche inversée y a pour nœud central une composante de X_K ou l'un de ses ascendants.

Considérons alors le sous-graphe D'' constitué, comme dans l'exemple du graphique :

- du sentier de ce type comportant le plus petit nombre de fourches inversées ;
- des sentiers qui relient le nœud central de chacune de ses fourches inversées à la composante la plus proche de X_K . Ces sentiers ne se coupent pas entre eux, et ne coupent le sentier principal qu'en un point unique. Dans le cas contraire, en effet, il serait possible de définir un nouveau sentier principal comportant moins de fourches inversées que le précédent.

Considérons enfin la distribution de probabilité P' ainsi définie :

$P'(X_h) = \delta_h$ si X_h est une racine de D'' , où δ_h est une distribution de probabilités non dégénérée sur χ_h

$P'(X_h) = \delta_{Pa(h)}^f$ sinon, où $\delta_{Pa(h)}^f$ est la mesure image de la distribution jointe $\delta_{Pa(h)}$ des parents de X_h dans D'' par une fonction mesurable non dégénérée f_h définie de $\chi_{Pa(h)}$ dans χ_h .

Désignons les nœuds du sentier par la suite d'indices $h_1, \dots, h_s$, avec $X_{h_1} = X_i$ et $X_{h_s} = X_j$. Nous savons que X_{h_2} est dépendant de X_i conditionnellement à X_K pour P' (puisque'il est soit enfant, soit parent de X_i , en d'autres termes soit son image soit sa pré-image par f_h). Supposons que la dépendance conditionnelle soit vérifiée jusqu'au nœud X_{h_t} , et considérons le nœud suivant :

- si l'arc entre X_{h_t} et $X_{h_{t+1}}$ est orienté vers ce dernier, alors :

$$P(X_{h_{t+1}} \mid X_i, X_K) = P(f_{h_{t+1}}(X_{h_t}) \mid X_i, X_K)$$

- si le triplet de nœuds $(X_{h_{t-1}}, X_{h_t}, X_{h_{t+1}})$ est une chaîne orientée vers $X_{h_{t-1}}$, alors :

$$P(X_{h_{t+1}} \mid X_i, X_K) = P(f_{h_t}^{-1}(X_{h_t}) \mid X_i, X_K)$$

- enfin, si le triplet de nœuds $(X_{h_{t-1}}, X_{h_t}, X_{h_{t+1}})$ est une fourche inversée, alors :

$$X_{h_t} = f_{h_t}(X_{h_{t-1}}, X_{h_{t+1}})$$

Nous savons en outre que tous les sommets de fourches inversées sur le sentier principal sont des ascendants d'une composante X_k de X_K , de sorte que $X_{h_t} = g_{h_t}^{-1}(X_k)$, où g_{h_t} est une fonction mesurable non dégénérée définie de χ_{h_t} dans χ_k .

Dans les trois cas, la connaissance de X_K suffit donc à créer une dépendance entre X_i et $X_{h_{t+1}}$. Par récurrence, X_j est dépendant de X_i conditionnellement à X_K pour P' .

Il suffit, pour finir, d'étendre P' au reste du graphe D en posant $P'(X_h) = \delta_{x_h}$ pour $x_h \in \chi_h$, où δ est la mesure de Dirac. La distribution de probabilités P' est Markov relativement à D , et X_j est dépendant de X_i conditionnellement à X_K pour P' . $\square$

Corollaire 5.5. *P est Markov par rapport à D ssi conditionnellement à ses parents, toute variable est indépendante de toutes les variables autres que ses descendants dans un ordre compatible avec D .*

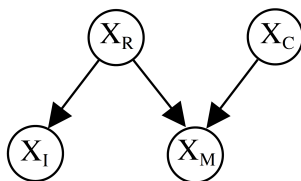
Démonstration. Suffisance : Il suffit de constater que chaque variable est d-séparée par ses parents des variables autres que ses descendants, et d'appliquer le théorème.

Nécessité : Pour chaque ordre compatible avec D , les prédécesseurs de chaque variable sont un sous-ensemble des variables autres que ses descendants. $\square$

Le théorème 5.4 est dû à Verma et Pearl (1988) pour la première partie (suffisance) et à Geiger et Pearl (1988) pour la seconde (nécessité). Notre démonstration s'inspire largement des leurs.

Dans un réseau bayésien, la propriété de séparation directionnelle est donc la traduction graphique de l'indépendance conditionnelle. Considérons l'exemple de quatre variables X_R, X_I, X_C et X_M dont la distribution de probabilités jointe P serait Markov par rapport au graphe de la figure 5.2.1. X_R désigne ici l'activité d'une usine qui rejette une substance cancérigène dans l'environnement, X_I les bénéfices économiques liés à cette activité, X_C la consommation de produits agricoles locaux et X_M le développement de cancers dans la population. X_M est alors dépendante de X_I selon P , mais le conditionnement par X_R fait disparaître cette dépendance. Sur le graphe, on constate que X_R d-sépare X_M et X_I . L'une et l'autre de ces propriétés ont la même signification en termes de pertinence de l'information apportée par l'observation d'une variable pour la prévision de l'autre : la connaissance du niveau de bénéfices est utile pour prévoir l'ampleur de la maladie, mais elle ne l'est plus dès lors que l'on connaît aussi le niveau de rejets de l'usine. Inversement, X_R et X_C sont indépendantes, mais ne le sont plus conditionnellement à X_M et, de façon équivalente, tout ensemble de variable qui ne contient pas X_M d-sépare X_R et X_C : l'observation du niveau de

FIGURE 5.2.1 – Un graphe dirigé acyclique



rejets n'apporte aucun renseignement au sujet de la consommation de produits locaux, sauf si l'on connaît également l'ampleur de la maladie.

Ces deux cas de figure illustrent l'avantage des réseaux bayésiens par rapport aux graphes dits de Markov, où les relations de dépendance entre variables sont représentées par des arcs non dirigés. Dans ces graphes, on appelle ensemble de coupe (*cutset*) entre deux groupes de variables tout ensemble contenant au moins un nœud de chaque sentier allant d'un nœud du premier groupe à un nœud du second. La coupe, qui est donc l'équivalent de la d-séparation pour un graphe non dirigé, ne permet cependant pas de distinguer une situation où deux variables sont dépendantes, mais indépendantes conditionnellement à une troisième (la fourche $X_I \leftarrow X_R \rightarrow X_M$), d'une autre où deux variables sont indépendantes, mais dépendantes conditionnellement à la troisième (la fourche inversée $X_R \rightarrow X_M \leftarrow X_C$). Pour représenter l'indépendance conditionnelle dans un graphe, il est donc nécessaire que les arcs de celui-ci soient dirigés. On voit par là-même que la propriété d'indépendance conditionnelle est plus informative que celles de dépendance et d'indépendance classiques, puisqu'elle permet de distinguer deux formes de dépendance.

Deux aspects des réseaux bayésiens interdisent d'interpréter systématiquement la d-connection comme une marque de dépendance conditionnelle. Le premier est que la définition d'un réseau bayésien n'exclut pas l'existence d'un arc entre deux nœuds représentant des variables indépendantes. Ceci constitue une faiblesse évidente, que l'on corrigera en admettant une condition de minimalité³.

Définition 5.6. Un graphe D représentant une distribution de probabilités P est dit minimal s'il n'admet pas de sous-graphe D' tel que D' représente P .

Un graphe est donc minimal lorsque l'on ne peut en retirer aucune flèche sans contrevenir à la condition de Markov.

Le second aspect est qu'il est possible, lorsque deux variables ont un ancêtre commun, que leurs relations de dépendance à son égard se compensent exactement, de sorte que ces variables soient indépendantes entre elles. Les principaux auteurs s'accordent à écarter ces cas exceptionnels par

3. La dénomination est due à Spirtes, Glymour and Scheines (1993). Chez Pearl (2000), elle est intégrée à la définition des parents markoviens, tandis qu'une condition plus restrictive, que nous mentionnerons dans la suite, est dite de minimalité.

le biais d'une condition supplémentaire dite de fidélité⁴, qui exige explicitement que la distribution de probabilités P ne contienne que les relations d'indépendance conditionnelle reflétées par la d-séparation dans D . Nous mentionnons cette clause par souci d'exhaustivité, et pour y faire brièvement référence dans la suite. Mais nous n'aurons pas à la supposer vérifiée.

5.2.1.3 D'une distribution de probabilités à un réseau bayésien : le problème de la dépendance à l'ordre

Supposons connue la distribution de probabilités de $X = (X_1, \dots, X_n)$. La remarque suivante permet de construire un graphe minimal qui la représente.

Remarque 5.7. Soit D un graphe tel que :

1. X_1 est une racine.
2. Pour $i \geq 2$, X_i reçoit un arc entrant de tous les éléments du plus petit sous-ensemble de ses prédécesseurs dont la connaissance la rend indépendante de ses autres prédécesseurs selon P .

Alors, par construction, D représente P et D est minimal.

L'ensemble des parents de chaque variable est unique lorsque P est strictement positive (autrement dit lorsque toutes les valeurs de X sont permises). Dans ce cas, le graphe D construit selon la procédure décrite est unique *pour l'ordre considéré*⁵. Si toutefois on n'impose pas à P d'être positive (par exemple lorsque certaines valeurs d'une variable sont exclues par certaines réalisations de ses parents), l'unicité pour un ordre donné n'est plus garantie.

Un problème plus sérieux encore est posé par la dépendance à l'ordre $(1, \dots, n)$. Comme l'ordonnement en parents et enfants, le sens des flèches et donc le graphe D lui-même dépendent de l'ordre dans lequel les variables ont été prises en compte, et que l'on ne dispose d'aucun critère pour préférer un ordre à un autre, on est donc en mesure d'associer une multitude de graphes différents à la même distribution de probabilités.

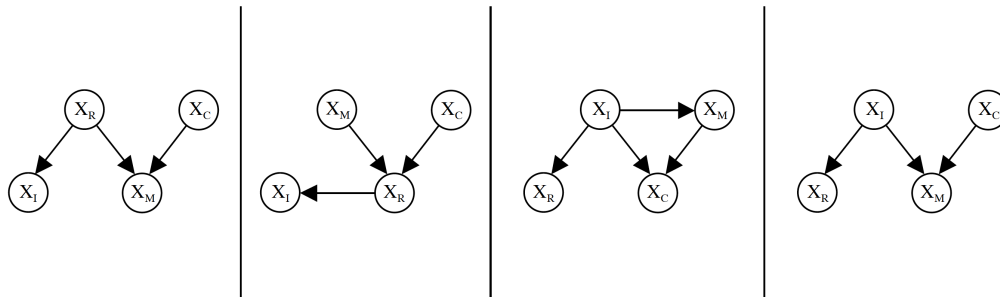
Considérons à nouveau l'exemple précédent. Le diagramme de gauche de la figure 5.2.2 représente la structure initiale, obtenue en considérant les variables dans l'ordre X_R, X_I, X_C, X_M . Les trois autres peuvent représenter la même distribution jointe que le premier si l'on change simplement l'ordre de prise en compte des variables (respectivement en X_M, X_C, X_R, X_I , puis X_I, X_M, X_C, X_R et enfin X_C, X_I, X_R, X_M). Le deuxième diagramme présente une situation remarquable : comme nous l'avons relevé plus haut, X_R n'est pas indépendante de X_C conditionnellement à X_M ; l'ensemble de ses parents dans (X_M, X_C) est donc (X_M, X_C) . Le troisième diagramme présente un cas similaire de dépendance factice, cette fois entre X_C et (X_I, X_M) .

Constatons cependant que certaines de ces structures n'ont pas le même pouvoir explicatif que

4. Ce terme est également dû à Spirtes, Glymour and Scheines (1993). Chez Pearl (2000), la condition équivalente est dite de stabilité.

5. voir Pearl (1988, sections 3.2. (théorème 4) et 3.3. (théorème 9)).

FIGURE 5.2.2 – Influence de l'ordre sur la structure d'un graphe



d'autres, au sens où elles représentent un ensemble plus vaste de distributions de probabilités. Contrairement au premier diagramme de la figure 5.2.2, le deuxième autorise ainsi une dépendance inconditionnelle entre X_R et X_C , et le troisième entre X_I et X_M et entre X_I et X_C .

Le quatrième diagramme, en revanche, représente exactement le même ensemble de distributions de probabilités que le premier. Verma et Pearl (1990) appellent de tels graphes équivalents du point de vue de l'observation (*observationally equivalent*), et montrent qu'ils sont obtenus à chaque fois que l'on peut modifier la direction des flèches d'une structure sans créer ni détruire de fourche inversée.

Dans notre exemple, on observe en revanche que cela n'est pas le cas des arcs entre X_R et X_M et entre X_C et X_M : pour les renverser, il est nécessaire d'introduire des arcs supplémentaires. Le principe fondamental des algorithmes de construction de réseaux bayésiens causaux est d'identifier, à partir d'un ensemble d'observations, les contraintes de ce type, et de choisir alors, parmi les structures possibles, la plus simple.

5.2.2 Les interprétations causales des réseaux bayésiens

Le formalisme des réseaux bayésiens est d'une grande efficacité lorsque les liens entre variables peuvent être interprétés comme des rapports causaux auxquels on peut prêter, si l'on se réfère aux théories dites interventionnistes de la causalité⁶, une certaine invariance aux changements intervenant dans leur environnement. On parle alors de réseaux bayésiens causaux (RBC). L'interprétation causale rend les relations de dépendance et d'indépendance conditionnelle plus faciles à appréhender, utiliser et tester. En particulier, une modification de la relation existante entre deux variables n'affecte un RBC que localement, alors qu'elle remet entièrement en question une représentation fondée sur la seule association probabiliste, telle que celle que nous venons d'exposer (c'est-à-dire

6. Sur les théories de la causalité fondées sur l'intervention ou sur le concept voisin de manipulation, voir Woodward (2008). Pour un exemple de théorie alternative, mais à notre avis moins expressive, de la causalité, voir Shafer (1996).

que la connaissance de la loi jointe d'avant le changement n'apporte rien à celle d'après). De ce fait, les RBC sont des instruments performants d'analyse et d'évaluation des actions.

5.2.2.1 La causalité comme réponse d'un système stochastique à une intervention

Une première approche des RBC consiste précisément à caractériser la causalité comme la propriété d'une représentation qui reste stable lorsque des changements sont imposés à son objet⁷. Le changement consiste ici à rompre les liens entre une variable et ses parents, et à lui affecter une valeur fixe. La condition de stabilité est que les relations de dépendance déterminant les autres variables ne soient pas affectées.

Définition 5.8. On appelle intervention sur une variable X_i une modification du système X qui consiste à rompre les relations de dépendance entre X_i et ses parents et à lui affecter une valeur constante x_i qui soit élément du support de P ⁸.

Une telle intervention est notée $do(X_i = x_i)$ pour la distinguer de l'événement $(X_i = x_i)$ ⁹. Son principe peut être étendu aux interventions sur un sous-ensemble de variables repérées par l'ensemble d'indices I .

Considérons ainsi une distribution de probabilités P sur le système X , notons P_{x_I} la distribution correspondant à l'intervention $do(X_I = x_I)$ sur un sous-ensemble de variables¹⁰ et P^* l'ensemble de toutes les distributions P_{x_I} ainsi obtenues pour tous les sous-ensembles de variables (y compris P , qui correspond au sous-ensemble vide).

Définition 5.9. On dit qu'un graphe D est un réseau bayésien causal compatible avec P^* si les trois conditions suivantes sont vérifiées pour toute distribution P_{x_I} de P^* :

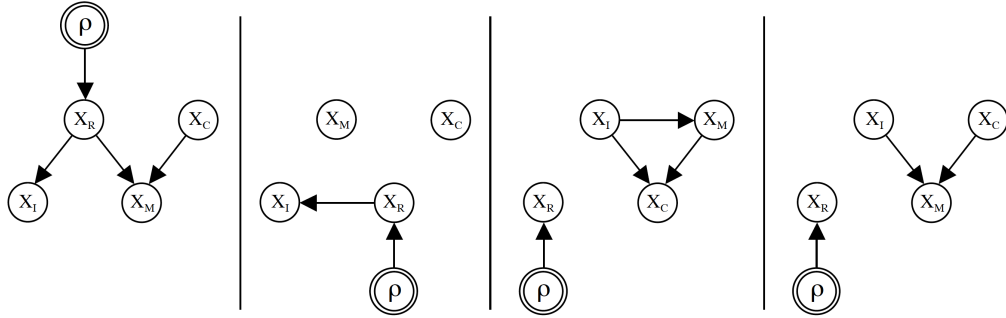
1. P_{x_I} est Markov relativement à D ;
2. $P_{x_I}(x_i) = 1$ pour tout $i \in I$, dès lors que $X_i = x_i$ est compatible avec $X_I = x_I$;
3. $P_{x_I}(x_i | x_{pa(i)}) = P(x_i | x_{pa(i)})$ pour tout $i \notin I$, dès lors que $X_{pa(i)} = x_{pa(i)}$ est compatible avec $X_I = x_I$.

7. Cette approche, qui constitue la première formalisation historique des RBC, est exposée en détail par Spirtes, Glymour and Scheines (1993).

8. Cette dernière condition est fréquemment omise dans la littérature sur les RBC. Or elle est cruciale : si x_i a une probabilité nulle de se réaliser avant l'intervention, alors la dérivation de la distribution de probabilités postérieure à l'intervention à partir de P , telle que la décrit la définition suivante, est impossible. Cette dérivation est en effet l'équivalent d'une révision de P après observation de l'événement $(X_i = x_i)$ selon la règle de Bayes, qui ne peut être appliquée si l'événement en question a une probabilité a priori nulle.

9. Dans de nombreuses disciplines scientifiques, l'intervention est un outil d'inférence essentiel, souvent indissociable de l'observation, par exemple lorsqu'elle consiste à affecter une valeur à une variable de contrôle (voir Rothman, Greenberg et Lash (2008) pour les applications épidémiologiques). Or le langage probabiliste classique ne contient aucun terme approprié pour l'intervention. En particulier, la révision bayésienne s'appuie sur un conditionnement par des événements observés, et n'offre pas de possibilité de prise en compte d'événements provoqués par une intervention. L'opérateur $do(X_i = x_i)$, parfois aussi noté $(X_i = \hat{x}_i)$, crée cette possibilité. L'écriture $do(X_i = x_i)$ est due à Goldzmidt et Pearl (1992).

10. P_{x_I} est donc la distribution de probabilités conditionnelle à l'intervention x_I , soit $P(\cdot | do(X_I = x_I))$.

FIGURE 5.2.3 – Graphes modifiés par une intervention sur X_R 

La distribution observée à l'issue de l'intervention doit être conforme à l'intervention pour les variables directement affectées (point 2), présenter les mêmes dépendances qu'auparavant pour toutes les autres (point 3), et de ce fait, admettre la factorisation (5.2.2) (point 1). Cette dernière se ramène à :

$$P_{x_I}(x_1, \dots, x_n) = \begin{cases} \prod_{i \notin I} P(x_i | x_{\text{Pa}(i)}) & \text{si } x_i \text{ est compatible avec } x_I \forall i \in I \\ 0 & \text{sinon} \end{cases} \quad (5.2.3)$$

Notons que si D représente P_{x_I} , il n'est en revanche pas minimal puisque tout sous-graphe de D où les arcs entre une composante de X_I et ses parents auraient été supprimées continuerait à représenter P_{x_I} .

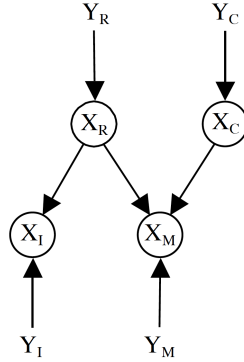
Cette définition nous permet de distinguer le premier diagramme de la figure 5.2.2 de ses trois congénères. Considérons en effet l'intervention sur X_R qui consisterait à imposer à l'usine un niveau de rejets ρ . Dans la distribution de probabilités qu'elle engendrerait sur les variables X_R, X_I, X_M et X_C , cette dernière (seule) aurait une distribution marginale inchangée, alors que les *relations* de dépendance que l'on observerait entre X_I et X_R d'une part, entre X_M et (X_R, X_C) seraient les mêmes que précédemment. Des quatre diagrammes modifiés, le seul qui serait représentatif de la nouvelle distribution serait le premier (voir figure 5.2.3).

Définition 5.10. Soient X_I et X_J les vecteurs correspondants à deux sous-ensembles disjoints de variables. On appelle effet causal de X_I sur X_J , que l'on note $P_{X_I}(X_J)$, la fonction qui attribue à chaque valeur x_I de X_I la distribution de probabilités $P_{x_I}(x_J)$ sur l'ensemble des valeurs x_J de X_J .

5.2.2.2 La causalité comme propriété fondamentale de la Nature

Une seconde approche s'appuie sur la représentation explicite de relations causales entre variables, qui est commune dans la plupart des disciplines scientifiques ou techniques (de la géologie

FIGURE 5.2.4 – Réseau bayésien causal



à l'économétrie) où l'on cherche à rendre compte des valeurs observées d'un « système » et à prévoir ses évolutions, notamment en réponse à des interventions extérieures (Woodward, 2003). Ces modèles causaux peuvent s'écrire sous une forme fonctionnelle générale :

$$X_i = f_i(X_{pa(i)}, Y_i) \quad (5.2.4)$$

où $X_{pa(i)}$ représente un ensemble des variables identifiées comme causes observables de X_i , Y_i est un terme d'erreur agrégeant toutes les causes de X_i qui sont extérieures à X et inobservables, et f_i est la fonction, également appelée *mécanisme*, de détermination de X_i .

Les modèles à équations structurelles linéaires, par exemple, sont un cas particulier de modèles causaux :

$$X_i = \sum_{k \in Pa(i)} \alpha_{ik} X_k + Y_i \quad (5.2.5)$$

Le signe d'égalité dans (5.2.4) et (5.2.5) est remplacé par certains auteurs par un signe d'assignation ($:=$) pour souligner le caractère causal des équations, et plus spécifiquement le fait que la valeur du terme de gauche est déterminée par celle du terme de droite quel que soit le mécanisme selon lequel cette dernière a été déterminée.

Le système X des sections précédentes est ici constitué de l'ensemble des variables endogènes X_i , et son caractère aléatoire résulte de l'action de variables stochastiques exogènes Y_i dans les équations (5.2.4). En reliant chaque élément de $X_{pa(i)}$ à la variable X_i correspondante dans chacune de ces équations et en omettant les variables exogènes, on associe un diagramme D au modèle causal. La figure 5.2.4 représente un tel diagramme dans un cas où notre problème précédent aurait été décrit par un ensemble de relations du même type que (5.2.4).

Formellement, un modèle est défini par la donnée du diagramme D , de la famille f de fonctions

f_i et de la distribution jointe P_Y des variables Y_i :

$$M = \langle D, f, P_Y \rangle$$

Le modèle est dit semi-markovien lorsque D est acyclique. Les X_i sont alors déterminés de façon univoque par les Y_i (autrement dit la loi jointe P est entièrement déterminée par la distribution des erreurs P_Y). Le modèle est dit markovien lorsqu'en outre, les termes d'erreur sont mutuellement indépendants.

Théorème 5.11. *Tout modèle markovien engendre une distribution (unique) P qui est Markov relativement à son diagramme D .*

Démonstration. La distribution marginale de chaque X_i est bien définie en tant que mesure-image de P_{Y_i} par f_i . Il suffit alors de procéder par récurrence en supposant la condition de Markov vérifiée à l'ordre $i - 1$ ($i \geq 2$) et en observant qu'elle découle, pour X_i , de l'indépendance entre Y_i et $(Y_1, \dots, Y_{i-1})$. $\square$

L'indépendance des termes d'erreur, condition critique du théorème, est à rapprocher de l'absence de cause commune latente (c'est-à-dire extérieure au système). Pour clarifier la relation entre ces deux hypothèses, nous devons introduire le principe dit de cause commune, formulé par Hans Reichenbach (Reichenbach, 1956).

Conjecture 5.12. *Principe de cause commune : Soient X_1 et X_2 deux variables telles que $P(X_1, X_2) \neq P(X_1).P(X_2)$. Alors il existe une variable vectorielle X et un graphe D de X tels que X_1 et X_2 sont des composantes de X , d est Markov par rapport à P et au moins l'une des conditions suivantes est remplie :*

- $1 \in De(2)$
- $2 \in De(1)$
- X a une composante X_3 telle que $1, 2 \in De(3)$

où $De(i)$ est l'ensemble des descendants de X_i dans D .

Le principe proclame donc qu'il ne peut y avoir de corrélation en l'absence d'un rapport de cause à effet : si deux variables sont corrélées, alors soit l'une est une cause de l'autre, soit elles ont une cause commune. Si l'on admet ce principe, alors l'indépendance mutuelle des termes d'erreur Y_i découle de l'hypothèse selon laquelle, parmi toutes les causes inobservables des X_i , aucune n'est commune à deux variables ou plus. L'indépendance des termes d'erreur n'implique pas, en retour, l'absence de cause commune latente, puisque, comme nous l'avons vu en introduisant la condition de fidélité, les relations de dépendance vis-à-vis de cette cause peuvent se compenser.

Notons aussi que la condition de Markov n'implique, pour sa part, ni l'indépendance mutuelle des termes d'erreur, ni l'absence de cause commune extérieure. À titre d'exemple, l'existence d'une cause commune latente à une variable et l'un de ses parents conduit à une corrélation des deux termes d'erreur, mais n'entraîne pas de violation de la condition de Markov.

La logique des interventions et la mesure des effets causaux exposés à la section précédente peuvent être reproduits pour les modèles markoviens. Dans un modèle markovien, une intervention du type $do(X_i = x_i)$ consiste à remplacer la fonction f_i par la constante x_i . On peut facilement constater que les trois conditions de la définition d'un réseau bayésien causal donnée à la section précédente sont vérifiées, et donc que le diagramme de tout modèle markovien est un RBC.

Inversement, on peut montrer que toute distribution de probabilités représentée par un réseau bayésien D peut être engendrée par un modèle markovien ayant D pour diagramme.

5.2.2.3 Quelques remarques sur les présupposés des deux approches

Soubassements philosophiques

Les deux approches des RBC se rattachent à deux courants philosophiques distincts et même opposés.

La première suppose l'existence de relations probabilistes entre variables aléatoires, et construit une *interprétation* causale de ces relations. Elle se conforme ainsi à la doctrine empiriste, pour laquelle seules les grandeurs observables sont pertinentes pour la connaissance scientifique : alors que les corrélations entre événements sont des caractéristiques mesurables de la Nature, leur mise en relation causale est un artifice de l'esprit humain¹¹.

Dans l'approche « probabiliste », le lien entre causalité et interventions n'est pas seulement épistémique, mais ontologique. Le rapport causal est celui que l'on constate entre une variable que l'on modifie par intervention et une autre sur laquelle on mesure l'effet de cette intervention. Sans intervention, la notion même de causalité n'existerait pas. En cela, l'approche probabiliste constitue (ou, tout au moins, est compatible avec) une version radicale de la vision interventionniste.

À l'inverse, l'approche « fonctionnelle » des RBC stipule l'existence de relations causales préalablement à toute notion probabiliste. Les probabilités ne sont même introduites que pour pallier l'absence de certaines variables explicatives dans des relations qui seraient entièrement déterministes si toutes les causes pouvaient être observées. Cette approche est donc en ligne avec la vision laplacienne d'un monde déterministe où les probabilités ne reflètent que notre ignorance des lois de la Nature.

De même, le rôle des interventions dans cette approche est seulement d'ordre épistémique. Les relations causales sont des données de la Nature qui nous sont cachées, et que les interventions nous

11. Voir l'introduction générale.

permettent de révéler.

La condition d'autonomie

Dans un cas comme dans l'autre, la définition des RBC s'appuie sur une hypothèse que nous n'avons pas encore discutée : que la structure de dépendances entre variables reste inchangée à la suite d'une intervention (à l'exception, bien sûr, des dépendances qui sont directement supprimées par l'intervention). Ceci suppose que le système représenté par un RBC soit constitué d'un ensemble de mécanismes stables et autonomes (les relations de chaque variable avec ses parents), de sorte que la modification de l'un de ces mécanismes n'affecte pas les autres. C'est à cette condition, que l'on appelle indifféremment modularité ou autonomie, que l'influence d'événements induits du type $do(X_i = x_i)$ sur une variable X_j peut être assimilée à celle d'événements observés du type $(X_i = x_i)$.

La notion d'autonomie trouve ses sources dans les travaux de Frisch et Haavelmo relatifs à l'estimation des modèles à équations structurelles¹² : dans un système de relations structurelles, tout coefficient d'une relation peut être changé *de façon institutionnelle* sans entraîner de modification dans une autre relation. Contrairement aux autres systèmes de relations, dont la validité est limitée aux situations observées, les systèmes de relations structurelles peuvent être extrapolées aux situations où une ou plusieurs variables viendraient à évoluer d'une façon nouvelle. Elles se prêtent donc au raisonnement contrefactuel.

L'hypothèse d'autonomie s'applique donc à un système constitué d'éléments suffisamment disjoints pour pouvoir être considérés séparément les uns des autres, et suffisamment cohérents pour être gouvernés par des mécanismes invariants. Elle suppose que toutes les relations entre deux variables X_i et X_j soient explicitées dans le modèle, ce qui n'est pas garanti par la condition de Markov mais découle, dans le cas de l'approche fonctionnelle, de l'indépendance mutuelle des termes d'erreur. L'hypothèse d'autonomie institue alors un critère de qualité de la spécification d'un modèle causal.

Il faut enfin noter que l'autonomie est une approximation, puisque rien ne permet de s'assurer que deux mécanismes en apparence autonomes ne sont pas liés par une variable latente. La validité pratique de l'hypothèse d'autonomie peut donc toujours prêter à débat. En économie, la célèbre critique de Lucas a été fondamentalement dirigée contre les hypothèses implicites d'autonomie acceptées dans le modèle d'une courbe de Phillips (relation entre inflation et niveau d'activité), et plus généralement dans les modèles à équations structurelles. Pour Lucas et les autres tenants des anticipations rationnelles, les anticipations des agents économiques constituaient la variable latente créant une liaison entre inflation et activité (Lucas, 1976).

12. Sur la genèse et le rôle de l'hypothèse d'autonomie en économétrie, voir Aldrich (1989).

L'interprétation causale de la condition de Markov

Comme nous l'avons vu, les relations déterministes 5.2.4 d'un modèle fonctionnel, associés à des termes d'erreur mutuellement indépendants, engendrent un ensemble de relations d'indépendance conditionnelle et une décomposition de la loi jointe identiques à ceux d'un réseau bayésien, quels que soient les paramètres du modèle (fonctions f_i et distribution P_Y). L'indépendance de X_i par rapport à son prédécesseur X_j conditionnellement à ses parents $X_{pa(i)}$ dans un RBC trouve son équivalent dans l'indépendance *inconditionnelle* entre Y_i et X_j pour $i > j$ (elle-même impliquée par l'indépendance entre Y_i et $(Y_1, \dots, Y_j)$) dans un modèle fonctionnel. Les questions relatives à l'indépendance conditionnelle entre grandeurs observables reviennent ainsi à des interrogations sur l'existence d'une cause commune inobservable (sous réserve d'acceptation du principe de la cause commune).

Il y a cependant deux grands avantages à construire un RBC à partir d'un modèle causal plutôt qu'une distribution de probabilités.

Le premier est qu'une fois les relations de dépendance entre variables spécifiées, les modèles causaux sont beaucoup plus économes en termes d'observations. Pour établir qu'un réseau bayésien est causal, il n'est plus nécessaire, comme dans l'approche précédente, d'observer et de tester les distributions de probabilités associées à toutes les interventions possibles sur les variables du système, mais simplement de s'assurer que les variables exogènes sont mutuellement indépendantes. La connaissance de la loi jointe ne passe plus par celle de n distributions conditionnelles (de chaque X_i par rapport à ses prédécesseurs), mais inconditionnelles (des Y_i). Les effets d'une intervention peuvent être évalués *a priori*, par le biais des relations de dépendance entre la variable sur laquelle on agit et les autres, et il n'est pas nécessaire d'estimer la distribution de probabilités résultante *a posteriori*.

Le second grand avantage des modèles causaux est qu'ils permettent d'élargir considérablement la notion d'intervention, puisque l'on peut considérer comme telle toute modification d'une fonction f_i en g_i , c'est-à-dire tout aménagement par un agent extérieur de la relation entre une variable et ses causes. Cet élargissement s'avère crucial d'une part pour étudier les problèmes d'intervention avec les outils de la théorie de la décision, et de l'autre pour considérer à l'intérieur d'un même cadre d'analyse les démarches déterministe et probabiliste de prise en compte du risque.

5.2.3 Le choix d'un réseau bayésien

5.2.3.1 Principes généraux

Le problème de choix d'une structure causale pour rendre compte d'une situation donnée n'est qu'un cas particulier de celui du choix d'un modèle. Deux critères sont généralement jugés pertinents pour orienter ce choix : l'adéquation des modèles proposés à l'information dont on dispose, mesurée

par leur vraisemblance, et leur simplicité, souvent mesurée par le nombre de paramètres qui sont à estimer à partir de l'information disponible.

Dans la plupart des analyses formelles du problème, l'information considérée consiste en une série d'observations des grandeurs endogènes, que le modèle est censé reproduire ou prévoir. Toute autre connaissance pertinente est supposée neutre à l'égard des différents modèles en présence, et laissée à l'arrière-plan. La vraisemblance conduit alors à préférer les modèles qui accordent la plus grande probabilité *a priori* aux observations disponibles. Ce faisant, elle favorise les spécifications les plus détaillées, puisque la probabilité d'un ensemble d'observations peut toujours être accrue en augmentant le nombre de paramètres libres du modèle¹³. Ceci engendre un risque d'excès d'ajustement (*overfitting*), où, à force de se rapprocher des observations, le modèle estimé finit par s'éloigner du processus qui les a engendrés¹⁴.

Le principe d'induction dit du rasoir d'Occam apporte une réponse pragmatique à cette question en demandant que l'on pondère la recherche de vraisemblance par une préférence pour la simplicité. L'hypothèse, ici, est que la simplicité élargit la portée d'un modèle en limitant la possibilité qu'il soit calibré de façon *ad hoc* pour rendre compte des observations existantes. C'est donc par une double appréciation de la vraisemblance et de la simplicité d'un modèle que l'on juge de sa plausibilité.

Si ces deux notions sont communes à une grande variété de méthodes de sélection de modèles, la façon de les mesurer et de les combiner au sein d'une règle de choix synthétique est, en revanche, propre à chaque méthode. Gilboa et Schmeidler (2010) ont récemment proposé un traitement axiomatique du problème d'un décideur qui doit choisir une théorie à l'intérieur d'un ensemble de possibles en s'appuyant sur une base d'observations. Ils obtiennent une représentation de ses préférences pour les théories par une fonction qui est la somme d'une log-vraisemblance et d'une probabilité *a priori*. Ce dernier terme peut, par hypothèse, refléter la préférence du décideur pour la simplicité. La représentation de Gilboa et Schmeidler généralise un certain nombre de règles de choix existantes.

Dans le cas des réseaux bayésiens, nous considérons un n -uplet de variables $(X_1, \dots, X_n)$ pour lequel on dispose d'un ensemble d'observations e . Le problème peut être séparé en deux composantes : celui du choix de la structure D et, à structure donnée, celui du choix du vecteur de paramètres $\theta = (\theta_1, \dots, \theta_n)$ qui détermine la loi jointe P . Chaque θ_i représente ici un paramétrage de la distribution de probabilités conditionnelles $P(X_i \mid X_{Pa(i)})$, dont l'estimation à partir d'un

13. À l'extrême, un modèle disposant d'autant de paramètres libres qu'il y a d'observations peut avoir une vraisemblance égale à un.

14. Dans le cas d'un RBC fonctionnel, à titre d'exemple, une estimation parfaite des mécanismes f_i doit préserver une différence entre valeurs observées et valeurs estimées de X_i ; ces écarts mesurent les réalisations de la variable résiduelle Y_i .

échantillon d'observations ne pose pas de difficulté¹⁵ :

$$\theta_i = \left(\theta_{ijk} \right)_{j=1, \dots, |\chi_i|; k=1, \dots, |\chi_{Pa(i)}|}$$

$$\theta_{ijk} = P(X_i = x_j \mid X_{Pa(i)} = x_k)$$

Une recherche extensive consiste à parcourir l'ensemble des structures possibles et retenir celle pour laquelle cette estimation produit les résultats les plus satisfaisants¹⁶. Nous en exposons les principes pour l'approche dite classique ci-dessous. Il faut noter, cependant, que cette recherche devient très vite impraticable lorsque que le nombre de variables augmente, en raison de la croissance plus qu'exponentielle du nombre de structures possibles : pour représenter six variables, le nombre de structures envisageables est de 3,8 millions ; pour dix variables, il dépasse $4 \cdot 10^{18}$. Une grande variété de méthodes et d'algorithmes de recherche ont été proposés pour faire face à ce problème, mais il n'entre pas dans notre propos de les exposer ici¹⁷.

Notons aussi que le développement de l'interprétation causale des réseaux bayésiens a été en large part la conséquence de l'élaboration d'algorithmes de construction automatique de structures par identification d'un type particulier de relation d'indépendance conditionnelle entre trois variables, auquel on a *attribué* une signification causale¹⁸ (Pearl, 2000, pp.49 et suivantes). L'un des avantages de ces algorithmes était la réduction drastique du nombre de structures candidates.

5.2.3.2 Approche classique

La fonction de vraisemblance est une notion centrale de l'approche classique. Celle des paramètres θ s'écrit :

$$L(\theta \mid e) = P(e \mid \theta)$$

Si e consiste en un échantillon de N observations indépendantes des valeurs de X notées $(x_1^1, \dots, x_n^1), \dots, (x_1^N, \dots, x_n^N)$, on a :

15. Au moins dans le cas le plus simple où les données sont complètes (c'est-à-dire qu'il n'y a pas d'observation manquante pour un sous-groupe de variables) et où l'on ne cherche pas à rendre compte de l'action de variables latentes.

16. Comme nous l'avons vu, l'unicité de cette structure n'est toutefois pas assurée.

17. On pourra se reporter à Neapolitan (2004) pour une exposition détaillée et à Darwiche (2009, chapitres 17 et 18) pour une vue d'ensemble.

18. Il s'agit des situations où X et Z sont tous deux statistiquement dépendants de Y , et indépendants entre eux, ce que l'on a traduit par des relations causales en forme de fourche inversée $X \rightarrow Y \leftarrow Z$.

$$\begin{aligned}
L(\theta \mid e) &= \prod_{j=1}^N P(X_1 = x_1^j, \dots, X_n = x_n^j \mid \theta) \\
&= \prod_{j=1}^N \prod_{i=1}^n P(X_i = x_i^j \mid X_{Pa(i)} = x_{Pa(i)}^j, \theta_i)
\end{aligned}$$

On démontre que cette fonction atteint son maximum pour une valeur unique θ^* , dite estimateur du maximum de vraisemblance, dont les composantes sont les fréquences conditionnelles observées :

$$\begin{aligned}
\theta^* &= \operatorname{argmax} L(\theta \mid e) \\
&= (\theta_{ijk}^*) \\
\theta_{ijk}^* &= P_e(X_i = x_j \mid X_{Pa(i)} = x_k)
\end{aligned}$$

où P_e est la distribution empirique de X dérivée des observations e .

On peut alors définir la fonction de vraisemblance d'une structure à partir de l'estimateur du maximum de vraisemblance des paramètres qui lui sont associés :

$$L(D \mid e) = L(\theta^* \mid e)$$

La fonction de log-vraisemblance correspondante s'écrit alors :

$$\begin{aligned}
LL(D \mid e) &= \sum_{j=1}^N \sum_{i=1}^n \operatorname{Log} P(X_i = x_i^j \mid X_{Pa(i)} = x_{Pa(i)}^j, \theta_i^*) \\
&= -N \sum_{i=1}^n \operatorname{Ent}_e(X_i \mid X_{Pa(i)})
\end{aligned}$$

où l'entropie de X_i conditionnellement à la valeur de ses parents est définie par :

$$\operatorname{Ent}_e(X_i \mid X_{Pa(i)}) = - \sum_{x_i, x_{Pa(i)}} P_e(x_i, x_{Pa(i)}) \cdot \operatorname{Log} P(X_i = x_i \mid X_{Pa(i)} = x_{Pa(i)}, \theta_i^*)$$

Par ailleurs, le degré de complexité d'une structure est mesuré par le nombre de ses paramètres indépendants :

$$\|D\| = \sum_{i=1}^n \|X_i, X_{Pa(i)}\|$$

où :

$$\|X_i, X_{Pa(i)}\| = (Card\chi_i - 1) \cdot Card\chi_{Pa(i)}$$

En combinant ces deux critères, on se donne une mesure synthétique, ou *score*, de la performance de chaque structure :

$$Score(D | e) = LL(D | e) - \psi(N) \cdot \|D\|$$

La fonction ψ est croissante. Elle est destinée à compenser l'augmentation linéaire de la log-vraisemblance avec N , de sorte que le poids du critère de simplicité dans le score ne diminue pas à mesure que le nombre d'observations augmente. Pour $\psi = cste$, le score correspond au critère d'information d'Akaike (*Akaike Information Criterion*), d'usage courant pour comparer des structures ayant des vraisemblances voisines. Pour $\psi = \frac{LogN}{2}$, le score correspond au critère dit de longueur de description minimale (*Minimum Description Length*) ou au critère d'information bayésien (*Bayesian Information Criterion*).

5.3 Une application à la sûreté

Nous cherchons dans cette partie à construire un cadre formel afin de rendre compte des démarches d'étude et de construction de la sûreté que nous avons décrites dans le cas des centrales nucléaires. Depuis une dizaine d'années, différents travaux ont commencé à explorer les possibilités des réseaux bayésiens dans ce domaine, en montrant notamment qu'ils permettaient de généraliser les méthodes d'analyse de la fiabilité (Bobbio et al., 2001; Langseth et Portinale, 2007). À leur différence, toutefois, nous nous appuyons sur la variante fonctionnelle des réseaux bayésiens causaux, qui nous semble avoir des avantages considérables sur l'approche simplement probabiliste. Certains de ces avantages ont été amplement mis en avant par les promoteurs des RBC fonctionnels¹⁹. Les RBC fonctionnels décrivent les relations entre grandeurs de façon beaucoup plus directe et intuitive que les représentations fondées sur les seules probabilités conditionnelles. En outre, ils permettent d'anticiper l'effet d'un changement spontané ou induit de la valeur de certaines variables, ce que ne fait pas la donnée d'une distribution de probabilités jointe. Les conséquences de tels changements y sont en effet circonscrits, et ne remettent pas en cause la structure du modèle comme elles le font d'une distribution de probabilités dérivée à partir d'observations passées. Nous ne reviendrons pas ici sur ces avantages, mais nous chercherons à en établir deux autres : en premier lieu, l'approche fonctionnelle permet de représenter de façon plus générale et plus précise les interventions sur le système à l'étude ; en second lieu, elle intègre dans un cadre unique les démarches déterministes et probabilistes, et met ainsi en lumière leurs rapports.

19. Voir notamment le premier chapitre de Pearl (2000).

Dans toute la suite, nous considérons que le problème de sûreté est décrit par un système $X = (X_1, \dots, X_n)$, pour lequel on dispose d'un modèle markovien $M = \langle D, f, P_Y \rangle$. D est un graphe du même type que celui de la figure 5.2.4 page 173, et la famille de fonctions f vérifie :

$$X_i = f_i(X_{Pa(i)}, Y_i) \quad i = 1, \dots, n \quad (5.3.1)$$

où $X_{Pa(i)}$ est le vecteur des parents de X_i dans D et les Y_i sont des variables aléatoires mutuellement indépendantes dont les distributions sont données par P_Y . X_i et Y_i prennent leurs valeurs respectives dans les ensembles χ_i et $\mathcal{Y}_i$, que nous supposons finis²⁰.

Nous appelons *état du système* chaque valeur du n-uplet $x = (x_1, \dots, x_n)$, et *état du monde* chaque valeur du n-uplet $y = (y_1, \dots, y_n)$. Nous définissons les ensembles $\chi = \bigotimes_{i=1, \dots, n} \chi_i$ et $\mathcal{Y} = \bigotimes_{i=1, \dots, n} \mathcal{Y}_i$ des états du système et du monde, et nous donnons à leurs sous-ensembles respectifs les noms d'*événements internes* et d'*événements externes*. Dans la suite, il nous arrivera, par abus de langage, d'utiliser les dénominations d'événement interne ou externe pour des ensembles $\varepsilon_i \subset \chi_i$ ou $\eta_i \subset \mathcal{Y}_i$; il s'agira en fait des événements $\{(x_1, \dots, x_n) \mid x_i \in \varepsilon_i\}$ et $\{(y_1, \dots, y_n) \mid y_i \in \eta_i\}$.

Comme nous l'avons vu, la loi jointe P_Y des variables exogènes engendre une distribution de probabilités sur les états du système qui vérifie la condition de Markov relativement à D :

$$P(x) = P(x_1, \dots, x_n) = \prod_{i=1}^n P(x_i \mid x_{Pa(i)})$$

Nous supposons que le bon fonctionnement de l'installation est représenté par un événement interne $\chi^s \subset \chi$ et qu'un décideur cherche à assurer sa réalisation afin d'éviter un événement redouté χ^e ($\chi^s \cap \chi^e = \emptyset$).

5.3.1 Extension de la notion d'intervention

Nous avons introduit plus haut les interventions particulières de type $do(X_i = x_i)$, qui suspendent les liens entre X_i et ses parents et lui attribuent la valeur x_i . Nous allons à présent considérer une classe plus large d'interventions.

La notion d'intervention a été formulée à l'origine dans le cadre de modèles structurels. Marschak proposait de représenter ainsi les mesures de politique économique, en distinguant les politiques structurelles, qui changent une relation entre variables économiques, et politiques non structurelles, qui fixent la valeur d'une variable exogène (Marschak, 1950). Le terme intervention a probablement été utilisé la première fois pour désigner une modification exogène d'une relation structurelle par Simon, également dans un sens très général (Simon, 1953). L'approche plus restrictive qui consiste à ne considérer comme intervention que l'affectation d'une valeur constante à une variable est

20. Voir la note 1.

plus récente (Strotz et Wold, 1960), mais elle a rapidement gagné en popularité (Fisher, 1970; Sobel, 1990; Spirtes, Glymour and Scheines, 1993), notamment en raison de sa simplicité²¹. Dans l'approche « probabiliste » des RBC, en particulier, la distribution de probabilités résultant d'une intervention sur la variable X_i ne peut être dérivée de la distribution initiale P que si l'on dispose de la nouvelle distribution marginale de X_i et si le support de cette dernière est inclus dans le support de la distribution initiale. Le choix, à travers l'intervention $do(X_i = x_i)$, d'une distribution marginale dégénérée simplifie ce problème²². Dans l'approche « fonctionnelle », comme nous allons le voir, cette difficulté ne se pose pas.

Nous proposons donc de revenir à la notion générale d'intervention, qui s'avère beaucoup plus riche dans notre cadre, et de considérer comme telle toute modification de la relation (5.3.1) entre une variable et ses causes. L'intervention sur la variable X_i peut alors être représentée comme valeur spécifique d'une *variable d'intervention* G_i au sein d'une structure causale « augmentée » (Pearl, 1993). Nous pouvons ainsi considérer, en chaque nœud de la structure, une variable d'intervention dont chaque réalisation $G_i = g_i$ détermine une relation $x_i = g_i(x_{Pa(i)}, y_i)$ (voir la figure (5.3.1)).

Définition 5.13. Soit $\mathcal{G}$ un ensemble de vecteurs de mécanismes tel que pour tout $g \in \mathcal{G}$, $M = \langle D, g, P_Y \rangle$ soit un modèle markovien pour le système X . On appelle intervention le choix d'un vecteur particulier de mécanismes g . On appelle structure causale augmentée un graphe D^+ obtenu en associant à chaque nœud de D une variable d'intervention G_i et $M^+ = \langle D^+, \mathcal{G}, P_Y \rangle$ un modèle markovien augmenté pour le système X .

Le vecteur de mécanismes préexistants f représente alors une intervention particulière que nous appellerons *laisser-faire*. Nous nous autoriserons à parler d'intervention g_i relative à (ou sur) une variable X_i au lieu de l'intervention $(f_1, \dots, f_{i-1}, g_i, f_{i+1}, \dots, f_n)$.

L'ensemble $\mathcal{G}$ permet de rendre compte de contraintes exogènes²³ pesant sur l'intervention. Dans les cas où aucune intervention au sens propre n'est possible en un nœud i , par exemple, la variable d'intervention G_i est dégénérée et prend pour seule valeur f_i .

Cette structure augmentée peut être considérée selon deux points de vue : celui du décideur qui détermine les interventions ou celui d'un observateur extérieur.

5.3.1.1 Point de vue du décideur

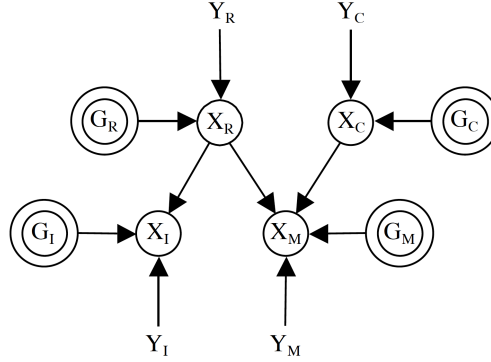
Le décideur n'a pas d'incertitude au sujet de ses propres choix, mais seulement à l'égard de leurs conséquences : il considère les alternatives à sa disposition et en choisit une en fonction des effets qu'il en attend. Ce point de vue correspond au cadre de la théorie de la décision, dans lequel nous

21. Pour Pearl, un grand avantage de cette forme est son économie, puisqu'il suffit d'une valeur pour la décrire entièrement (communication personnelle).

22. Voir la définition 5.9 et l'équation 5.2.3 page 172.

23. On entend par là des contraintes qui ne seraient liées à aucune variable du système, faute de quoi elles seraient explicitées par les relations fonctionnelles constitutives du réseau.

FIGURE 5.3.1 – Réseau bayésien causal augmenté de variables de décision



nous situerons au cours des chapitres suivants. Ici, il permet de rendre compte des modifications de sûreté.

Pour le décideur, les variables d'intervention G_i ne sont donc pas des variables aléatoires (ce que reflète leur représentation par un double cercle dans la figure (5.3.1)). À chaque intervention g correspondent une structure D_g et une distribution P_g qui vérifie :

$$P_g(x_1, \dots, x_n) = \prod P_g(x_i \mid x_{Pa(i)})$$

où :

$$P_g(x_i \mid x_{Pa(i)}) = \sum_{y_i} P_Y(y_i) \quad \text{pour } (y_i, x_{Pa(i)}) \in g_i^{-1}(x_i)$$

En particulier, si l'intervention porte sur la seule variable X_i et laisse inchangées $n - 1$ relations (5.3.1), P_g admet la factorisation :

$$P_g(x_1, \dots, x_n) = P_g(x_i \mid x_{Pa(i)}) \cdot \prod_{j \neq i} P(x_j \mid x_{Pa(j)})$$

où l'on retrouve l'expression habituelle de la transformation de P par une intervention (voir l'équation (5.2.3)).

5.3.1.2 Point de vue d'un observateur extérieur

Pour un observateur extérieur, au contraire, les choix du décideur constituent une source d'incertitude qui vient s'ajouter aux autres données du problème. L'observateur peut constituer à leur sujet des croyances probabilistes. Sous leur forme la plus basique, ces croyances seront dérivées d'une représentation purement aléatoire, mais il serait également concevable pour l'observateur d'identifier et d'intégrer à son raisonnement des déterminants du choix du décideur tels que sa

psychologie ou le contexte dans lequel il est placé, en augmentant ainsi le RBC. Ce point de vue correspond au cadre de la théorie de l'agence, où un principal considérerait les choix possibles d'un agent. Il permet de représenter les considérations d'erreur humaine et de succès ou d'échec des missions de conduite, qui constituent un élément critique des évaluations probabilistes de la sûreté des centrales nucléaires²⁴.

Formellement, la structure augmentée est associée à une distribution de probabilités P' obtenue en combinant la distribution P_g relative à chaque intervention particulière avec une distribution a priori P_G du vecteur des variables d'intervention $G = (G_1, \dots, G_1)$:

$$\begin{aligned} P'(x_1, \dots, x_n, g_1, \dots, g_n) &= \sum_{g \in \mathcal{G}} P'(x_1, \dots, x_n, g_1, \dots, g_n \mid G = g) \cdot P'(G = g) \\ &= \sum_{g \in \mathcal{G}} P_g(x_1, \dots, x_n) \cdot P_G(g) \end{aligned}$$

5.3.2 L'approximation déterministe

Nous cherchons à présent à reconstituer à partir de ce modèle les grandes lignes de l'approche déterministe de la sûreté présentée au chapitre 3. Notre visée étant uniquement illustrative, il va de soi que nous privilégierons ici la simplicité par rapport au réalisme.

La démarche déterministe consiste à simuler le comportement du système par le biais d'une représentation simplifiée, où la prise en compte des variables aléatoires exogènes est repoussée aux frontières du système et où, hormis les variables initiatrices, chaque variable X_i du système ne dépend que d'un petit ensemble de causes identifiées²⁵, représenté par le vecteur $X_{Pa^0(i)}$. Les relations 5.3.1 de la section précédente s'écrivent alors sous la forme :

$$\begin{cases} X_i = g_i(Y_i) & \text{si } i \in In \\ X_i = g_i(X_{Pa^0(i)}) & \text{sinon} \end{cases} \quad (5.3.2)$$

L'objectif est de s'assurer du bon fonctionnement du système sous l'intervention g considérée. Les moyens sont l'intégration de marges de sûreté et les hypothèses pénalisantes, qui doivent permettre compenser l'approximation déterministe.

5.3.2.1 Les marges de sûreté

En fixant la valeur des variables initiatrices dans le système pseudo-déterministe d'équations (5.3.2), le décideur établit une estimation *a priori* de l'état du système. L'intégration de marges de

24. Voir en particulier le premier paragraphe de la sous-section 3.4.2.1 page 93.

25. Cet ensemble est en général plus restreint que dans l'approche probabiliste, où l'effort d'inférence est plus important.

sûreté consiste, en premier lieu, à se donner, pour chacune des variables initiatrices, une valeur $\bar{x}_i$ qu'il estime prudente. Pour porter une telle appréciation, le décideur doit disposer d'une distribution de probabilités sur les valeurs de X_i . Dans le domaine de la sûreté nucléaire, comme nous l'avons vu dans la sous-section 3.3.3, il exclut les événements qu'il juge trop improbables. À titre d'exemple, si X_i est le mouvement au sol associé à l'occurrence d'un séisme, $\bar{x}_i$ peut être la valeur-limite correspondant à un événement millénal :

$$P_g(x_i > \bar{x}_i) = \alpha$$

avec $\alpha = 0,001$. Ces valeurs sont alors transmises au reste du réseau. Lorsque tous les parents d'une variable X_j ont été déterminés, on obtient :

$$\bar{x}_j = g_j^0(\bar{x}_{Pa^0(j)})$$

où g_j^0 est une estimation prudente de g_j retenue par le décideur pour compenser l'absence d'un terme aléatoire et le faible nombre de parents de X_i . Il peut s'agir, par exemple, d'un calcul majorant des contraintes exercées sur les différentes structures suite à la survenue d'une accélération au sol $\bar{x}_i$.

Une fois l'ensemble des valeurs déterminées²⁶, on vérifie que $(\bar{x}_1, \dots, \bar{x}_n) \in \chi^s$.

Notons que si l'on peut calculer la marge de sûreté pour les variables initiatrices (par exemple comme la différence entre $\bar{x}_i$ et la moyenne de la distribution $P_g(x_i)$), ce n'est pas le cas pour les autres variables, en raison du cumul d'estimations conservatrices.

5.3.2.2 Les hypothèses pénalisantes

Le second mode de raisonnement dans une approche déterministe de la sûreté est de vérifier la robustesse du bon fonctionnement à un dépassement des marges. En d'autres termes, on suppose que pour une ou plusieurs variables, on a $x_i > \bar{x}_i$, et l'on réestime la valeur des autres variables sous cette hypothèse.

La valeur $(\bar{x}_1, \dots, \bar{x}_n)$ obtenue, où les valeurs ayant fait l'objet d'hypothèses pénalisantes sont réajustées à la hausse, détermine le *niveau de dimensionnement* de l'installation (en supposant bien sûr elle correspond à un bon fonctionnement).

On sait alors, *sous réserve de monotonie des fonctions g_i et de justesse des marges calculées*, que :

$$P_g(x \in \chi^s) > \alpha$$

26. La notion de marge de sûreté, au sens traditionnel, ne s'applique qu'aux variables pour lesquelles une métrique peut être définie. La représentation de la sûreté comporte également, nous l'avons vu, des variables qualitatives telles que « l'opérateur chargé de la manutention fait chuter une barre de combustible ». Ici, nous traitons de ces variables dans le cas général, en posant que les marges sont nulles dans leur cas et en notant qu'elles font, en revanche, fréquemment l'objet d'hypothèses pénalisantes (voir la sous-section suivante).

Mais naturellement, cette évaluation est incertaine.

5.3.3 L'identification des causes

Nous abordons à présent les deux méthodes à la base de l'évaluation probabiliste de la sûreté, c'est-à-dire l'identification des causes, puis, dans la section suivante, celle des conséquences.

5.3.3.1 Résultat général

Considérons à présent un événement interne ε_i auquel nous portons un intérêt particulier. Il est possible d'utiliser notre modèle causal pour caractériser les conditions de réalisation de ε_i d'abord à partir des parents de X_i et de la variable résiduelle Y_i , puis en progressant pas à pas vers l'amont de la représentation causale. De façon générale, cette identification des causes de ε_i peut s'étendre à tout sous-ensemble des ascendants de X_i qui soit cohérent avec le graphe D au sens de la définition suivante.

Définition 5.14. On dit qu'un sous-graphe D_0 de D est cohérent avec D pour l'identification des causes d'un événement interne $\varepsilon_i \subset \chi_i$ si D_0 est constitué d'un ensemble de nœuds représenté par l'ensemble d'indices J_0 et par tous les arcs qui relient ces nœuds dans D , et :

- $\{i\} \in J_0$
- $J_0 \setminus \{i\} \subset As(i)$
- $\forall j \in J_0 \quad En(j) \cap As(i) \subset J_0$.

On exige donc de D_0 qu'il soit constitué de X_i et d'un groupe de ses ascendants, et qu'il ne contienne pas d'ascendant X_j de X_i sans contenir aussi ceux parmi ses enfants qui sont ascendants de X_i . Cette dernière condition permet d'assurer qu'une part de l'effet causal de X_j sur X_i ne s'échappe pas à l'extérieur de D_0 .

Sous ces conditions, le processus d'exploration permet de dégager deux groupes de causes de ε_i qui peuvent être qualifiées d'ultimes pour le sous-graphe considéré : des causes externes liées à chacun des nœuds visités, et des causes internes définissant les limites²⁷ du sous-graphe. Le théorème suivant établit formellement ce résultat.

Théorème 5.15. Si $\varepsilon_i \subset \chi_i$ est un événement interne, D_0 un sous-graphe de D cohérent pour l'identification des causes de ε_i , J_0 l'ensemble d'indices des nœuds de D_0 et J_1 l'ensemble d'indices des nœuds de D qui sont parent d'une racine de D_0 (ou plus), alors il existe un entier h , des sous-ensembles $\varepsilon_j^1, \dots, \varepsilon_j^h$ de χ_j pour chaque $j \in J_1$ et des sous-ensembles $\eta_j^1, \dots, \eta_j^h$ de $\mathcal{Y}_j$ pour chaque $j \in J_0$ tels que :

27. Plus précisément les racines.

- pour tout $k \leq h$, la conjonction $\bigwedge_{j \in J_1} (x_j \in \varepsilon_j^k) \wedge \bigwedge_{j \in J_0} (y_j \in \eta_j^k)$ est une condition suffisante de la réalisation de ε_i ;
- pour l'ensemble des $k \leq h$, ces conditions sont mutuellement exclusives ;
- la disjonction $\bigvee_{k=1, \dots, h} \left[\bigwedge_{j \in J_1} (x_j \in \varepsilon_j^k) \wedge \bigwedge_{j \in J_0} (y_j \in \eta_j^k) \right]$ est une condition nécessaire de la réalisation de ε_i .

Démonstration. Le résultat s'obtient facilement par une récurrence sur le nombre de nœuds explorés, c'est-à-dire le cardinal de l'ensemble J_0 .

Considérons le cas $|J_0| = 1$, soit $J_0 = \{i\}$ et $J_1 = Pa(i)$. L'image inverse de ε_i par g_i est un sous-ensemble de l'espace produit $\chi_{Pa(i)} \times \mathcal{Y}_i$ qui peut toujours être écrit comme union finie d'ensembles disjoints de la forme $\bigotimes_{j \in Pa(i)} \varepsilon_j \times \eta_i$, où $\varepsilon_j \subset \chi_j$ et $\eta_i \subset \mathcal{Y}_i$ puisque, au minimum, chacun de ses éléments constitue un singleton de cette forme. Il existe donc un h tel que :

$$f_i^{-1}(\varepsilon_i) = \bigcup_{k=1}^h \left(\bigotimes_{j \in Pa(i)} \varepsilon_j^k \times \eta_i^k \right) \quad (5.3.3)$$

Les trois propositions du théorème découlent immédiatement de cette décomposition.

Supposons à présent que le résultat du théorème soit établi à l'ordre $s - 1$. Considérons un sous-graphe D_0 contenant s nœuds et cohérent avec D , les ensembles J_0 et J_1 associés et une racine de D_0 dont on notera l'indice j_0 . Alors le sous-graphe D'_0 obtenu en amputant D_0 du nœud X_{j_0} est cohérent avec D . En définissant les ensembles J'_0 et J'_1 comme précédemment, on sait qu'il existe un entier h' et des ensembles $\varepsilon_j^{k'}$ et $\eta_j^{k'}$ tels que :

$$x_i \in \varepsilon_i \Leftrightarrow \exists k' \leq h' \begin{cases} \forall j \in J'_1 & x_j \in \varepsilon_j^{k'} \\ \forall j \in J'_0 & y_j \in \eta_j^{k'} \end{cases}$$

En isolant dans la proposition de droite le terme $\varepsilon_{j_0}^{k'}$ et en lui appliquant la décomposition (5.3.3), on sait qu'il existe, pour chaque $\varepsilon_{j_0}^{k'}$, un entier $h_0(k')$ et des ensembles $\varepsilon_j^{k_0}(k')$ et $\eta_{j_0}^{k_0}(k')$ tels que :

$$x_i \in \varepsilon_i \Leftrightarrow \exists k' \leq h' \begin{cases} \forall j \in J'_1 \setminus \{j_0\} & x_j \in \varepsilon_j^{k'} \\ \forall j \in J'_0 & y_j \in \eta_j^{k'} \\ \exists k_0 \leq h_0(k') & \forall j \in Pa(j_0) \quad x_j \in \varepsilon_j^{k_0}(k') \\ & y_{j_0} \in \eta_{j_0}^{k_0}(k') \end{cases}$$

Notons que $J_0 = J'_0 \cup \{j_0\}$ et $J_1 = J'_1 \setminus \{j_0\} \cup Pa(j_0)$. Posons $h = \sum_{k'=1}^{h'} h_0(k')$ et pour $k \leq h$,

définissons des ensembles ε_j^k et η_j^k à partir des ensembles précédents, le seul cas non trivial étant celui où $j \in J'_1 \cap Pa(j_0)$, pour lequel on définit : $\varepsilon_j^k = \varepsilon_j^{k'} \cap \varepsilon_j^{k_0}(k')$. On a alors :

$$x_i \in \varepsilon_i \Leftrightarrow \exists k \leq h \quad \forall j \in J_1 \quad x_j \in \varepsilon_j^k \quad \forall j \in J_0 \quad y_j \in \eta_j^{k'}$$

En outre, les propositions du terme de droite sont mutuellement exclusives pour les différents k par construction. $\square$

Le théorème énonce formellement la condition INUS²⁸ exposée par Mackie en 1965 et qui a trouvé depuis des applications dans de multiples domaines²⁹ : la condition portant sur un ascendant X_j ou un terme résiduel Y_j constitue une partie insuffisante mais nécessaire d'une condition portant sur l'ensemble des variables $((X_j)_{j \in J_1}, (Y_i)_{i \in J_0})$, elle-même non nécessaire mais suffisante à la réalisation de la condition finale sur X_i (Mackie, 1965). On peut résumer ce résultat par la formule suivante : pour éviter la survenue de l'événement interne ε_i , il faut et il suffit que pour chaque k , l'un des événements internes ε_j^k ou externes η_j^k ne se produise pas.

Deux cas sont d'un intérêt particulier : lorsque $J_0 = \{i\}$, les causes identifiées sont les causes immédiates de ε_i , définies à partir des parents de X_i et du seul terme résiduel Y_i (voir l'équation (5.3.3)) ; lorsque $J_0 = \{i\} \cup As(i)$, l'ensemble des variables pertinentes du réseau a été intégré, et les causes identifiées portent sur les seules réalisations des termes résiduels Y_j . On en déduit le corollaire suivant.

Corollaire 5.16. *Tout événement interne ε_i a pour cause ultime nécessaire et suffisante une union finie d'intersections d'événements externes.*

5.3.3.2 Application aux arbres de défaillance

Cette logique d'identification des causes est exactement celle des arbres de défaillances, dont nous avons exposé les principes plus haut³⁰ : décrire un événement interne (l'événement de tête) à partir de causes que l'on choisit de considérer élémentaires (les événements de base), en passant par des causes médiatrices (les événements intermédiaires).

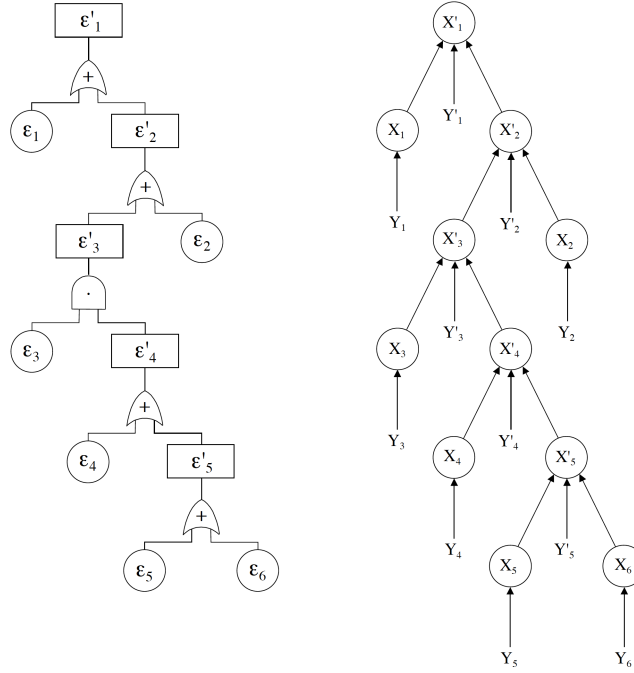
La figure 5.3.2 présente l'arbre de défaillance de la figure 3.4.2 page 87 accompagné du RBC correspondant. Dans celui-ci, la variable X'_1 décrit l'état de la cuve après le démarrage de la pompe ; elle engendre l'événement de tête de l'arbre de défaillance. Les variables X_1 à X_6 correspondent aux événements de base ; elles décrivent, dans l'ordre, l'état de la cuve, du second relais, du pressostat, de l'interrupteur, du premier relais et de la minuterie dans des conditions normales de fonctionnement. Les variables X'_2 à X'_5 correspondent, de la même façon, aux événements intermédiaires. Les

28. Acronyme de *insufficient but necessary part of a unnecessary but sufficient condition*.

29. Par exemple, en épidémiologie, dans l'article fondateur de Rothman (1976).

30. Voir la sous-section 3.4.1.1.

FIGURE 5.3.2 – RBC de la défaillance d'une cuve pressurisée



variables sont booléennes ; elles prennent les valeurs 0 en cas de fonctionnement normal et 1 en cas de défaillance. Les variables résiduelles Y_i (ou Y'_i) agrègent dans tous les cas l'action des causes de défaillance autres que celles identifiées. Les fonctions f_i (ou f'_i) sont toutes des sommes booléennes $Y_i + \sum_{j \in Pa(i)} X_j$, représentées par une porte *OU* dans l'arbre de défaillance, hormis f'_3 qui s'écrit $Y_i + \prod_{j \in Pa(i)} X_j$, le produit entre X_3 et X'_4 correspondant à la porte *ET* entre les événements $S3$ et $E4$ de l'arbre.

Il suffit alors d'appliquer le théorème 5.15 pour caractériser l'événement $\varepsilon'_1 = \{x'_1 = 1\}$ à partir d'événements intervenant en amont dans la structure causale, notamment, dans les conditions du corollaire 5.16, à partir des seuls événements externes η_i et η'_i . Cela permet de discerner l'ensemble des combinaisons qui peuvent conduire à la défaillance, et d'en calculer la probabilité $P(x'_1 = 1)$ à l'aide de la distribution jointe des Y_i . La quantification des arbres de défaillance consiste précisément en la mise en œuvre de ces étapes.

$$\begin{aligned}
 \varepsilon'_1 &= \varepsilon_1 \cup \varepsilon'_2 \cup \eta'_1 \\
 &= \eta_1 \cup \eta_2 \cup (\eta_3 \cap \eta_4) \cup (\eta_3 \cap \eta_5) \cup (\eta_3 \cap \eta_6) \\
 &\quad \cup \eta'_1 \cup \eta'_2 \cup \eta'_3 \cup (\eta_3 \cap \eta'_4) \cup (\eta_3 \cap \eta'_5) \cup (\eta_3 \cap \eta'_6)
 \end{aligned}$$

La représentation en RBC fournit un cadre informatif pour l'analyse des hypothèses de construction et de quantification des arbres de défaillance bien connues des analystes de la sûreté.

En premier lieu, il faut noter l'approximation que constitue l'emploi de variables binaires pour mesurer l'état des éléments du système. À titre d'exemple, même défaillante, une composante peut rester disponible en partie ou par intermittence. Dans la mesure où toute indisponibilité partielle serait considérée comme une défaillance, la forme binaire correspondrait alors à un parti pris pénalisant.

En second lieu et surtout, les relations d'indépendances conditionnelles traduites par la structure du graphe reposent, comme nous l'avons vu, sur l'hypothèse d'indépendance mutuelle des termes d'erreur Y_i . Cette dernière suppose qu'il n'existe pas d'autre interaction entre les X_i que celles répertoriées par les fonctions f_i et qu'il n'existe pas, en dehors du système, de cause commune à deux X_i ou plus.

Troisièmement, les seuls termes aléatoires retenus dans les arbres de défaillance sont ceux relatifs aux événements primaires, c'est-à-dire dans notre cas les Y_i . On fait par là l'hypothèse que les termes résiduels Y'_i sont identiquement nuls et que les causes représentées dans l'arbre sont exhaustives. L'usage est alors de représenter l'événement de tête à partir des combinaisons d'événements de base, que l'on nomme *termes minimaux* (*mintérms*) et dont la simplification fournit les *coupes minimales* (*minimum cutsets*) de l'événement de tête.

5.3.4 L'identification des conséquences

Nous nous limitons ici à la question de l'identification des conséquences d'un événement interne ε_i telle qu'elle est abordée dans les arbres d'événements des évaluations probabilistes de sûreté³¹.

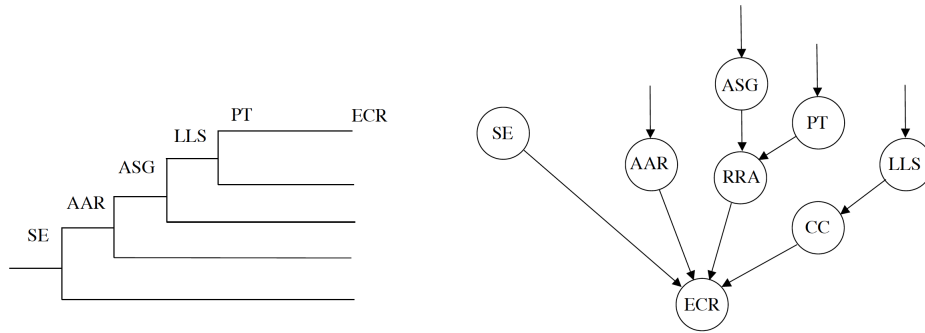
Le formalisme des RBC permet de représenter la démarche des arbres d'événements et ses conditions de validité de façon intuitive : il s'agit de parcourir une branche de descendants de X_i menant à l'événement redouté χ^e , en évaluant à chaque nœud la probabilité de survenue d'un événement ε_j (succès d'une mission de sauvegarde) permettant de prévenir la réalisation de χ^e .

L'arbre d'événement et le RBC de la figure 5.3.3, à titre d'exemple, correspondent à la séquence accidentelle décrite dans la sous-section 3.4.1.1³². La variable SE y désigne les sources électriques dont la défaillance totale initie la séquence. La variable ECT représente l'état du cœur du réacteur, dont on redoute la valeur « fusion ». Pour l'éviter, la conjonction de trois événements a été jugée nécessaire : l'arrêt d'urgence du réacteur AAR ; le refroidissement du réacteur à l'arrêt RRA , dont les deux conditions sont le fonctionnement de l'alimentation de secours des générateurs de vapeur ASG et celui de la pompe de test PT ; enfin, le maintien du contrôle-commande CC , qui repose sur

31. Dans les chapitres 7 et 8, nous reviendrons de façon plus précise sur la prise en compte des conséquences dans un RBC dans le cadre de la théorie de la décision.

32. On a supprimé ici la dimension temporelle afin de simplifier le graphe.

FIGURE 5.3.3 – RBC de la perte des alimentations électriques



le groupe turbo-alternateur *LLS*. Les événements considérés portent sur les variables *ASG*, *PT* et *LLS*. Les flèches pointées vers ces variables indiquent qu'elles ont des ascendants dans le graphe. Les probabilités des événements sont alors évaluées par une inférence des causes selon la méthode que nous venons de détailler, et la construction de l'arbre probabiliste ne pose aucune difficulté.

Seuls deux aspects nous semblent appeler un commentaire.

En premier lieu, les probabilités considérées sont conditionnelles à une certaine réalisation de *SE*. La représentation graphique du RBC présente de grands avantages pour vérifier l'indépendance de *ASG*, *PT* et *LLS* vis-à-vis de *SE*.

En second lieu, dans le cas où la variable initiatrice aurait une autre branche descendante menant à un événement redouté, l'usage veut que la séquence correspondante soit associée à un arbre d'événements séparé. Il est crucial, toutefois, de vérifier l'indépendance des deux séquences. Dans ce cas également, la forme graphique s'avère particulièrement avantageuse.

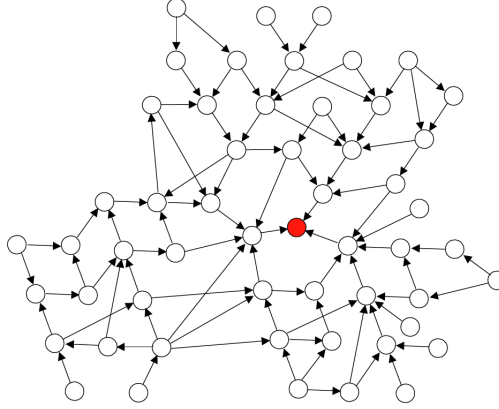
5.3.5 Une définition du risque

Une fois enrichi d'une conception plus générale des interventions, le formalisme des RBC permet donc de reproduire aisément l'ensemble des outils et raisonnements de sûreté dans un cadre unifié : changements de l'installation par le biais de modifications de sûreté, évaluation probabiliste des interventions humaines, marges de sûreté et hypothèses pénalisantes, identification des causes et des conséquences. Nous proposons à présent une synthèse des démarches probabiliste et déterministe de la sûreté comme stratégies d'atteinte d'une certaine cible (l'ensemble d'états χ^s) par l'étude et la modification d'un réseau partiellement identifié.

5.3.5.1 Un RBC idéal

En élaborant des modèles, ou représentations simplifiées d'un système, nous aspirons à une connaissance toujours croissante des grandeurs considérées et de leur rapport au reste du monde, qui puisse, à la limite, nous conduire à une description parfaite. Un tel idéal laplacien ne serait cependant

FIGURE 5.3.4 – Graphe idéal du problème de sûreté



pas déterministe, puisqu'il ne s'étendrait pas à l'ensemble du monde. Les facteurs qu'il ignorerait de ce fait formeraient, ensemble, sa part aléatoire. Mais cet aléa serait caractérisé, limité dans son ampleur et stable dans le temps, à la manière d'un bruit blanc dans un modèle économétrique. Cette idée est l'abstraction d'une abstraction, mais elle nous est utile en tant que point de référence. Elle est, nous semble-t-il, le « vrai modèle » par rapport auquel nous nous situons lorsque nous considérons l'incertitude d'une modélisation.

Supposons donc qu'un tel modèle markovien existe dans le cas de notre problème de sûreté, que son graphe soit celui de la figure 5.3.4 et qu'après l'ensemble des interventions du décideur, ses équations soient de la forme :

$$X_i = g_i^*(X_{Pa^*(i)}, Y_i^*)$$

Supposons enfin, en simplifiant le problème de la sûreté à des fins d'illustration, que l'ensemble d'états χ^s soit caractérisé par les valeurs d'une seule variable, indiquée par un cercle plein sur la figure. Dans le cas de la sûreté nucléaire, il pourrait s'agir de l'état du cœur du réacteur.

Les démarches déterministe et probabiliste constituent alors deux stratégies d'étude et de modification d'une partie identifiée de ce RBC idéal.

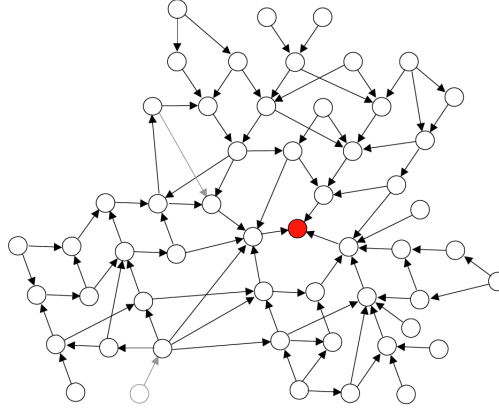
5.3.5.2 L'approche probabiliste

L'approche probabiliste consiste à chercher à identifier le réseau de façon systématique et à estimer l'ensemble d'équations :

$$X_i = g_i(X_{Pa(i)}, Y_i)$$

qui conduit à une mesure naturelle du risque.

FIGURE 5.3.5 – Démarche probabiliste de sûreté



Définition 5.17. On appelle risque sous l'intervention g la donnée de l'ensemble des états du monde accompagnés de leurs probabilités de réalisation et de l'ensemble des états du système accompagnés de leurs probabilités de réalisation sous g :

$$R_g = \langle \mathcal{Y}, P_Y, \chi, P_g \rangle$$

En particulier, on appelle risque *a priori* le risque sous le laisser-faire f .

Le bon fonctionnement d'une installation, ou plus généralement la réalisation d'un ensemble d'états que l'on se donne pour cible, représente un événement interne $\chi^s \subset \chi$. Le risque acceptable, et l'objectif de sûreté qui est sa différence par rapport à 1, portent sur la probabilité de cet événement.

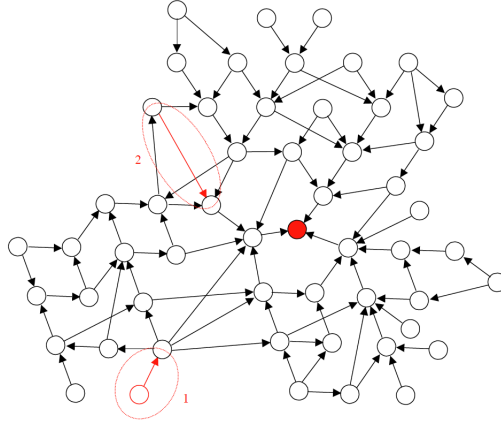
Définition 5.18. On appelle risque acceptable pour la réalisation d'un événement χ^s un réel $\alpha^s \in [0, 1]$, et l'on dit alors qu'une intervention g est acceptable si $P_g(x \in \chi^s) \geq 1 - \alpha^s$.

Le risque, tel que nous venons de le définir, est une mesure de l'incertitude aléatoire³³ portant sur les différentes grandeurs du système. Elle ignore cependant une part de l'incertitude, que l'on appelle épistémique. L'incertitude épistémique est elle-même de deux formes : elle peut, d'une part, porter sur la forme de la fonction f_i ou celle de la distribution de probabilités associée à la variable Y_i , qui sont les paramètres du modèle markovien ; elle peut, d'autre par, être relative à la structure causale du modèle, c'est à dire à l'ensemble d'indices $Pa(i)$ des parents de X_i .

L'incertitude paramétrique peut être mesurée par des distributions de probabilités et « propagée » dans la structure, ce qui revient formellement à assimiler les paramètres incertains à des variables et les intégrer à la structure (à l'image de la variable manquante (1) du graphe de la figure 5.3.6). Comme nous l'avons dit auparavant, la distinction entre incertitude aléatoire et incertitude

33. Voir la sous-section 3.4.2.2.

FIGURE 5.3.6 – Incertitudes de la démarche probabiliste



paramétrique, bien qu'utile sur un plan pratique, ne nous semble pas reposer sur un fondement théorique solide. Nous préférons donc considérer que cette part d'incertitude est déjà intégrée à la mesure du risque, c'est-à-dire que toutes les grandeurs incertaines observables ont été intégrées au système.

Il reste alors l'incertitude de modélisation, c'est-à-dire celle portant sur la structure causale du modèle (à l'image de l'arc manquant (2) de la figure), qui peut s'écrire sous la forme d'un vecteur $(\mu_1, \dots, \mu_n)$, avec :

$$\mu_i = g_i(x_{Pa(i)}, y_i) - g_i^*(x_{Pa^*(i)}, y_i^*)$$

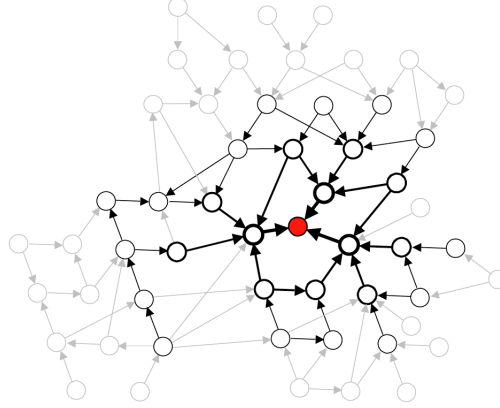
On peut certes imaginer procéder ici de la même façon pour l'incertitude paramétrique, en attribuant des probabilités *a priori* à différentes structures causales en présence. Cette procédure ne peut toutefois pas épuiser notre incertitude dès lors que l'on suppose que la *vraie* structure nous est inaccessible³⁴.

5.3.5.3 L'approche déterministe

L'approche déterministe est adaptée à une situation de grande incertitude du décideur (voir la figure 5.3.7). Celui-ci procède par identification de la variable-cible, de ses parents, puis de ses ascendants d'ordre supérieur. Dans le cas de la sûreté nucléaire, il s'agirait de l'état du cœur, des trois fonctions fondamentales de sûreté (contrôle de la réaction, refroidissement, confinement), puis de fonctions plus élémentaires, du fonctionnement de systèmes spécifiques, etc. Par ses interventions, le décideur cherche à mettre en place des mécanismes permettant d'éviter l'événement redouté. Il sait cependant que le réseau sur lequel il s'appuie est très partiel, et cherche à assurer la robustesse de l'installation contre la survenue d'événements externes potentiellement importants. Pour cela, il

³⁴. Autrement dit, nous ne pouvons pas formuler une partition de l'ensemble des modèles possibles (Winkler, 1994).

FIGURE 5.3.7 – Démarche déterministe de sûreté



intègre aux mécanismes élaborés des marges de sûreté et en organise la succession le long de toutes les chaînes causales menant à la variable cible, en faisant à chaque fois l'hypothèse pénalisante d'un dépassement des marges pour le mécanisme situé en amont. Ces mécanismes successifs forment les niveaux de la défense en profondeur, figurés par des cercles d'épaisseurs différentes sur le graphe.

L'approche déterministe est soumise à deux formes principales d'incertitude pouvant conduire au dépassement des marges pour une ou plusieurs variables du système. L'expression formelle de ces marges est :

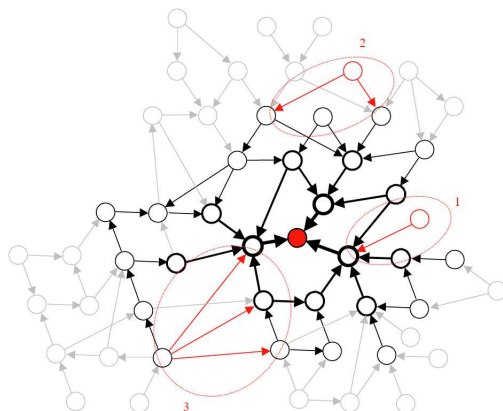
$$\mu_i^0 = g_i^0(\bar{x}_{Pa^0(i)}) - g_i^*(x_{Pa^*(i)}, y_i^*)$$

La première forme d'incertitude, symbolisée par la variable manquante (1) sur le graphe de la figure 5.3.8, est liée à la survenue d'un événement exogène dépassant directement les marges prévues. Les mécanismes situés en aval, s'ils disposent eux-mêmes de marges suffisantes, peuvent compenser ce dépassement mais, comme nous l'avons noté, il est impossible de savoir *a priori* si cette compensation aura lieu. Nous verrons au chapitre suivant que les marges intégrées à la conception des centrales japonaises leur ont permis par le passé de subir sans dommage majeur des séismes dépassant largement leurs niveaux de dimensionnement.

La seconde forme principale d'incertitude est le mode commun représenté par la variable manquante (2) du graphe. L'existence d'une cause commune latente peut entraîner la survenue simultanée d'événements supposés indépendants sur des chaînes causales distinctes, et finir par provoquer l'événement redouté sans qu'il y ait eu dépassement des marges de sûreté en amont.

Une forme particulière de mode commun est celui qui affecte différents niveaux de la défense en profondeur (arcs manquants (3) sur le graphe) et suffit de ce fait à causer l'événement redouté. Le tsunami qui a affecté la centrale de Fukushima Dai-Ichi en mars 2011 était un tel événement (voir chapitre suivant).

FIGURE 5.3.8 – Incertitudes de la démarche déterministe



5.4 Conclusion

Le modèle proposé nous a permis de rendre compte de l'ensemble des outils des démarches déterministe et probabiliste de la sûreté, de poser ainsi un cadre dans lequel ces approches peuvent être comparées, et finalement de reprendre la question de la définition du risque, que nous avons laissée en suspens au premier chapitre.

La définition 5.17 généralise et clarifie toutes celles que nous avons passées en revue dans les chapitres précédents, notamment celle de Kaplan et Garrick (1981), où le risque est un ensemble de triplets (scénario, probabilité, conséquences). La force de la définition de Kaplan et Garrick, qui sert comme nous l'avons vu de référence dans le domaine de la sûreté, est qu'elle explicite en partie la construction causale sur laquelle repose la mesure du risque. Les scénarios, en effet, y sont définis comme les états d'un système qui décrivent la réalisation d'événements tels que des accidents. La faiblesse de la définition, cependant, est que cette explicitation n'est que partielle, puisqu'elle ne s'étend ni aux événements en amont du système, que nous avons appelés ici les états du monde et auxquels se rapportent les notions d'aléa ou de danger, ni aux événements en aval que sont les conséquences. La séparation entre conséquences et scénarios semble arbitraire dès lors que le calcul des premières relève d'un prolongement de la représentation causale du système, ce prolongement fut-il déterministe.

Notre définition repose sur une seule distinction : celle entre les aspects du monde que le décideur choisit de représenter et ceux qu'il assimile à un aléa. Elle découle d'une représentation générale de ses interventions sur les premiers aspects, et intègre de ce fait l'effet causal de son action. Elle nous invite, dès lors, à envisager une extension de notre cadre vers les modalités du choix de ces actions, c'est-à-dire vers la théorie de la décision.

Chapitre 6

Les leçons de Fukushima

*"You acted unwisely," I cried, "as you see
by the outcome." He calmly eyed me :
"When choosing the course of my action," said he,
"I had not the outcome to guide me."
Ambrose Bierce, A Lacking Factor*

6.1 Introduction

Dans ce chapitre, nous cherchons à analyser l'accident nucléaire de Fukushima et les leçons qui peuvent en être tirées à l'aide du modèle causal exposé au chapitre dernier. Nous n'essaierons pas de construire une représentation exhaustive des installations et des événements qui y sont survenus, mais plutôt d'éclairer, par quelques applications simples, un problème qui nous semble être au cœur de l'accident : les formes de prise en compte d'événements exogènes incertains. Ceci nous fournira l'occasion d'examiner de plus près les frontières de notre modèle causal et les conditions qui doivent y régner pour que le modèle soit un outil d'inférence fiable, et de nous interroger sur les modalités de passage d'un mode de raisonnement déterministe à un mode probabiliste (ou inversement). Enfin, le retour au thème de la sûreté nucléaire, nous permet de préciser et d'illustrer l'analyse critique que nous avons proposée à la fin des chapitres 3 et 4.

La première partie du chapitre retrace la chronique des événements survenus entre le 11 et le 25 mars dans les réacteurs de la centrale de Fukushima Dai-Ichi, à un niveau de détail suffisant pour distinguer quelques faits caractéristiques de l'accident. Dans la deuxième partie, nous proposons une analyse formelle de ces faits, suivie d'une étude descriptive de l'évaluation et de la gestion des risques sismiques au Japon, ce qui nous permet de conclure la partie en tirant certains enseignements de l'accident. Dans la troisième partie, enfin, nous examinons dans quelle mesure les enseignements de

FIGURE 6.2.1 – Emplacement des centrales de la région de Tohoku



Fukushima sont pertinents dans le cas de la France, notamment en nous positionnant par rapport au processus de retour d'expérience qui a commencé à être mis en place avec les Évaluations complémentaires de sûreté.

6.2 Chronique des événements

Le 11 mars 2011 à 14h46, un tremblement de terre de magnitude Mw 9, le quatrième séisme le plus puissant jamais enregistré sur la planète, se produisit à 25 *km* de profondeur au large de la côte est du Japon, à 130 *km* de la ville de Sendai. Le séisme causa des dommages importants dans une vaste part de Honshu, la plus grande île de l'archipel japonais. Il provoqua surtout des tsunamis d'une ampleur exceptionnelle qui dévastèrent la côte nord-est de l'île. Fin août 2011, les pertes humaines de cet événement étaient estimées à 15 726 morts, dont près de 90 % semblaient avoir péri de noyade, et 4 593 disparus (Croix rouge japonaise, 2011), le nombre d'évacués à 84 000, le nombre d'immeubles détruits ou endommagés à 810 000 (Gouvernement du Japon, 2011b) et la valeur totale des seules destructions physiques à *EUR* 160 *mrds*, soit 3 % du produit intérieur brut japonais (Gouvernement du Japon, 2011c). Le tremblement de terre de Tohoku est la catastrophe naturelle la plus meurtrière subie au Japon depuis près d'un siècle¹ et la plus coûteuse au monde.

Le séisme et les tsunamis entraînèrent la défaillance d'une grande partie du réseau de distribution d'électricité dans le nord de Honshu. D'importance mineure à première vue à l'échelle des conséquences de la catastrophe naturelle, l'interruption de la fourniture d'électricité eut pourtant

1. Précisément depuis le grand tremblement de terre de Kanto en 1923.

des répercussions importantes sur le fonctionnement de nombreuses infrastructures, notamment les quatre centrales nucléaires de la côte nord-est de l'île de Honshu (voir la carte de la figure 6.2.1). Dans les trois centrales de Onagawa, Fukushima Daini, et Tokai Daini, cette succession d'événements causa des dommages importants, mais pas d'accident impliquant la fusion du cœur d'un réacteur. Le site de Fukushima Daini, en particulier, fut très affecté par le tsunami mais conserva une ligne d'alimentation électrique externe, ce qui permit d'éviter un accident grave dans l'un de ses quatre réacteurs d'une puissance de 1 100 MWe .

Nous nous concentrons, dans la suite, sur la centrale de Fukushima Dai-ichi qui subit les conséquences les plus désastreuses.

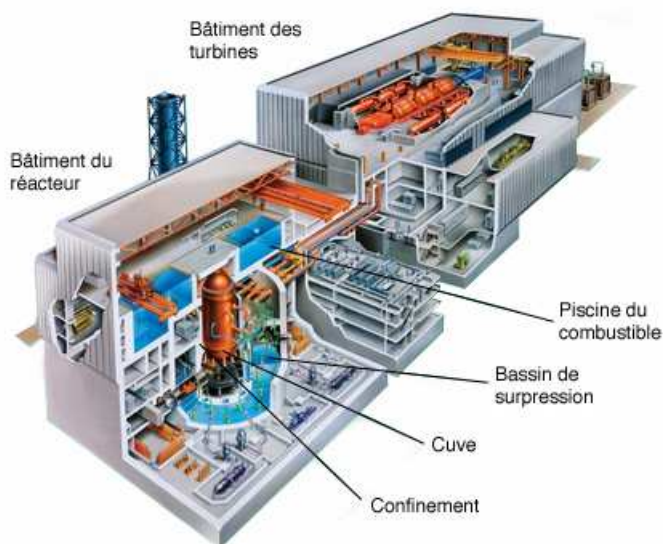
6.2.1 Données générales

Le site de Fukushima Dai-ichi était exploité par l'entreprise Tokyo Electric Power Company (TEPCO), par ailleurs propriétaire des centrales nucléaires de Fukushima Daini et de Kashiwazaki-Kariwa. Le site comprenait six réacteurs, dont les quatre premiers étaient disposés côte à côte sur une plate-forme située dix mètres au-dessus du niveau de la mer et les deux derniers un peu plus au nord et en léger surplomb. Le réacteur 1, doté d'une puissance de 460 MWe , avait été mis en fonctionnement en 1971 ; il s'agissait de l'une des premières tranches construites au Japon. Les réacteurs 2 à 5, entrés en fonctionnement en 1974, appartenaient à la même génération, mais étaient d'une puissance plus élevée (784 MWe). Le réacteur 6, enfin, était de conception plus récente et sa puissance était de 1 100 MWe . Les six unités étaient des réacteurs à eau bouillante (REB), dont la différence fondamentale par rapport aux réacteurs à eau pressurisée (REP) présentés au chapitre 3 est l'absence de fluide secondaire (les turbines génératrices d'électricité y sont directement actionnées par la vaporisation du fluide primaire). La figure 6.2.2 présente un schéma synthétique d'un REB du type de ceux de Fukushima Dai-ichi.

L'accident de Fukushima Dai-ichi a fait l'objet de plusieurs rapports officiels, notamment celui du gouvernement japonais à l'AIEA (Gouvernement du Japon, 2011) et ceux des missions d'enquête ou d'observation dépêchées par l'AIEA (AIEA, 2011), la NRC américaine (NRC, 2011) et l'INPO² (INPO, 2011). Nous nous appuyons dans la suite de cette section sur ces rapports, dont les éléments factuels sont largement convergents, pour décrire de façon synthétique le déroulement des événements depuis l'arrivée des premières ondes sismiques jusqu'à la stabilisation des réacteurs après l'accident. Nous nous intéresserons essentiellement aux causes de l'accident, et non à ses conséquences, en analysant en premier lieu son déclenchement, puis en établissant une chronologie détaillée des faits survenus dans le réacteur 1, et enfin en ajoutant quelques éléments concernant

2. L'*Institute of Nuclear Power Operations* est une organisation fondée par l'industrie nucléaire civile américaine après l'accident de Three Mile Island pour développer des normes de sûreté et promouvoir l'échange d'informations (<http://www.inpo.info>).

FIGURE 6.2.2 – Schéma d'un réacteur à eau bouillante avec un confinement de type Mark I



Source : General Electric.

les autres unités.

Il convient de noter que les données techniques reprises dans les rapports mentionnés proviennent dans une très large mesure de l'entreprise exploitante TEPCO. Ces informations, qu'elles soient issues des mesures effectuées sur place ou de simulations informatiques, demeurent à ce jour partielles et entachées d'incertitudes. À titre d'exemple, il est impossible de déterminer avec précision à quel moment et dans quelle mesure les cœurs ont été dénoyés. Les indications de niveaux d'eau rapportées ci-dessous sont issues des simulations de TEPCO, que l'entreprise elle-même présente comme conservatrices. Les niveaux constatés par les opérateurs sur les appareils de mesure ont été généralement plus élevés, mais ils ont été faussés par l'ébullition de l'eau dans les cuves des réacteurs 1 à 3.

De surcroît, il ne peut être exclu que TEPCO fasse un usage stratégique de l'asymétrie d'information dont elle bénéficie. L'entreprise a en particulier cherché à privilégier de façon constante la thèse selon laquelle les installations de Fukushima Dai-ichi n'avaient subi aucun dommage sérieux suite au séisme et que toutes leurs défaillances avaient été causées par le tsunami³. Le récit qui suit reproduit donc cette thèse. Mais il convient de noter qu'au moment où ces lignes étaient écrites, en décembre 2011, celle-ci laissait un certain nombre d'éléments inexplicables⁴ et ne pouvait pas être

3. Voir, à titre d'exemple, le rapport de TEPCO à l'autorité de sûreté japonaise NISA (*Nuclear and Industrial Safety Agency*) au sujet de la tenue de la tranche 1 au séisme du 11 mars (TEPCO, 2011).

4. À titre d'exemple, en essayant de réconcilier les concentrations de radionucléides relevées en différents points de la planète avec le terme source des réacteurs de Fukushima Dai-ichi, Stohl et al. (2011) parviennent à la conclusion que les émissions de gaz noble Xénon 133 ont dû commencer immédiatement après le séisme, et non le 12 mars comme les faits rapportés par TEPCO le laissent à penser. De telles émissions peuvent certes être provenues d'une

considérée comme établie.

6.2.2 Les initiateurs

Des six réacteurs de la centrale de Fukushima Dai-ichi, trois étaient à l'arrêt le 11 mars 2011 pour des activités normales de maintenance et de renouvellement du combustible. Dans l'unité 4, le combustible avait été entièrement retiré du cœur et déposé dans la piscine de refroidissement du combustible usé, de façon à permettre la conduite de travaux de maintenance de la cuve. Dans les unités 5 et 6, qui étaient en attente de redémarrage, le combustible avait été rechargé dans le cœur.

Selon les informations transmises par l'exploitant TEPCO⁵, dans les trois autres unités, la chute automatique des grappes de contrôle aurait interrompu la réaction en chaîne dès la détection des premières secousses sismiques⁶ (c'est-à-dire avant la défaillance des sources électriques externes). Dans l'ensemble des tranches, les circuits normaux de refroidissement du cœur se seraient fermés et les enceintes de confinement se seraient placées en isolement (*containment isolation*). Tous les générateurs au diesel disponibles se seraient mis en marche, rétablissant l'alimentation des systèmes importants de chaque unité, notamment ceux destinés au refroidissement d'urgence du cœur. Dès la fin des secousses, les opérateurs semblent avoir engagé leurs plans d'urgence, avec l'activation du Centre de contre-mesures pour catastrophe au siège de l'entreprise TEPCO à Tokyo et celle du Centre de réponse d'urgence sur le site. L'ensemble des procédures destinées à prévenir un accident majeur en cas de perte des sources extérieures d'alimentation électrique semble donc avoir bien fonctionné.

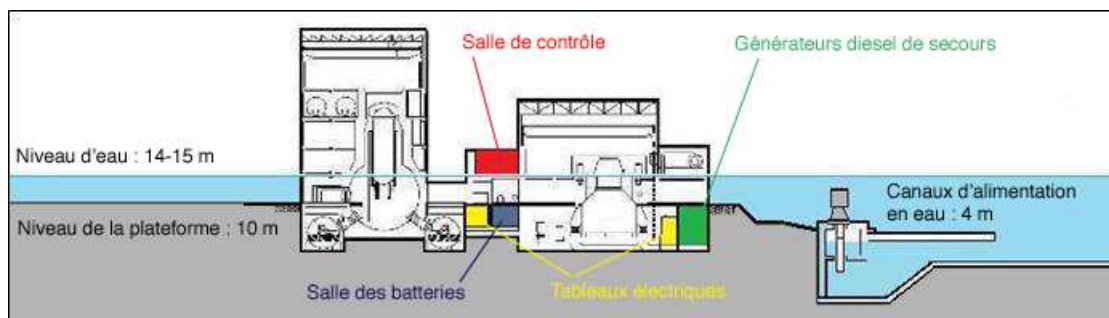
À 15h27, quarante et une minutes après le séisme, une vague d'une hauteur de quatre mètres vint se briser contre la digue protégeant le site de Fukushima Dai-ichi. Cette première vague fut suivie à 15h35 d'une seconde, qui submergea cette fois la digue de 6,5 mètres ainsi que le marégraphe, dont la hauteur d'observation maximale était de 7,5 mètres. Cinq autres vagues suivirent, portant l'eau à une hauteur totale de plus de 14 mètres et inondant les plates-formes sur lesquelles les installations centrales ont été construites (voir figure 6.2.3). En contrebas, l'eau s'engouffra dans les canaux d'alimentation, endommageant les pompes, causant des courts-circuits et entraînant une perte immédiate de la source froide pour les six tranches. Dans les bâtiments principaux, elle inonda les parties inférieures, notamment les salles abritant les tableaux électriques et la plupart des groupes électrogènes. Deux membres des équipes de conduite périrent noyés dans le sous-sol du bâtiment des turbines de l'unité 4.

Les unités 1 à 5 possédaient deux groupes électrogènes au diesel chacune, contre trois pour autre source que Fukushima Dai-ichi, mais cette hypothèse semble peu vraisemblable.

5. Voir par exemple la synthèse récente de cette information par l'*Institute of Nuclear Power Operations* créée par les exploitants américains de centrales nucléaires après l'accident de Three Mile Island (INPO, 2011).

6. Comme nous l'expliquons plus bas, les centrales japonaises sont équipées d'un dispositif d'arrêt d'urgence en cas de détection de mouvements sismiques supérieurs à une certaine valeur.

FIGURE 6.2.3 – Situation des unités 1, 2, 3 et 4 de la centrale de Fukushima Dai-ichi



Source : INPO (2011)

l'unité 6, la plus récente. Dix de ces groupes, situés en sous-sol et refroidis à l'eau de mer, étaient hors d'usage. Une amélioration de sûreté avait conduit à placer les trois groupes restants, attachés aux unités 2, 4 et 6, dans des bâtiments séparés, à les élever au-dessus du niveau du sol et à préférer, dans leur cas, un mode de refroidissement par air. Le groupe de l'unité 4 était en arrêt planifié pour maintenance. Celui de l'unité 2 resta opérationnel, mais le tableau auquel il était relié se situant pour sa part en sous-sol, il se trouva hors d'usage. Au total, seul un groupe du réacteur 6 put finalement fonctionner.

À 15h41, cinq des six réacteurs de Fukushima Dai-ichi étaient totalement privés de courant alternatif et leurs systèmes de refroidissement ne fonctionnaient plus. Pour la première fois dans l'histoire de l'industrie nucléaire, les opérateurs de la centrale faisaient face simultanément à plusieurs situations d'accident grave.

L'évacuation de la puissance résiduelle des cinq réacteurs affectés ne reposait plus que sur des systèmes de refroidissement d'urgence faisant appel à des batteries et, selon les unités, des systèmes passifs de condensation ou des turbines. D'après les Plans d'urgence interne, ces sources devaient permettre aux équipes de conduite de contenir la progression de l'accident pendant quelques heures, en espérant un rétablissement de l'alimentation extérieure. Dans l'unité 5, les équipes réussirent à maintenir un refroidissement suffisant de la matière fissile située dans le cœur et dans la piscine du combustible et, le 15 mars, au cinquième jour de la catastrophe, à établir une ligne avec le générateur diesel opérationnel de l'unité 6. Cela permit de parvenir à l'arrêt à froid des deux réacteurs cinq jours plus tard, le 20 mars.

Dans les tranches 1 à 3, la dynamique accidentelle ne put être maîtrisée et dans la tranche 4, elle ne le fut qu'à la faveur des circonstances.

6.2.3 La progression des accidents

6.2.3.1 Réacteur 1

Dans les unités 1 et 2, les batteries furent inondées, rendant inopérants les systèmes d'instrumentation et de contrôle. Dès 15h37, la salle de contrôle commune aux deux unités fut plongée dans l'obscurité. À partir de 15h50, elle ne reçut plus d'indication relative au fonctionnement de de refroidissement d'urgence du réacteur 1. À 16h36, les indicateurs de niveau d'eau et de pression à l'intérieur de la cuve cessèrent également de fonctionner. Par la suite, les opérateurs n'eurent plus eu accès qu'à des bribes d'information concernant ces paramètres essentiels, soit à la faveur de retours ponctuels de l'instrumentation, soit, le plus souvent, en se rendant dans le bâtiment du réacteur pour y effectuer des relevés. Pour cela, les équipes devaient faire usage de batteries d'automobile et, par manque de moyens de communication, revenir jusqu'à la salle de contrôle pour y faire leur rapport.

Le système de refroidissement d'urgence de l'unité 1 comprenait une injection de sécurité (*high pressure core injection*) et deux condensateurs d'isolement (*isolation condenser*). La première, qui seule aurait permis d'ajouter de l'eau dans le circuit de refroidissement, ne pouvait fonctionner sans électricité. Mais ce n'était pas le cas des condensateurs, qui devaient en principe faire circuler le fluide à travers un réservoir placé en surplomb de la cuve par simple usage de la gravitation, et permettre ainsi d'évacuer une partie de la chaleur du cœur sans apport d'eau pendant huit heures.

Entre 14h52 et 21h30, l'un des deux condensateurs fut actionné par intermittence et à chaque fois arrêté, dans certains cas, conformément aux procédures, pour éviter une baisse trop brutale de la température du cœur, et dans d'autres sans raison apparente. Selon les conclusions d'une enquête conduite en septembre, le condensateur ne semble pas avoir eu le rendement attendu même avant l'arrêt de son fonctionnement à 21h30.

Il est probable que dès la soirée du 11 mars, après plusieurs heures sans injection d'eau, le cœur du réacteur 1 soit entré en fusion. Selon les simulations réalisées par TEPCO, le niveau d'eau dans la cuve aurait atteint le sommet des barres de combustible (*top of active fuel*) environ trois heures après l'arrêt d'urgence, et le cœur aurait été totalement dénoyé après deux heures supplémentaires. On estime ainsi que le combustible dans la partie centrale du cœur aurait commencé à fondre quatre heures après l'arrêt, qu'il aurait été largement endommagé un peu plus d'une heure plus tard et que quinze heures après l'arrêt, lorsqu'eut lieu la première injection d'eau, le corium avait en partie coulé en fond de cuve. On considère que celle-ci a probablement été endommagée, mais l'instrumentation disponible dans la centrale ne permet pas, à l'heure actuelle, d'en connaître l'état précis.

À 20h50, la préfecture de Fukushima prononça un ordre d'évacuation de la population résidant à moins de deux kilomètres de la centrale.

À 21h23, alors que les nouvelles en provenance de l'unité 2 faisaient état d'un dénoyage imminent du cœur du réacteur, le Premier Ministre étendit la zone d'évacuation à un rayon de trois kilomètres et demanda aux personnes vivant à moins de dix kilomètres de la centrale de se confiner chez elles.

Au cours de la soirée, les niveaux de radioactivité sur le site augmentèrent fortement pour des raisons qui restent inexplicables. À 21h51, au vu des doses enregistrées, il fut décidé de restreindre l'accès des opérateurs au bâtiment du réacteur.

Dans le même temps, la pression relevée dans la cuve chuta et celle à l'intérieur de l'enceinte de confinement augmenta. Il y avait donc eu une perte d'étanchéité de la cuve, dont l'origine reste inconnue à ce jour, et fuite d'un mélange gazeux vers l'enceinte. Outre des gaz radioactifs, ce dernier comprenait vraisemblablement de grandes quantités d'hydrogène produites par l'oxydation du gainage en zirconium des barres de combustible qui n'étaient plus noyées.

La gestion de la pression dans l'enceinte constitue l'un des points délicats de la conduite accidentelle, puisqu'il s'agit d'arbitrer entre des relâchements volontaires dans l'atmosphère et le risque d'une perte de confinement qui conduirait à des rejets incontrôlés. À Fukushima, ce dilemme se doublait de la nécessité d'apporter de l'eau au cœur, puisque le seul moyen d'injection disponible, une pompe à incendie que les opérateurs avaient reliée au système d'aspersion, nécessitait une baisse supplémentaire de la pression dans la cuve, et donc dans l'enceinte.

Le plan d'urgence spécifiait que dans un cas de fusion du cœur, les vannes de dépressurisation devaient être activées dès lors que la pression dans l'enceinte avait atteint le double de sa valeur de dimensionnement. Les autorités nationales et locales devaient être informées de la décision avant sa mise en œuvre, notamment afin d'organiser l'évacuation des populations avoisinantes, mais leur autorisation n'était pas requise. Dès minuit, le responsable du site demanda aux équipes de conduite de préparer l'ouverture des vannes. Le 12 mars à 1h30, il en informa le gouvernement, qui décida de la tenue d'une conférence de presse à ce sujet à 3h00. Le Centre de contre-mesures de TEPCO ordonna alors au responsable du site d'attendre la fin de la conférence pour procéder à la dépressurisation. Vers 02h30, les équipes constatèrent que la pression était à peu près identique dans la cuve et dans l'enceinte de confinement, et qu'elle avait atteint le seuil prescrit.

Vers 4h00, alors que l'ordre d'ouverture des vannes n'avait toujours pas été donné, la pression relevée dans l'enceinte baissa brusquement, tout en restant largement supérieure au niveau de dimensionnement. Cette baisse pouvait être le signe d'une perte de confinement sous l'effet d'une surpression prolongée.

À 4h51, de premières traces de radioactivité de $1 \mu Sv \cdot h^{-1}$ furent relevées en bordure du site. À l'intérieur, la radioactivité progressait de nouveau. À 5h00, toutes les personnes présentes sur le site reçurent l'instruction de revêtir des combinaisons et des masques filtrants.

À 5h44, le Premier Ministre donna l'ordre d'étendre la zone d'évacuation à un rayon de 10 km

autour de la centrale.

À 05h46, bien que la dépressurisation n'ait toujours pas été engagée, la pression dans la cuve avait suffisamment baissé pour que les opérateurs puissent commencer à alimenter le cœur en eau fraîche, dans un premier temps à un faible débit. Durant neuf heures, ils y déversèrent l'ensemble des réserves d'eau dont ils disposaient, soit 80 tonnes.

À 6h50, le Ministre de l'économie ordonna à TEPCO de dépressuriser l'enceinte de la tranche 1. Mais les responsables de l'entreprise, apprenant que certains habitants de la zone d'évacuation ne savaient pas où se diriger, décidèrent de temporiser encore. Après discussion avec le Premier Ministre qui s'était déplacé sur le site, TEPCO annonça que les vannes seraient ouvertes à 9h00.

À 9h03, les responsables du site furent informés que l'évacuation était achevée, et donnèrent l'ordre de dépressurisation. Cependant, les vannes ne pouvaient pas être actionnées à distance en raison de la panne des batteries et de l'absence d'air comprimé, et aucune disposition n'était prévue pour leur ouverture manuelle. Durant les heures suivantes, les équipes de conduite tentèrent de les ouvrir par différents moyens de fortune, mais leurs efforts se trouvaient contrariées par les niveaux de radiation très élevés dans le bâtiment du réacteur.

Ce n'est finalement que vers 14h30 que les opérateurs eurent confirmation qu'ils avaient réussi et que la pression avait sensiblement déchu dans l'enceinte de confinement.

À 15h36, une explosion eut lieu dans le bâtiment du réacteur 1, vraisemblablement causée par l'accumulation d'hydrogène qui s'était échappé depuis une dizaine d'heures de l'enceinte de confinement. Des quantités importantes de gaz radioactifs furent relâchés dans l'atmosphère et des débris contaminés projetés dans le site, causant de nouveaux dommages. La radioactivité mesurée à l'extérieur du site passa à légèrement plus de $1\text{ mSv} \cdot \text{h}^{-1}$.

À 18h25, le Premier Ministre annonça un nouvel élargissement de la zone d'évacuation, cette fois à un rayon de 20 km.

Environ trois heures et demie après l'explosion, les opérateurs réussirent à établir un moyen d'injecter de l'eau de mer, régulièrement enrichie en bore, dans le cœur. Le 20 mars, l'alimentation électrique de l'unité 1 fut rétablie. À partir du 25 mars, le système de refroidissement fut remplacé par un apport d'eau douce provenant d'un réservoir, lui-même alimenté par conduite depuis un barrage situé 10 km en amont. On injecta également du nitrogène dans l'enceinte de confinement afin d'y prévenir une nouvelle explosion d'hydrogène.

6.2.3.2 Réacteur 2

L'unité 2 était équipée d'un dispositif de refroidissement d'urgence différent de la première, combinant une injection de sécurité et un système de refroidissement du cœur en isolement (*reactor core isolation cooling*). Dans ce dernier, une turbine est alimentée par la vapeur d'eau provenant

de la cuve ; l'électricité qu'elle engendre est utilisée, selon le principe de l'îlotage, pour pomper l'eau du bassin de surpression situé sous la cuve et la réinjecter dans celle-ci. Le système fut mis en marche manuellement après le tremblement de terre, s'arrêta parce que le niveau d'eau dans la cuve était suffisant, et fut relancé après le passage du tsunami, à 15h39. N'ayant aucune indication à son sujet, les opérateurs craignirent un dénoyage du cœur pendant toute la soirée, jusqu'à ce qu'ils parviennent, après plusieurs tentatives, à s'assurer qu'il fonctionnait bien vers 02h00 le 12 mars.

Avec l'aide d'autres entreprises exploitantes et le soutien des services gouvernementaux, le Centre de contre-mesures de TEPCO était parvenu à faire acheminer des groupes électrogènes mobiles sur le site tard dans la soirée du 11 mars. Les systèmes électriques généraux étant trop endommagés, les opérateurs cherchèrent à relier ces groupes directement à certains systèmes vitaux. Il fallait pour cela localiser des équipements en état de marche, puis faire venir jusqu'à eux un câblage pesant plus d'un tonne en se frayant un chemin entre les débris et les parties inondées. Quarante opérateurs furent à pied d'œuvre pendant la nuit et la matinée du 12 mars. À 15h30, elles réussirent finalement à établir une connexion avec des pompes du système de refroidissement de l'unité 2. Six minutes plus tard, alors qu'elles étaient sur le point de mettre les pompes en marche, une explosion eut lieu dans bâtiment du réacteur 1. Elle endommagea le groupe électrogène et les câbles, blessa cinq opérateurs, créa une ouverture dans le bâtiment du réacteur 2 et entraîna une telle hausse de la contamination sur le site que tous travaux durent être interrompus.

Le 12 mars à 17h30, anticipant une dégradation des conditions similaire à celle qui venait de déboucher sur explosion dans la tranche 1, le responsable du site ordonna aux équipes de se préparer à ouvrir les vannes de dépressurisation de l'enceinte de confinement du réacteur 2. Le 13 mars à 10h15, la principale vanne fut ouverte, mais la pression dans l'enceinte était trop faible pour la maintenir en position. À 11h00, le responsable de site demanda la mise en place de moyens d'injection d'eau de mer dans le cœur.

Le 14 mars à 11h01, une explosion d'hydrogène eut lieu dans le bâtiment du réacteur 3, détériorant la pompe et les lances à incendie déployés dans l'unité 2, bloquant la vanne de dépressurisation en position fermée et entraînant un nouvel arrêt des travaux.

À 13h25, les opérateurs constatèrent que le système de refroidissement s'était arrêté. Des répliques sismiques survinrent alors, obligeant toutes les équipes à évacuer le site. Vers 16h00, lorsque l'ordre d'évacuation fut levé, la pression à l'intérieur de la cuve et de la chambre de surpression avait fortement augmenté. Les équipes durent alors collecter toutes les batteries d'automobile qu'elles trouvèrent pour essayer, par leur entremise, d'ouvrir une vanne de sécurité permettant de réduire la pression de la cuve. Lorsque, vers 18h00, ils réussirent, le cœur était en grande partie découvert et des quantités importantes de gaz radioactifs et d'hydrogène s'échappèrent. La pression de l'enceinte de confinement varia peu, ce qui indiquait que des fuites importantes vers le bâtiment et,

par l'ouverture de celui-ci, vers l'environnement extérieur.

Ce n'est qu'à 19h54 qu'avec la baisse de la pression dans le réacteur, l'injection d'eau de mer dans le cœur put commencer à l'aide d'une, puis de deux nouvelles pompes à incendie.

Entre le 14 et le 26 mars, on apporta ainsi plus de 9 000 tonnes d'eau de mer au cœur, en l'enrichissant en bore par intermittence pour écarter tout risque de surcriticalité. La vapeur accumulée dans la cuve continua de s'échapper vers le bassin de surpression (méthode appelée *feed and bleed*) puis, par l'intermédiaire des vannes laissées ouvertes, vers l'enceinte de confinement et l'extérieur. Le 20 mars, l'alimentation du réacteur en électricité fut rétablie. À partir du 26 mars, le réacteur fut approvisionné en eau douce selon le même système que l'unité 1.

Il semblerait que le niveau d'eau dans la cuve soit resté au-dessus du sommet des barres de combustible tant que le système de refroidissement du cœur en isolement fonctionnait. Il descendit ensuite brutalement avant de remonter jusqu'à mi-hauteur des barres avec l'injection d'eau de mer, et de se stabiliser à ce niveau. La température dans la cuve augmenta rapidement et 87 heures après la mise à l'arrêt, le centre du cœur était en fusion. La fusion s'étendit jusqu'à 120 heures après l'arrêt, avant que la température ne commençât à baisser. Il n'est pas possible, à l'heure actuelle, d'avoir une idée précise de l'état de la cuve. Les calculs indiquent qu'il doit être très dégradé, mais ils conduisent aussi à des estimations de températures nettement supérieures aux niveaux observés.

6.2.3.3 Réacteur 3

Contrairement aux unités 1 et 2, le réacteur 3 garda l'usage d'une partie de ses batteries, qui alimentèrent le système d'injection de sécurité, permirent d'ouvrir certaines vannes et fournirent un éclairage de secours au poste de contrôle commun aux unités 3 et 4. Les batteries s'épuisèrent toutefois au bout de 30 heures, alors que leur système de recharge avait été inondé.

Le système de refroidissement de secours de l'unité 3 était similaire à celle de l'unité 2. Après le tsunami, il fut actionné à 16h03 et fonctionna jusqu'au 12 mars à 11h36. Environ une heure après son arrêt, le faible niveau d'eau de la cuve déclencha automatiquement le système d'injection à haute pression (*high pressure coolant injection*), qui fonctionne également grâce à une turbine. Ce dernier s'interrompt le 13 mars à 02h44.

Sept heures plus tard, les opérateurs réussirent à alimenter la cuve en eau de mer et à laisser échapper sa vapeur selon le même procédé que pour le réacteur 2. Au total, près de 4 500 tonnes d'eau y furent injectées jusqu'à la mise en place d'une alimentation en eau douce le 25 mars. Pour évacuer la pression de l'enceinte de confinement, des vannes purent être ouvertes à 8h41 le 13 mars. À 11h01 le 14 mars, une explosion d'hydrogène se produisit dans le bâtiment du réacteur.

Le combustible dans le cœur du réacteur semble être resté noyé tant que l'un ou l'autre des deux

systèmes d'injection de secours fonctionnait. En particulier, la transition entre les deux systèmes semble avoir été suffisamment brève pour que le cœur ne subisse pas de dommage majeur. Par la suite, cependant, l'injection d'eau de mer aurait laissé les barres de combustible dénoyées sur une hauteur d'environ trois mètres, ce qui aurait conduit à la fusion d'une partie du cœur. L'ampleur de la fusion devrait être moins importante que dans le réacteur 1 et, selon les scénarios, une quantité de corium aurait ou non coulé au fond de la cuve. L'état de cette dernière reste inconnu.

6.2.3.4 Réacteur 4

Comme l'ensemble de la matière fissile du réacteur 4 se trouvait le 11 mars dans la piscine de refroidissement du combustible, sa situation ne parut pas, aux yeux des responsables de la gestion de crise, présenter le même degré d'urgence que celle des trois premiers. Ils ne s'attendaient donc pas à l'explosion qui se produisit le 15 mars vers 06h00 dans le bâtiment du réacteur, et la première explication qu'ils lui trouvèrent fut la fusion des barres de combustible placés dans la piscine. Différents événements pouvaient avoir conduit à une telle situation : une partie du liquide pouvait avoir quitté la piscine du fait des secousses sismiques, comme elle pouvait s'être évaporée ; la piscine pouvait aussi avoir perdu son étanchéité. Ils pouvaient d'ailleurs être en voie de se produire, ou même s'être déjà produits, dans les autres tranches. Les piscines de refroidissement n'étant pas équipées d'instruments de mesure du niveau de liquide, ces hypothèses étaient impossibles à vérifier.

En raison du volume très important de matière déposée dans la piscine du réacteur 4⁷, les équipes d'urgence consacrèrent des ressources considérables dans les jours suivants à essayer d'y ajouter de l'eau. Le 15 mars, on utilisa des canons à eau et des lances à incendie, mais leur portée était trop limitée. Le 17, des hélicoptères survolèrent le site pour tenter d'arroser la piscine, mais les niveaux de radiation les forcèrent à se tenir à une altitude trop élevée pour qu'ils puissent être efficaces. Le 18, finalement, on parvint, grâce à des pompes à béton, à injecter de l'eau dans la piscine de l'unité 4, ainsi d'ailleurs que dans celles des trois premières unités.

Par la suite, des investigations menée par TEPCO montrèrent que le combustible n'avait été dénoyé dans aucune piscine, et que ces efforts étaient donc superflus. La piste aujourd'hui privilégiée pour expliquer l'explosion de l'unité 4 est un reflux de l'hydrogène évacué de l'enceinte de confinement du réacteur 3.

6.3 Les causes d'une catastrophe

La description qui précède suffit à montrer que le séisme et le tsunami qu'il a déclenché ont mis à bas tous les mécanismes de prévention et de maîtrise des accidents graves pour au moins trois des

7. 1535 crayons de combustible, quand il y en avait 400 dans le cœur du réacteur 1 et 548 dans ceux des réacteurs 2 et 3.

réacteurs de la centrale, menant pas à pas à l'arrêt du refroidissement, la fusion du cœur et enfin la perte de confinement avec rejets précoces et massifs. Trois aspects des séquences ont illustré cette défaillance généralisée :

- l'étendue des pertes causées par le séisme et l'inondation : un très grand nombre d'équipements et de systèmes importants pour le maintien du refroidissement du cœur et du confinement de la matière radioactive dans les unités 1 à 3 furent affectés ; la multiplication des situations d'urgence entraîna une dispersion des équipes et des moyens de gestion de crise.
- l'extrême difficulté de la situation pour les équipes en charge de la gestion de crise : sur le terrain des opérations, les conditions de travail furent aggravées par la dégradation des structures, l'accumulation de débris, l'absence d'éclairage, la limitation des moyens de communication et, par la suite, plusieurs explosions et des niveaux élevés de radiation. À titre d'exemple, le centre de contrôle d'urgence du site ne disposait, pour communiquer avec les équipes de conduite, que d'une ligne téléphonique par poste de contrôle, chaque poste étant commun à deux tranches.
- l'absence de procédure pour faire face à certaines contingences ; certains dispositifs de dernier recours, tels que l'ouverture des vannes de dépressurisation, n'avaient visiblement pas été conçus sous l'hypothèse de perte totale des alimentations électriques.

Contrairement aux cas de Three Mile Island et Tchernobyl, l'accident de Fukushima ne résulta pas d'une succession d'erreurs humaines ou de défaillances mineures, mais d'un événement extérieur unique. Une fois survenue l'agression extérieure, la dynamique accidentelle était implacable.

Nous cherchons dans cette partie à analyser de plus près cet aspect de l'accident et à en tirer quelques enseignements de portée générale. Nous commençons par représenter le problème l'aide des outils formels exposés au chapitre 5, avant de nous intéresser à la façon dont le risque sismique avait été évalué et pris en compte dans la sûreté de la centrale.

6.3.1 Analyse formelle

6.3.1.1 Une représentation causale

Considérons le problème de protection d'une centrale contre une agression extérieure telle qu'un séisme pouvant se doubler d'un tsunami. Nous adoptons ici le point de vue des responsables de la sûreté d'une centrale comme Fukushima Dai-ichi. Les interventions envisageables sont de deux ordres : des normes sismiques imposées au dimensionnement de la centrale et une digue censée protéger l'ensemble du site contre les inondations⁸. Considérons un modèle peu raffiné, mais qui suffira pour nos besoins, des phénomènes en jeu. Supposons ainsi que l'ampleur du tsunami soit

8. Pour simplifier la représentation, la digue représente ici à la fois le mur anti-tsunami et la surélévation du site, de sorte que son débordement est synonyme d'inondation de la centrale.

déterminée par les seules caractéristiques du séisme ; que les tsunamis de hauteur inférieure à celle de la digue soient simplement effacés ; que les normes sismiques déterminent une valeur limite de résistance des structures et que les effets du séisme sur la sûreté de la centrale soient nuls tant que les secousses ne dépassent pas cette limite ; qu'il existe une mesure de la sûreté telle que son passage en-dessous d'un certain seuil serait équivalent à un accident grave ; que les effets de l'inondation et du séisme sur cette mesure soient cumulatifs ; et enfin que toutes les relations causales décrites soient déterministes⁹. Si l'on note X les caractéristiques du séisme, T la hauteur du tsunami affectant la centrale, S l'état des systèmes importants pour la sûreté de la centrale, A l'occurrence d'un accident grave, et $P = (N, D)$ l'intervention consistant à appliquer des normes de conception protégeant la centrale contre des secousses sismiques inférieures à une limite N et à l'équiper d'une digue de hauteur D :

$$\begin{aligned} T &= f(X) \\ S &= S_0 - g_N(X) - g_D(T) \\ A &= 1_{S < S_{min}} \end{aligned}$$

où S_0 est le niveau de sûreté avant le séisme, S_{min} le seuil de déclenchement de l'accident, f est une fonction et g_N et g_D sont des familles de fonctions croissantes avec $g_N(X) = 0$ si $X \leq N$ et $g_D(T) = 0$ si $T \leq D$.

Le caractère aléatoire des grandeurs du modèle provient de la seule variable X . Cette approximation, sur laquelle nous reviendrons sous peu, nous permet de mettre en exergue un aspect important du problème : X est ici la seule variable déterminée de façon exogène ; les autres variables sont toutes descendantes de X et les choix de protection de la centrale doivent tenir compte de ces dépendances, notamment de la relation entre X et T . Dans le cas qui nous intéresse, tout tremblement de terre dans la zone de subduction au large de Fukushima a de fortes chances de s'accompagner d'un tsunami. De ce fait, il est inapproprié de parler de deux aléas naturels. Le seul aléa est le séisme.

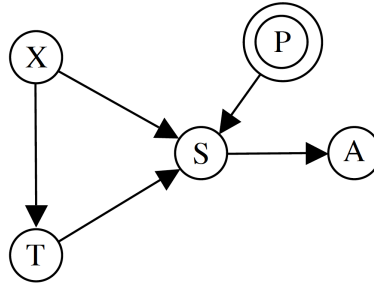
Si nous disposons de la distribution de probabilités de X , nous avons un réseau bayésien causal entièrement spécifié, dont le graphe est présenté dans la figure 6.3.1.

Supposons que notre modèle soit le reflet fidèle de la réalité et que nous en connaissions toutes les grandeurs. Alors le risque résiduel, c'est-à-dire la probabilité d'accident après intervention, peut être facilement dérivé de la distribution de X :

$$P_{N,D}(A = 1) = P(g_N(X) + g_D(f(X)) > S_{min} - S_0)$$

9. Au sens où l'on considérerait que les causes suffisent à déterminer totalement les effets.

FIGURE 6.3.1 – Une représentation du problème de sûreté



Dans ce cadre simplifié où l'on n'a considéré ni les coûts ni les limites techniques des protections, il suffit de choisir une combinaison d'interventions (N, D) permettant d'atteindre le risque résiduel que l'on se donne pour objectif ou, si l'on renverse le raisonnement, de protéger la centrale contre un événement d'une certaine fréquence.

Le problème devient plus délicat si nous relâchons l'hypothèse selon laquelle nous avons une connaissance parfaite de tous les phénomènes et grandeurs en jeu hormis les caractéristiques du séisme.

Considérons, à titre d'exemple, une situation où l'on est incertain de la valeur de différents paramètres de la fonction g_N , qui mesure la propagation des ondes sismiques jusqu'aux équipements de la centrale et la réponse de ceux-ci. Il faut alors tenir compte de ces sources supplémentaires d'incertitude lors de la détermination de l'intervention (N, D) , ce que l'on fait généralement en y intégrant des marges de sûreté.

Comme nous l'avons vu¹⁰, ces marges peuvent être fondées sur des hypothèses réalistes. Dans ce cas, la démarche équivaut à envisager, pour chaque valeur de X , l'ensemble des valeurs possibles de $g_N(X)$, y définir un sous-ensemble des valeurs admissibles (par exemple les quatre-vingt quinze premiers centiles), et enfin à estimer un majorant de ce sous-ensemble (par exemple le quatre-vingt quinzième centile). La distribution de probabilités associée à ces valeurs majorantes est une estimation de l'enveloppe supérieure des distributions de probabilités admissibles. Le risque résiduel, enfin, est une estimation du risque résiduel maximum. L'ensemble de ces estimations sont toutefois relatives à la représentation utilisée et nous ne pouvons pas être certains de leur validité.

Alternativement, les marges peuvent découler de l'usage d'hypothèses pénalisantes à chaque fois qu'une composante de g_N est incertaine. Dans ce cas, la valeur de chaque marge, et *a fortiori* celle de leur résultante, sont inconnues. Nous ne savons pas comment se situent nos estimations de l'effet du séisme $g_N(X)$ ou du risque résiduel $P_{N,D}(A = 1)$ par rapport à leurs ensembles respectifs de valeurs possibles. Naturellement, ces estimations restent soumises à l'incertitude de modélisation.

La confrontation de cette représentation théorique avec l'ensemble des observations existantes

10. Nous rappelons ici quelques développements de la section 3.5.1.

est essentielle pour éclairer l'étendue de nos incertitudes et décider de nos interventions en conséquence. Considérons, à titre d'exemple, la relation $T = f(X)$ qui doit nous permettre de déterminer une distribution théorique de probabilités sur les hauteurs de tsunami à partir de celle sur les caractéristiques du séisme. Supposons que nous disposions d'un ensemble d'observations historiques nous permettant de construire une distribution empirique de hauteurs de tsunami, et que les deux distributions ne concordent pas. Il est alors possible que la représentation théorique nous indique la possibilité de survenue d'événements qui n'ont pas été observés dans le passé, par exemple parce que les phénomènes qui en sont à l'origine ne se sont pas exprimés durant la période d'observation. Il est aussi possible, à l'inverse, que certains événements historiques, dont le modèle théorique ne permet pas de rendre compte, mettent en évidence les limites de notre représentation des phénomènes sous-jacents.

6.3.1.2 Une cause et une multitude de défaillances

Pour en venir à l'analyse de l'accident proprement dit, il peut être utile de rendre notre représentation un peu moins abstraite en considérant une catégorie particulière d'équipements. Dans le cas de Fukushima Dai-ichi, les sources d'alimentation électrique, dont la perte a joué un rôle central dans la dynamique accidentelle, sont un choix tout indiqué.

Comme nous l'avons vu, les réacteurs 1, 2 et 3 de la centrale de Fukushima étaient équipés de quatre sources principales d'électricité : l'alimentation extérieure $S1$, deux groupes électrogènes au diesel $S2$ et $S3$, et l'ensemble $S4$ de turbines et de batteries mobilisées lors de l'îlotage. Ces quatre sources étaient redondantes : chacune devait en principe permettre de maintenir les fonctions de sûreté du réacteur pendant un temps suffisant. Pour qu'il y ait un accident grave par manque d'alimentation électrique, il fallait donc qu'il y ait une défaillance cumulée des quatre sources (voir la figure 6.3.2). Ces équipements constituaient, sur la question particulière de l'alimentation électrique, les quatre premiers niveaux de la défense en profondeur.

Les séismes constituent une cause commune de défaillance majeure, puisque leur survenue peut affecter tous les équipements d'une centrale. C'est notamment le cas des sources externes et internes d'approvisionnement électrique. Comme nous le verrons dans la suite, les risques sismiques font l'objet de dispositions particulièrement contraignantes dans les centrales japonaises, qui doivent être construites sur un sous-sol rocheux, être en partie enterrées, être équipées de sismographes reliés aux systèmes d'arrêt d'urgence des réacteurs et respecter des normes sismiques parmi les plus restrictives qui soient.

À Fukushima Dai-ichi comme dans les autres sites nucléaires japonais, ces normes venaient d'être relevées en application d'une réglementation adoptée en 2006. Elles assuraient en principe le maintien des fonctions de sûreté pour des accélérations au sol maximales comprises entre 0,45 et 0,50 g

selon les tranches dans le sens horizontal, légèrement moins dans le sens vertical¹¹. L'événement ayant servi de référence pour leur calcul était un séisme de magnitude Mw 7,1 survenant au large de Fukushima. Outre ces dispositions générales, les équipements électriques tels que les groupes électrogènes, les batteries et les tableaux faisaient partie des systèmes prioritaires qui avaient été placés en sous-sol pour les protéger contre les effets d'amplification des mouvements sismiques causés par les structures.

Nous pouvons donc supposer que du point de vue du gestionnaire de la sûreté, l'impact du séisme lui-même sur les sources électriques avait été entièrement pris en compte dans le dimensionnement de la centrale ; que des secousses sismiques d'intensité inférieure ou égale à celles retenues dans ce dimensionnement ne devaient donc pas nuire au fonctionnement des sources internes mais tout au plus à celui de l'alimentation extérieure. L'impact de X sur $S2$, $S3$ et $S4$ relevait donc du risque résiduel après prise en compte de ces mesures, 1, ce que nous avons représenté par des flèches pointillées sur le graphe.

Le même raisonnement s'applique aux risques liés aux tsunamis.

Les dispositions prises en la matière à Fukushima étaient d'une part l'emplacement du site, à dix mètres au-dessus du niveau de la mer, et de l'autre une digue de près de six mètres protégeant les canaux d'approvisionnement en eau de la centrale, situés en contrebas. Lors de la demande d'autorisation de construction sur le site de Fukushima Dai-ichi, le tsunami choisi comme référence de dimensionnement avait été celui provoqué par un tremblement de terre de magnitude Mw 9,5 en 1960 au Chili. Dans la zone d'emplacement de la centrale, l'eau avait alors atteint une hauteur de 3,1 m. En 2002, l'exploitant TEPCO avait choisi de réévaluer l'aléa et avait alors retenu, sur la base d'un séisme de magnitude Mw 7,9 survenu en 1938 au large de Fukushima, une hauteur corrigée à 5,4 – 5,7 m. Il avait alors été décidé de surélever de vingt centimètres deux moteurs de pompe situés au niveau du canal d'approvisionnement de l'unité 6, et de placer trois groupes électrogènes refroidis par air à l'écart et en hauteur. Ces groupes appartenaient aux unités 2, 4 et 6, mais comme nous l'avons vu, les équipements électriques auxquels ils étaient reliés n'avaient pas fait l'objet de protections adéquates.

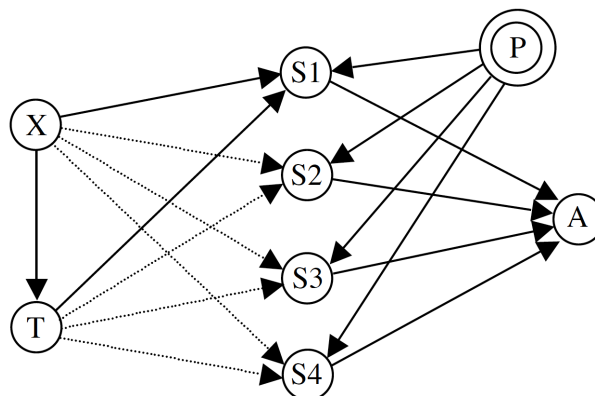
Après ces interventions, l'exposition des sources $S2$, $S3$ et $S4$ à l'aléa tsunami faisait également partie du risque résiduel.

L'accident du 11 mars a mis à jour deux failles fondamentales dans ce dispositif de sûreté.

La première est sa très grande vulnérabilité à des événements hors dimensionnement. L'événement constitue une cause commune qui a mis en échec l'ensemble des lignes de la défense en profondeur de cinq réacteurs.

11. L'accélération au sol est considérée comme la mesure la plus significative de l'intensité de secousses sismiques. Ses deux unités les plus courantes sont le *Gal* ($1 \text{ Gal} = 1 \text{ cm.s}^{-2}$) et l'accélération de la gravité terrestre g ($1 g = 980,665 \text{ Gal}$).

FIGURE 6.3.2 – Une représentation du risque d'accident lié à la perte totale des sources électriques



L'identification des aléas et l'estimation de leurs probabilités de survenue est en principe propre à chaque site. Il est important de noter que les conséquences d'une agression dépassant l'événement de référence (et dont la probabilité serait donc, en principe, inférieure au seuil retenu) ne sont pas explorées, ni pour l'ensemble du site ni même pour chaque réacteur isolément. En outre, les responsables de sûreté n'ont pas d'idée précise de l'ampleur des marges disponibles. Ceci rend l'ensemble de la démarche déterministe particulièrement sensible à l'estimation des probabilités d'agression et laisse ouverte la possibilité d'un *effet falaise**.

La seconde est l'insuffisance manifeste des valeurs de dimensionnement de la centrale de Fukushima Dai-ichi. Ceci est évident pour le tsunami, puisque les hauteurs d'eau prises comme référence ont été dépassées de près de neuf mètres. Mais les valeurs d'intensité sismique ont également été excédées dans les réacteurs 2, 3 et 5, la plus forte accélération ayant été mesurée dans la tranche 2, à 0,56 *g*. Avant l'événement, l'exploitant TEPCO avait estimé que la probabilité d'observer des mouvements au sols plus intenses que les limites retenues était comprise entre 10^{-6} et 10^{-4} (Gouvernement du Japon, 2011, p.III-29). Le séisme et le tsunami du 11 mars étaient donc tous deux considérés comme des événements extrêmement improbables avant leur survenue. Cette estimation reflétait-elle l'état des connaissances ? Nous cherchons dans les sections suivantes à rassembler quelques éléments de réponse à cette question.

6.3.2 Sismicité et sismologie au Japon

6.3.2.1 Une exposition extrême

Le Japon est situé dans l'une des régions de la planète les plus exposées aux séismes, à l'intersection de quatre plaques tectoniques dont deux sont océaniques (plaques pacifique et philippine) et deux continentales (plaques eurasienne et d'Okhotsk). Alors que le pays et ses mers avoisinantes recouvrent moins de 1 % de la surface du globe, ils ont été le siège de 13 % des tremblements de

terre d'une magnitude supérieure ou égale à Mw 7 survenus dans le monde entre 2002 et 2011¹².

Peu de parties du territoire sont abritées des risques sismiques, même si ceux-ci sont particulièrement importants sur l'ensemble de la côte pacifique, où les plaques océaniques plongent sous les plaques continentales. Dans cette zone de subduction, le glissement des plaques engendre des déformations qui se relâchent lorsqu'un segment situé sur l'une d'elles ou à leur intersection rompt. Les séismes qui en résultent causent de fortes secousses à la surface et s'étendent à de vastes zones. Ils s'accompagnent très souvent de tsunamis qui peuvent toucher des côtes à une grande distance de leur épicentre. Leur magnitude peut excéder Mw 8 lorsque la période d'accumulation de tension est de l'ordre du siècle. C'était le cas du grand tremblement de terre du Kanto, qui dévasta Tokyo en 1923 en faisant plus de 100 000 victimes et des millions de sans-abri et marqua une étape dans la construction du Japon moderne. Lorsque l'énergie libérée a été emmagasinée pendant des millénaires, comme pour le séisme de Tohoku du 11 mars 2011, la magnitude Mw 9 peut être dépassée (Le Pichon, 2011).

Les pressions exercées sur les plaques tectoniques peuvent également engendrer des séismes loin des zones de subduction, à des profondeurs inférieures à 20 km, le long de zones de faille. Ces séismes peuvent avoir des périodes de récurrence de l'ordre du millénaire, voire de la dizaine de milliers d'années. Ils sont habituellement de plus faible durée et de magnitude plus limitée que les précédents, mais ils peuvent causer des dommages considérables à l'échelle d'une ville en raison de leur proximité à la surface. Le grand tremblement de terre de Hanshin-Awaji qui affecta le port de Kobe le 17 janvier 1995, fit plus de 6 400 morts et endommagea 400 000 immeubles, était un séisme superficiel causé par la faille dite de Nojima.

6.3.2.2 Une politique sismologique

Le Japon a réalisé, depuis des décennies, un effort continu pour identifier et étudier les sources de sismicité sur son territoire. Un grand nombre de centres universitaires et d'institutions publiques et privées sont engagées dans la recherche en géologie, en sismologie ou en tectonique. Leurs travaux sont souvent considérés de niveau mondial et ils disposent d'un équipement technique de pointe. L'une des spécificités du Japon est l'intensité du lien entre cette production scientifique et la politique publique en matière réglementaire, d'aménagement du territoire ou d'offre de services publics. Un cadre a été mis en place dès le début des années 1960 pour coordonner l'ensemble des actions de l'État et des collectivités territoriales ayant un rapport avec la prévention des catastrophes, ce qui constitue au Japon un domaine extrêmement vaste¹³. Il définit une hiérarchie détaillée de textes législatifs et réglementaires et d'instances de décision et organise, à chaque niveau, les interactions

12. Nos calculs, à partir la base de données sur les séismes de l'United States Geological Survey, <http://earthquake.usgs.gov/earthquakes/eqarchives/epic/>.

13. *Disaster Countermeasures Basic Act* du 15 novembre 1961.

entre la politique publique et la recherche scientifique¹⁴.

Le désastre de Kobe, qui fut vécu au Japon comme une grave défaillance de la politique publique de prévention, conduisit en fait à une relance et un renforcement de cette approche. Quelques mois après le séisme, une loi exceptionnelle¹⁵ lança un vaste ensemble de réformes, au nombre desquels était la création d'un Centre pour la promotion de la recherche sismique. Le Centre devait œuvrer à « une politique nationale compréhensive de prévention des désastres sismiques », notamment « en communiquant et en appliquant les résultats de la recherche »¹⁶. En 1999, il produisit un document de doctrine intitulé « La promotion de la recherche sismique » dans lequel il affichait sa volonté d'élaborer des cartes d'aléas sismiques pour l'ensemble du territoire national et d'en faire le principal outil d'information et d'aide à la décision en matière de gestion des risques de tremblements de terre :

'Based on the results of nation-wide surveys of active faults, the systematization of information on deep-sea earthquakes, and data on past earthquakes, at present, using methods now under investigation by the earthquake survey committee, we conduct long-term probability assessments on the possible occurrence of regional shallow earthquakes and deep-sea earthquakes. In order for the public to feel the danger and imminence of an earthquake, it is desirable for them to be given information on the probability of occurrence, in increments of several decades, with the imminence index for as short a period of time as possible' (HERP, 1999, p.15).

Dans ce but, le document identifiait trois domaines de recherche prioritaires : la collecte et l'analyse de données géologiques, sismiques et tectoniques ; l'identification aussi étendue que possible des failles actives ; et la recherche sur la prédiction des tremblements de terre, c'est-à-dire l'évaluation fiable des probabilités d'occurrence de séismes, spécifiés par leur localisation et leur magnitude, sur des périodes suffisamment longues. Des avancées rapides dans ces trois directions devaient s'enchaîner pour conduire à une prévision toujours plus précise et détaillée des risques sismiques. Le Centre décrivait en outre la façon dont les ressources publiques allaient être orientées vers ces domaines.

Dans les années qui suivirent, le Japon développa son appareillage technologique déjà conséquent, pour en faire le réseau d'observation des mouvements de la croûte terrestre le plus dense au monde¹⁷. Les groupes de travail du Centre, qui regroupaient un grand nombre d'experts de renom,

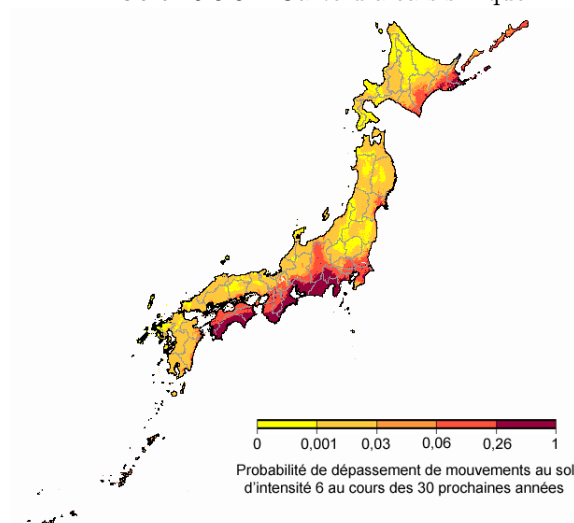
14. Pour une revue des politiques générales d'évaluation et de gestion des risques sismiques au Japon, on pourra se reporter à OCDE (2009).

15. *Special Measures Law on Earthquake Disaster Prevention* du 16 juin 1995.

16. Le Centre, dont le nom anglais est *Headquarters for Earthquake Research Promotion*, est rattaché au Ministère de l'éducation, de la culture, du sport, de la science et de la technologie (site : <http://www.jishin.go.jp/main/index-e.html>).

17. Au nombre des institutions engagées dans cet effort, l'Institut géographique national (*Geospatial Information Authority of Japan*) enregistre les mouvements des plaques tectoniques à partir d'un réseau de plusieurs milliers de stations géodésiques connectées par GPS, avec un degré d'imprécision estimé inférieur au centimètre (site

FIGURE 6.3.3 – Carte d'aléa sismique



Source : Headquarters for Earthquake Research Promotion

<http://www.jishin.go.jp/>

passèrent en revue quelques 2 000 failles actives identifiées dans le pays, et en sélectionnèrent un certain nombre qu'ils estimèrent particulièrement dangereux. Rassemblant ces failles avec les principaux séismes-type associés aux zones de subduction, ils constituèrent une liste de 98 scénarios principaux de tremblements de terre, dont ils évaluèrent les magnitudes et probabilités. En 2005, le Centre publia finalement le résultat de ces estimations sous forme de cartes de probabilité dont un exemple est présenté par la figure 6.3.3. Ces cartes d'aléas devinrent une pierre angulaire pour la politique gouvernementale de gestion des risques sismiques (OCDE, 2009).

6.3.2.3 Des orientations discutables mais guère discutées

L'orientation scientifique à laquelle les puissances publiques ont fortement contribué ne faisait pas l'unanimité dans la communauté des sismologues (et le fait encore beaucoup moins depuis la catastrophe de Fukushima). Deux points cristallisaient les divergences : la prévisibilité des séismes et le caractère représentatif du dernier siècle, sur lequel se concentre une grande partie des informations quantitatives disponibles.

Les modèles d'évaluation de probabilités de séisme se fondent sur une hypothèse dite de *seismic gap*, selon laquelle les mouvements des plaques tectoniques, lorsqu'ils s'accompagnent de déformations, provoquent des séismes de façon monotone. L'absence prolongée de séisme dans une zone de subduction est alors synonyme de l'augmentation de sa probabilité. Cependant, une partie impor-

<http://www.gsi.go.jp/ENGLISH/index.htm>). L'Institut national de recherche en sciences de la terre et en prévention des catastrophes (*National Research Institute for Earth Science and Disaster Prevention*) est, pour sa part, à la tête d'un réseau de plusieurs milliers de sismographes permettant de détecter et d'étudier les mouvements sismiques lents, de très faible ou au contraire de grande amplitude (<http://www.bosai.go.jp/e/>).

tante des sismologues s'est détournée de cette vision en vogue dans les années 1970, pour estimer que l'influence des mouvements tectoniques était plus complexe et pas nécessairement linéaire (Geller, 2011). Si l'approche reste fréquemment utilisée pour des séismes de faible profondeur et de magnitude élevée (supérieure à Mw 6 ou 6,5), c'est uniquement comme une solution empirique à la rareté des données.

Même si l'on accepte l'hypothèse de récurrence régulière des séismes majeurs, il reste à se donner une période d'inférence raisonnable. La sismologie officielle japonaise s'appuyait essentiellement sur travaux portant sur une période d'observation de l'ordre du siècle, pour en dériver un ensemble de séismes-type pour lesquels des prévisions étaient établies. Dans la zone de subduction appelée la fosse du Japon, en particulier, ces prévisions considéraient une succession de segments ayant chacun leur séisme caractéristique, dont la magnitude était directement liée à la taille du segment : Ibaraki, Fukushima, Miyagi, Sanriku, Tokachi. La possibilité d'une rupture en cascade de ces segments causant un séisme d'ampleur beaucoup plus grande était écartée. Or, depuis la fin des années 1970, tous les travaux ont indiqué seuls 20 à 25 % du mouvement de subduction de la plaque pacifique dans la région de la fosse du Japon avaient été compensés par les séismes enregistrés au cours du siècle dernier (Le Pichon, 2011). Ce déficit sismique pouvait avoir deux explications : soit le mouvement tectonique se faisait sans frottement, et était donc asismique, soit des séismes importants étaient ignorés dans l'histoire sismique japonaise. En 2001, des travaux universitaires ont établi que l'un des principaux séismes enregistrés dans cette région, celui de Jogan en 869, avait été accompagné d'un tsunami géant, dont les traces ont été retrouvées jusqu'à quatre kilomètres à l'intérieur des terres dans la baie de Sendai (Minoura et al., 2001). Trouvant des indications de la récurrence de tels événements environ un fois par millénaire depuis la Préhistoire, ils en ont déduit que la survenue d'un méga-séisme pouvait être imminente¹⁸.

Par ailleurs, les données géodésiques traditionnelles montraient qu'au nord du Japon, la croûte terrestre subissait une déformation importante depuis plusieurs siècles, ce qui contredisait l'hypothèse d'asismicité. C'est notamment ce fait troublant qui a poussé les autorités japonaises à mettre en place, à partir de 1994, un réseau dense de stations géodésiques satellitaires. Or les données fournies par ce réseau ont confirmé la déformation, et permis à plusieurs études de montrer, à partir de 1998, que la plaque eurasiennne était totalement bloquée sur la plaque pacifique le long du plan de subduction de la fosse du Japon (Le Pichon et al., 1998). Ceci confirmait la survenue à terme d'un très grand séisme.

Il existait, au total, un faisceau très important d'arguments théoriques et empiriques contre les estimations probabilistes de la sismologie officielle japonaise. Deux points nous semblent particulièrement remarquables dans cette situation de controverse scientifique.

18. Nott (2006) décrit comment les données paléontologiques pourraient être utilisées pour évaluer les probabilités de survenue d'une large gamme d'aléas naturels, notamment les séismes.

En premier lieu, les arguments opposés à la position officielle se sont rassemblés à partir de la deuxième moitié des années 1970, alors que la centrale de Fukushima était déjà construite. Au moment il eut lieu, le choix du site, dans un pays tel que le Japon, semblait très approprié du point de vue de la sismologie (Henry, 2011).

Le second aspect est que les instances publiques ne paraissent avoir fait aucune référence aux incertitudes existantes dans des documents officiels. Le document de doctrine du Centre pour la promotion de la recherche sismique ne fait pas mention de l'incertitude scientifique portant sur le phénomène sismique (HERP, 1999). Dans le rapport qui accompagnait la publication de ses cartes d'aléas et exposait la méthodologie suivie, le Centre discutait brièvement du caractère incertain de divers paramètres des modèles mis en œuvre, sans chercher à donner une indication, même qualitative, de son ampleur, et surtout sans la moindre évocation d'éventuelles incertitudes relatives à la validité de ces modèles eux-mêmes.

6.3.3 Modalités de prise en compte du risque sismique dans la sûreté nucléaire

6.3.3.1 Principes généraux

La résistance aux séismes ne faisait pas partie des préoccupations de sûreté initiales lors de la construction des premiers réacteurs civils aux États-Unis. La réglementation américaine dans ce domaine a vu le jour au cours des années 1960, et s'est renforcée avec l'adoption des critères généraux de conception des réacteurs en 1971¹⁹. Le deuxième de ces critères exigeait notamment que l'on tienne compte, dans la conception des systèmes et composants importants pour la sûreté, des aléas naturels les plus graves ayant été observés sur le site et dans sa région, en incorporant une marge suffisante pour compenser l'insuffisance en qualité et en quantité des données historiques.

L'approche suivie s'inscrivait bien sûr dans la démarche déterministe. Elle avait la particularité de distinguer deux événements de référence : un séisme dont l'occurrence pendant la durée de vie de l'installation était considérée crédible, et qui servait de base à la conception générale, et un séisme maximal que l'on retenait pour la construction de scénarios d'accident les plus graves. Ce dernier, que l'on a appelé initialement *Maximum hypothetical earthquake* (MHE) puis *Safe shutdown earthquake* (SSE), était basé sur les données historiques : c'était le séisme le plus important jamais enregistré dans un rayon de 300 km de la centrale, et dont l'épicentre était translaté sur le site de celle-ci (Stevenson, 2003). À partir des caractéristiques de cet événement, un spectre de réponse simulant le mouvement au sol sur le site était calculé. Sa valeur la plus élevée indiquait le pic d'accélération au sol, dont on se servait pour définir les contraintes subies par les installations. Le séisme de référence plus limité, appelé *Maximum credible earthquake* (MCE) et plus tard *Operating*

19. *General design criteria for nuclear power plants*, 10CFR50 annex A. Voir la sous-section 4.2.1.1.

basis earthquake (OBE), était généralement un événement conventionnel obtenu en divisant par deux cette valeur du pic d'accélération au sol. Ces modes de calcul ont été graduellement affinés, avec notamment la distinction des types de sols, celle des mouvements horizontaux et verticaux et la prise en compte de la réponse des structures à partir de la publication du guide réglementaire 1.60 de l'Atomic Energy Commission en 1973. L'une des innovations du guide était aussi de cumuler un écart-type au spectre moyen estimé à partir d'un échantillon d'événements passés afin de tenir compte de l'incertitude associée à l'estimation du mouvement sismique.

La méthodologie développée aux États-Unis au cours de ces années a été reprise par de nombreux pays²⁰ et par l'AIEA, qui a adopté en 1979 la dénomination internationale S1 et S2 pour désigner l'équivalent des séismes OBE et SSE (AIEA, 1979).

Une approche probabiliste de l'évaluation des aléas sismiques s'est développée à partir de l'article fondateur de Cornell (1968). Elle s'intéresse non pas à un événement de référence, mais à l'ensemble des événements pertinents à partir desquels elle estime une fonction de répartition des mouvements au sol (c'est-à-dire une fonction associant à chaque valeur de mouvement au sol sa probabilité annuelle de dépassement). Elle étudie ensuite les conséquences de ces mouvements pour la tenue des structures.

En outre, comme dans d'autres domaines, les évaluations déterministes aussi bien que probabilistes se sont progressivement départies de la juxtaposition indistincte d'hypothèses conservatrices, pour s'orienter vers des estimations plus réalistes assorties de prises en compte explicites de l'incertitude. Au fil des ans, l'amélioration des connaissances (tectonique des plaques, tenue des structures) et l'évolution des outils techniques (imageries satellite et Doppler, élargissement des bases de données, accroissement vertigineux des capacités de calcul) ont entraîné des révisions régulières des réglementations sismiques, qui se sont souvent traduites par un *rehaussement* des exigences, sans toutefois que celles-ci soient toujours appliquées de façon rétroactive aux centrales existantes²¹.

6.3.3.2 La réglementation sismique dans le secteur nucléaire japonais

Le secteur nucléaire japonais a dû lui aussi prendre des dispositions de sûreté particulièrement draconiennes en raison de l'extraordinaire exposition du territoire du Japon aux risques sismiques. En premier lieu, les centrales japonaises ont toutes été situées sur des terrains rocheux, où les bâti-

20. C'est notamment le cas de la France, où les deux événements de référence ont été appelés Séisme majeur de sécurité (SMS) et Séisme maximal historiquement vraisemblable (SMVH). En raison de la faible sismicité du territoire métropolitain, c'est le SMVH qui a été défini à partir des maxima historiques ; le SMS a été dérivé en augmentant d'un point l'intensité du SMVH sur l'échelle de Medvedev-Sponhauer-Karnik (MSK), ce qui correspond à peu près à un doublement des paramètres de mouvement au sol. Le rapport entre le SMVH et le SMS en France est donc similaire à celui entre SSE et OBE aux États-Unis.

En revanche, certains pays tels que le Royaume-Uni, la Suède, la Finlande et les pays d'Europe Centrale et Orientale, n'ont imposé aucune norme sismique à leurs centrales jusqu'au milieu des années 1980.

21. L'une des principales leçons tirées par la mission d'étude de la NRC à Fukushima est précisément la nécessité d'analyses sismiques spécifiques à chaque site aux États-Unis permettant de situer les installations par rapport aux normes en vigueur (NRC, 2011, pp.29-30).

ments abritant les réacteurs ont été en partie construits en sous-sol, souvent sur plusieurs dizaines de mètres, ce qui leur permet de bénéficier d'une atténuation importante des ondes sismiques. À la centrale de Kashiwazaki-Kariwa, par exemple, les mouvements enregistrés au niveau des fondations des bâtiments de réacteurs sont estimées représenter entre 40 et 60 % de ceux à la surface de l'assise rocheuse (NSC, 2009, p.48). Chaque centrale a en outre été dotée d'un réseau de capteurs sismiques pouvant commander le système d'arrêt d'urgence des réacteurs dès qu'un certain niveau de secousse est dépassé.

Les responsables japonais de la sûreté ont également développé une approche originale sur le plan méthodologique.

Telle qu'initialement élaborée en 1978, la réglementation japonaise relative au dimensionnement sismique des installations nucléaires retenait le principe américain du double séisme de référence pour chaque site. Mais outre la prise en compte de tous les tremblements de terre enregistrés dans la région, elle supposait, au moins dans ses principes, un effort particulier d'identification des failles susceptibles d'être actives, notamment par l'analyse géologique. La période d'activité considérée était de 10 000 ans pour le « séisme de dimensionnement maximal » (équivalent du S1) et de 50 000 ans pour le calcul du « séisme de dimensionnement extrême » (équivalent du séisme S2). Pour ce dernier, elle imposait de surcroît, la prise en compte d'un séisme « hypothétique » de faible profondeur et de magnitude Mw 6,5, représentant la famille d'événements que l'on ne parvenait pas à associer à une faille spécifique.

Par ailleurs, lors de la construction des premières centrales, les bâtiments, systèmes ou composants des installations nucléaires ont été répartis en quatre classes selon leur importance pour la sûreté. La première regroupait les éléments les plus critiques, dont la défaillance causerait une perte de liquide réfrigérant (y compris celui de la piscine d'entreposage du combustible usé), de confinement ou de contrôle de la criticité. Ces équipements devaient être capables de conserver leurs fonctions de sûreté après un séisme S2 et leur élasticité durant un séisme S1. La deuxième classe contenait les autres équipements importants pour la sûreté, tels que le système d'injection de secours ou celui de refroidissement à l'arrêt, auxquels seule la dernière des deux exigences précédentes s'appliquait. Bien que moins essentiels du point de vue de la sûreté, les équipements de la troisième classe renfermaient des produits radioactifs à haute activité, et faisaient à ce titre l'objet d'exigences renforcées par rapport aux normes sismiques applicables aux constructions non nucléaires. Enfin la quatrième classe relevait de ces normes générales.

En 2006, la réglementation japonaise a fait, à son tour, l'objet d'une réforme importante destinée à tirer des enseignements de certains événements sismiques et à renouveler son outillage scientifique (NSC, 2006). Le double séisme de référence a été abandonné au profit d'un seul événement « enveloppe », pour lequel sont calculées deux valeurs de mouvement au sol, l'une « de sûreté », considérée

maximale, et l'autre « de dimensionnement », égale à une fraction supérieure à la moitié de la première. La recherche des failles actives a été intensifiée par le biais de l'introduction de nouvelles méthodes de relevé et d'analyse géologique et d'une extension de la période d'activité considérée à la fin du Pleistocène (80 000 à 130 000 ans avant notre ère). L'approche empirique traditionnelle d'évaluation des spectres de réponse a été complétée par une modélisation de la réponse spécifique à chaque faille. Pour représenter le tremblement de terre « hypothétique » de faible profondeur, un spectre de réponse a été estimé par des méthodes probabilistes à partir de l'ensemble des séismes observés sans que leur source soit identifiée. Une composante verticale a été introduite dans les calculs dynamiques d'accélération. Les deux premières catégories de la classification des équipements ont été fusionnées et les exigences réglementaires relatives à la seconde ont été harmonisées avec celles de la première. On a prit le parti, enfin, d'évaluer la probabilité de dépassement de la valeur d'accélération au sol retenue et d'en déduire, à l'aide d'une EPS intégrant les effets du séisme, une valeur de risque résiduel.

Le guide réglementaire de 2006 introduisait également une exigence explicite de prise en compte des aléas associés aux séismes. Au sujet des tsunamis, il recommandait qu'un niveau d'eau maximal soit déterminé en tenant compte à la fois de l'ensemble des précédents historiques et des résultats de modélisation, avec pour ces derniers l'inclusion d'une marge raisonnable d'incertitude. Le guide reprenait en pratique les grandes lignes d'une méthode d'évaluation des tsunamis à usage des centrales électronucléaires publiée en 2002 par la société japonaise des ingénieurs civils. La méthode indique qu'en raison de l'ampleur des incertitudes affectant l'évaluation des tsunamis et de la diversité de leurs sources, il est important de procéder à un nombre important de simulations en faisant varier les séismes de référence et les valeurs paramétriques, et de retenir la valeur maximale entre toutes celles obtenues et les précédents historiques (JSCE, 2002).

6.3.3.3 Des doutes croissants

La mise en œuvre de cette réglementation et sa capacité à protéger les centrales nucléaires japonaises des risques sismiques a cependant fait l'objet d'interrogations croissantes au cours des dernières années.

Le désastre de Kobe ébranla quelque peu la confiance dans la capacité des scientifiques à identifier les aléas sismiques et à anticiper leurs effets en surface, puisque la faille, bien que connue, était jusqu'à lors considérée peu dangereuse. Ce doute fut renforcé par la survenue, au cours des années suivantes, de plusieurs séismes de faible profondeur et d'une magnitude proche de Mw 7 à partir de failles inconnues, notamment celui de Chūetsu en 2004, durant lequel des secousses exceptionnelles furent observées.

Le secteur nucléaire avait été relativement épargné par ces événements, les centrales situées

dans les zones affectées étant à chaque fois restées à l'intérieur de leurs limites de dimensionnement et n'ayant subi aucun dommage significatif. Il fut cependant touché à son tour le 16 juillet 2007, lorsqu'un séisme de magnitude Mw 6,6, dont l'épicentre était situé en mer du Japon, affecta la centrale nucléaire de Kashiwazaki-Kariwa. Avec 7 réacteurs et une puissance totale de 8 200 MW_e , la centrale était alors le plus grand site de production d'électricité au monde. Ses spécifications para-sismiques étaient comprises entre 0,17 g dans l'unité 2 et 0,28 g dans l'unité 1, où furent enregistrées des accélérations au sol de 0,62 g et 0,69 g respectivement. Le séisme eut pour conséquence directe des fuites mineures de liquide et de gaz radioactifs et un début d'incendie dans un bâtiment secondaire. Aucun dommage visible ne fut subi par les équipements importants pour la sûreté, ce qui montra que l'accumulation d'hypothèses conservatrices dans la conception de la centrale avait suffi à compenser l'insuffisance de la valeur de dimensionnement. Mais l'ampleur des secousses avait à ce point dépassé les prévisions des concepteurs et l'étendue des marges effectivement disponibles était si incertaine que la bonne tenue des installations constitua une forme d'heureuse surprise. Pour la première fois, le Japon fit appel à une mission d'enquête de l'AIEA pour qu'elle contribue à établir un retour d'expérience de l'événement. Bien que se montrant rassurante quant à la sûreté de la centrale, celle-ci souligna la nécessité de réévaluer l'aléa sismique à Kashiwazaki et en particulier de déterminer si une faille active était présente sous la centrale.

Des investigations ultérieures montrèrent en effet que la faille responsable du séisme de Niigataken Chūetsu-Oki était considérablement plus longue que ne l'avaient estimé les autorités de sûreté et que, contrairement à ce que pensaient ces dernières, elle était active. Il apparut surtout que l'exploitant des installations, l'entreprise TEPCO, avait découvert ces faits en 2003 mais avait omis de les communiquer aux instances publiques (NSC, 2009). Certains sismologues soutinrent que dans d'autres cas aussi, des failles connues n'avaient pas été prises en compte lors des procédures d'autorisation et d'inspection de centrales (Ishibashi, 2007).

La nouvelle réglementation de 2006 constituait un début de réaction des autorités japonaises à ces événements. Son adoption et la remise en cause provoquée par le séisme de Niigataken Chūetsu-Oki conduisirent à un relèvement des normes sismiques dans toutes les centrales japonaises, parfois de façon spectaculaire. À Kashiwazaki-Kariwa, qui fut la première centrale faisant l'objet d'une revue basée sur le nouveau guide réglementaire, l'exploitant accepta d'engager des modifications permettant d'augmenter de plus de 0,5 g les valeurs de référence en termes de pic d'accélération au sol, les portant à un niveau compris entre 0,75 g et 0,86 g selon les tranches²². Sous l'effet des

22. À titre de comparaison, la centrale ayant la valeur de dimensionnement la plus élevée aux États-Unis est celle de Diablo Canyon, située dans une zone fortement sismique de Californie, avec 0,75 g . Dans des zones à faible sismicité, la valeur d'accélération fréquemment retenue est de 0,1 g , que de nombreux pays (dont la France) ont imposée comme minimum pour les installations nucléaires. Il faut noter que cette exigence minimale n'est guère contraignante, puisque selon certaines estimations, une norme inférieure n'engendrerait aucune contrainte propre pour la conception de structures telles que des centrales nucléaires ou pour celle de leurs systèmes et composantes (Stevenson, 2003). En France, les valeurs retenues sont généralement comprises entre 0,1 et 0,2 g .

inspections nécessitées par le séisme et les travaux de renforcement, la centrale resta entièrement à l'arrêt pendant 21 mois, avant que quatre de ses tranches ne soient progressivement autorisées à redémarrer entre mai 2009 et novembre 2010. Les trois dernières tranches restent inactives à ce jour.

L'incertitude concernant les séismes de référence dans les zones de subduction et les tsunamis associés, qui était comme nous l'avons vu exclue *de facto* de la méthodologie officielle japonaise, avait également commencé à poindre sur le plan de la pratique. En 2009, huit ans après la publication des résultats scientifiques concernant le méga-tsunami de Jogan, la NISA, l'autorité de sûreté japonaise, demanda à TEPCO de rassembler des informations au sujet du séisme en vue d'une réévaluation de la protection des sites de Fukushima contre les tsunamis (Gouvernement du Japon, 2011, chapitre III, p.31). Le 7 mars 2011, au cours d'une réunion bilatérale, TEPCO exposa à l'Agence le résultat de simulations provisoires des conséquences d'un séisme majeur au large de Fukushima, qui faisaient état d'un tsunami de plus de dix mètres (Gouvernement du Japon, 2011b, chapitre II, pp. 69-70).

6.3.4 Synthèse : le double enseignement de Fukushima

Différentes évaluations probabilistes avaient présentées par l'exploitant et validées par les autorités japonaises avant la catastrophe de Fukushima. Elles comprenaient :

- une probabilité de dépassement des valeurs d'accélération au sol retenues pour le dimensionnement des installations comprise entre 10^{-6} et 10^{-4} ;
- une probabilité de fusion du cœur avec défaillance de l'enceinte de confinement de 10^{-6} ;
- une absence de prise en compte de la possibilité d'accidents simultanés sur plusieurs tranches, qui équivaut à une hypothèse d'indépendance de ces événements.

Une reconstitution *post hoc* de la probabilité de l'accident du 11 mars à partir de ces éléments conduit à une valeur inférieure à 10^{-18} occurrence par an²³, dont une fraction seulement serait attribuable à une agression externe par un séisme ou un tsunami. Il s'agissait donc d'une impossibilité virtuelle, que nombre d'éléments rappelés dans les sections précédentes tendaient pourtant à démentir.

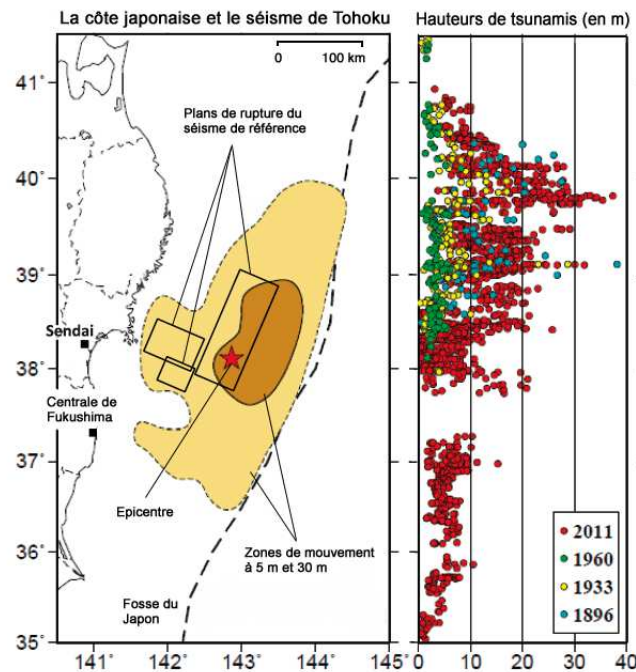
Cette erreur fondamentale peut être interprétée de deux manières, si on la considère du point de vue de l'incertitude.

Selon la première, au vu des précédents historiques, un méga-tsunami apparaît comme un événement millénal dont la possibilité aurait dû être prise en compte dans le dimensionnement de la centrale²⁴. À l'appui de cette thèse, il suffit de considérer la figure 6.3.4 pour mesurer la fragilité des hypothèses retenues dans le choix des événements de dimensionnement à Fukushima. En

23. Puisqu'il faudrait cumuler les défaillances observées dans les tranches 4, 5 et 6 à la fusion avec rejets massifs dans les unités 1, 2 et 3.

24. Certaines déclarations de commissaires de la NRC américaine au cours d'une réunion publique portant sur les enseignements de la catastrophe illustrent cette position (NRC, 2011b, pp.45-6).

FIGURE 6.3.4 – Dépassement des événements de référence à Fukushima



Source : Ando et al.(2011)

particulier, il apparaît que si les responsables japonais de la sûreté avaient adopté une perspective historique moins étroite, ils auraient été amenés à tenir compte de la possibilité d'occurrence de méga-tsunamis dans cette partie de la côte japonaise.

Cette conclusion s'appuie cependant sur une estimation de la probabilité de l'événement ; elle ne traite pas du caractère incertain de cette estimation, qui est également une donnée fondamentale du problème tel que nous l'avons décrit. L'interprétation alternative nous semble plus porteuse. Elle consiste à caractériser d'une part l'attitude des responsables de la sûreté japonais comme l'exact opposé d'une démarche de précaution, et à relever d'autre part la totale vulnérabilité de la centrale à un événement qui, bien que classé hors du dimensionnement, faisait l'objet de certaines indications incitant à la prudence.

6.3.4.1 L'absence de précaution ou les inconvénients d'une science officielle

La position des officiels japonais à l'égard des risques sismiques avant la catastrophe est un cas d'alignement excessif de décideurs publics sur une position scientifique controversée.

Alors que le choix du site de Fukushima pour l'implantation d'une centrale nucléaire semblait judicieux du point de vue de la sismologie du début des années 1970, il en était tout autrement après trois décennies d'évolution rapide des connaissances et des représentations scientifiques. L'attitude de la sismologie officielle japonaise, qui s'est trouvée à contre-courant de cette évolution, s'apparente

à un pari scientifique. Mais l'attention se porte surtout sur l'absence de prise en compte du caractère très incertain de la science officielle par les décideurs publics.

Cette attitude est d'autant plus surprenante que plusieurs événements sont venus au cours des années récentes rappeler l'ampleur des incertitudes dans ce domaine, notamment les tremblements de terre de Hanshin-Awaji en 1995 et de Niigataken Chūetsu-Oki en 2007. Ces séismes n'étaient certes pas du type de celui de Tohoku. Il semble néanmoins qu'ils auraient dû provoquer une remise en question profonde de la démarche de protection des centrales nucléaires contre les risques d'origine sismique. C'est, dans un premier temps au moins, le contraire qui s'est produit puisque la réaction au désastre de Kobe et la réforme réglementaire de 2006 ont renforcé la dépendance de la sûreté nucléaire japonaise à l'égard des modèles prédictifs et des estimations probabilistes. Les interrogations de la NISA au sujet du tsunami de Jogan indiquent qu'un début de prise en compte de l'incertitude affectant ces estimations s'était fait jour ces dernières années, mais elle fut trop tardive.

6.3.4.2 La vulnérabilité de la défense en profondeur

La démarche de défense en profondeur, que nous avons exposée au chapitre 3, consiste à s'assurer que le dimensionnement des installations est suffisant pour permettre, en cas de survenue d'un événement défavorable, le maintien des trois fonctions de sûreté (mise en arrêt et maintien du réacteur en situation sous-critique, évacuation de la puissance résiduelle, confinement des produits radioactifs) ou, à défaut, minimiser les risques liés à la perte de ces fonctions. La défense en profondeur rassemble pour cela un ensemble de structures, de systèmes et de procédures organisés en cinq niveaux. Comme nous l'avons expliqué, l'une des conditions fondamentales de son efficacité est que les défaillances de cause commune des systèmes engagés dans différents niveaux relèvent du risque résiduel.

La catastrophe de Fukushima a mis en évidence un grave défaut dans la conception de la défense en profondeur : le fait que les conséquences des événements « hors dimensionnement », c'est-à-dire ceux qui sont d'ampleur plus grande et de probabilité a priori moindre que les événements maximaux retenus pour le dimensionnement, ne sont pas étudiés. De ce fait, en présence de tels événements, des cas d'extrême vulnérabilité, que l'on qualifie d'« effet falaise », ne peuvent être exclus. Ceci s'est vérifié à Fukushima en dépit des modifications engagées pour renforcer la protection du site contre les risques d'inondation, notamment la surélévation de trois groupes électrogènes. L'efficacité réelle de ces modifications ne semble pas avoir fait l'objet d'un examen d'ensemble approfondi, si l'on peut en juger par le mode de défaillance de ces trois sources.

Notons également que cette faiblesse de la défense en profondeur n'est corrigée par les évaluations probabilistes que dans la mesure où celles-ci intègrent les événements en question. Si, par exemple,

ces événements relevaient de cas d'incertitude de modélisation, alors en l'état des pratiques d'EPS, ils auraient de grandes chances d'être ignorés tout autant que dans l'approche déterministe. C'est ce qui s'est produit à Fukushima.

Au-delà du cas de Fukushima, ces observations nous ramènent donc à nos conclusions des chapitres 3 et 4 : le manque, dans les doctrines de sûreté actuelles, d'une démarche systématique de reconnaissance et de prise en compte des incertitudes relatives aux choix de dimensionnement.

6.4 Après Fukushima

La catastrophe de Fukushima a eu un retentissement international considérable et sérieusement perturbé le développement de l'industrie électronucléaire dans le monde.

Sur le strict plan de la sûreté, les gouvernements des pays possédant une industrie nucléaire civile se sont montrés particulièrement actifs afin de tirer des enseignements de la catastrophe. Dès le 23 mars 2011, le Premier Ministre français demanda au Président de l'Autorité de Sûreté Nucléaire (ASN) de « réaliser une étude de sûreté des installations nucléaires, en priorité des centrales nucléaires, au regard de l'accident en cours ». En conclusion de sa réunion des 24 et 25 mars, le Conseil européen déclara également qu'il convenait « de vérifier la sûreté des toutes les installations nucléaires de l'UE, sur la base d'une évaluation globale et transparente des risques et de la sûreté ». Le Groupe des régulateurs européens dans le domaine de la sûreté nucléaire (ENSREG) définit le cahier des charges des tests de résistance (*stress tests*) auxquels les installations nucléaires des pays membres devraient être soumis. En France, des évaluations complémentaires de sûreté (ECS) furent lancées début mai dans un climat d'urgence, puisqu'il fut exigé des exploitants d'en délivrer les résultats avant le 15 septembre 2011, afin que ceux-ci soient immédiatement expertisés par l'Institut de Radioprotection et de Sûreté Nucléaire (IRSN) et que les dispositions nécessaires soient prises par l'ASN avant la fin de l'année.

Ces initiatives politiques ont encadré les premières étapes d'un processus de retour d'expérience fondé, comme après chaque accident grave, sur l'analyse minutieuse des données collectées. Ce n'est qu'au fil du déroulement de ce processus, qui prendra plusieurs années, qu'il deviendra possible de tirer les enseignements du désastre de Fukushima pour la sûreté nucléaire. Notre propos dans cette section sera donc beaucoup plus limité : mettre en évidence certains aspects des résultats des ECS français et essayer, à la lumière de l'analyse partielle menée dans les sections précédentes, de dégager quelques pistes de réflexion.

6.4.1 Les évaluations complémentaires de sûreté (ECS) en France

6.4.1.1 Grandes lignes des ECS

La méthodologie des évaluations complémentaires de sûreté a été mise au point conjointement par les exploitants et l'IRSN. Elle consiste à compléter l'analyse traditionnelle de la sûreté par l'identification d'un « noyau dur » de structures, systèmes et composants qui, placés dans des situations extrêmes, doivent demeurer robustes et suffire à préserver les trois fonctions de sûreté de la centrale : la maîtrise de la réaction en chaîne, le confinement permanent de la matière radioactive et le refroidissement des barres de combustible. Elle n'est donc pas, au moins dans ses principes, sans rappeler la démarche suivie au Japon en matière de résistance aux séismes. La palette de situations extrêmes envisagées est directement inspirée du scénario de Fukushima : agressions extérieures d'ampleur supérieure à celles considérées dans les référentiels de sûreté, agressions induites (incendie provoqué par une agression extérieure, explosion d'hydrogène), perte de fonctions de refroidissement ou perte de longue durée de sources électriques affectant l'ensemble des installations d'un site. La réalisation de chacun de ces scénarios est supposée afin d'analyser la tenue des équipements et les conséquences générales pour l'installation. La démarche s'inscrit donc pleinement dans la philosophie de la défense en profondeur.

Les résultats communiqués par les exploitants ont reçu un accueil globalement favorable de la part de l'IRSN (IRSN, 2011b) et, à sa suite, des groupes permanents d'experts pour les réacteurs et les usines et laboratoires chargés de conseiller l'ASN (Groupes Permanents Réacteurs et Usines, 2011). Il a notamment été estimé que les ECS ne révélaient pas de lacune significative à l'égard des agressions considérées et que les quelques insuffisances identifiées pourraient faire l'objet de mesures correctrices dans le cadre des réexamens décennaux, dont le rôle de pièce maîtresse de l'approche de sûreté française a été réaffirmé. Il a été décidé de laisser à chaque exploitant le soin de définir les éléments du « noyau dur » spécifiques à chacune de ses installations, ainsi que les niveaux d'aléas retenus pour leur dimensionnement. On peut estimer probable que les autorités de sûreté seront rapidement amenées à encadrer les pratiques dans ce domaine par le biais d'une réglementation spécifique, ou au moins d'un ensemble d'indications générales. Pour les groupes permanents, la *robustesse* de ce noyau dur doit désormais être considérée comme l'un des éléments de la défense en profondeur.

Un premier bilan du processus des ECS, nécessairement provisoire puisqu'aucune décision concrète n'est encore venue le sanctionner, pourrait donc être ainsi formulée : une évolution notable de la doctrine française de sûreté aux implications pratiques limitées. Mais cette appréciation superficielle manquerait de percevoir dans cet exercice l'écho de certaines interrogations profondes.

6.4.1.2 La question de l'incertitude au centre des ECS

De façon remarquable, le processus des ECS a été l'occasion de pointer certaines insuffisances de l'approche suivie jusqu'à ce jour à *l'intérieur* des référentiels de sûreté existants, c'est-à-dire sans qu'il soit nécessaire d'envisager la survenue d'événements plus graves. Trois exemples peuvent en être donnés ici : la question des cumuls conventionnels dans la définition de la défense en profondeur, le problème des écarts de conformité et celui de l'état des connaissances relatives aux aléas naturels considérés.

Dans l'appréciation initiale qu'il a portée sur la façon dont les exploitants proposaient de mettre en œuvre les ECS, l'IRSN a estimé que « qu'ÉdF devra[it] vérifier la robustesse des dispositions et matériels essentiels à la gestion d'une perte totale de la source froide et/ou d'une perte totale des alimentations électriques, ainsi qu'à la limitation des rejets en cas d'accident grave (accident avec fusion du cœur), en tenant compte des agressions ou des événements induits (incendie, explosion ...) par un séisme ou une inondation » (IRSN, 2011a, p.3). L'argument a été avancé de façon encore plus nette par les groupes permanents réacteurs et usines lorsqu'ils se sont penchés sur les résultats des ECS : « les référentiels actuels ne considèrent pas le cumul des situations de perte totale des alimentations électriques ou de sources de refroidissement avec les agressions externes prises en compte dans les référentiels : de ce fait, les équipements nécessaires à la gestion de ces situations ne sont pas tous protégés ou dimensionnés aux niveaux d'aléas du domaine de dimensionnement » (Groupes Permanents Réacteurs et Usines, 2011, p.3).

Cette lacune est une illustration typique des limites de l'approche déterministe que nous avons cherché à mettre en exergue au chapitre 3²⁵. Les cumuls conventionnels, qui permettent d'envisager la survenue simultanée de plusieurs événements défavorables, que ceux-ci soient liés par des relations de cause à effet ou que leur coïncidence soit fortuite, sont en effet choisis de façon arbitraire. En France, il a ainsi été convenu de se placer sous l'hypothèse d'occurrence conjointe d'une grosse brèche sur le circuit primaire, d'un séisme majoré de sécurité et de la perte des sources électriques extérieures. La perte totale des alimentations électriques et celle de sources de refroidissement ont été envisagées séparément, en tant que conditions de fonctionnement d'une grande gravité, mais dont la probabilité semblait suffisamment faible pour que l'on n'ait pas, en outre, à envisager leur cumul avec un autre événement. Seule une démarche d'inférence systématique et quantifiée telle que celle des EPS permettrait de déterminer si un tel cumul contribue significativement au risque total et, le cas échéant, d'explorer ses conséquences et de déterminer les actions correctrices nécessaires.

Dans son rapport d'évaluation du résultat des ECS, l'IRSN relève deux autres points de faiblesse du dispositif général de sûreté pour les aléas considérés dans le référentiel : premièrement, le risque que, sous l'effet des modifications successives apportées aux installations, leur comportement face à

25. Voir en particulier la discussion des conditions théoriques de la défense en profondeur dans la section 3.3.3.

une agression finisse par ne plus être conforme à celui stipulé dans les référentiels de sûreté (« écart de conformité ») ; deuxièmement, la possibilité que l'évolution des connaissances relatives aux aléas justifie un relèvement des exigences à l'intérieur du référentiel de sûreté, mais que ce relèvement tarde à être mis en place (IRSN, 2011b, p.3). Dans un cas comme dans l'autre, les réexamens décennaux doivent permettre d'apporter les corrections nécessaires. Mais dans l'intervalle entre deux réexamens, aucun dispositif ne permet aujourd'hui de combler ces deux faiblesses.

Un réexamen d'ensemble comme celui des ECS amène inévitablement à la question des écarts de conformité : avant de se demander comment des exigences supérieures pourraient être satisfaites, il s'agit en quelque sorte de veiller à ce que celles du référentiel de sûreté existant soit respectées. L'IRSN s'estime dans l'incapacité de l'affirmer sur la base des informations « peu détaillées » fournies par les exploitants dans le cadre de leurs obligations réglementaires. Il semble que les ECS aient fourni l'opportunité de réaliser quelques progrès dans ce domaine. Tous les exploitants se sont en effet engagés à confirmer avant la fin de l'année 2012 l'absence d'écart pour une partie des structures, systèmes et composantes du « noyau dur », ceux nécessaires à la gestion d'accidents de perte des sources électriques ou de perte de refroidissement. L'IRSN plaide, cependant, en faveur d'une évolution plus radicale qui verrait la mise en place d'un pilotage d'ensemble des modifications et de leur traçabilité, dans le but d'assurer le maintien constant de la conformité. Mais il ne semble pas qu'une telle évolution soit à l'ordre du jour.

La question de la connaissance des aléas de référence actuels s'impose tout aussi naturellement dans l'examen des ECS. Les exploitants ont en effet fondé leurs évaluations sur des scénarios d'aléas majorés et il faut déterminer si ceux-ci représentent une aggravation suffisante par rapport aux aléas de référence. La réponse de l'IRSN est généralement nuancée, et fait apparaître une prise en compte insuffisante, selon l'Institut, de l'incertitude relative aux événements naturels de référence dans les exigences de dimensionnement actuelles. Au sujet de l'aléa sismique, à titre d'exemple, l'IRSN estime que l'état de l'art justifierait une amélioration des pratiques dans trois directions : l'enrichissement des données utilisées, l'intégration explicite des incertitudes selon des méthodes déterministes et probabilistes et la prise en compte de la diversité des avis d'experts.

Comme on le voit, les trois sujets se ramènent à une discussion des incertitudes et limites de la démarche existante de défense en profondeur, qui finissent par occuper une place centrale dans l'évaluation des ECS. Sur chacun des sujets, les observations de l'Institut nous semblent ainsi éclairer un aspect particulier du manque plus général que nous avons identifié plus tôt : celui d'une démarche d'identification systématique des principales incertitudes relatives à la défense en profondeur de chaque installation, fondée sur l'approche probabiliste et prenant en considération les conditions spécifiques de l'installation.

La section suivante propose une analyse plus approfondie du rapport entre ECS, incertitudes et

limites de la démarche existante dans le cas particulier des risques liés aux inondations.

6.4.2 L'exemple des inondations

6.4.2.1 Le référentiel de sûreté : la RFS I.2.e

Un exemple de relation structurelle

La protection des centrales nucléaires contre les risques d'inondations d'origine externe est encadrée depuis 1984 par la règle fondamentale de sûreté (RFS) I.2.e (SCSIN, 1984). La règle distingue cinq types d'aléas pouvant causer une inondation d'origine externe : la crue fluviale, la crue en bord de mer par surcote, la crue en bord de mer par tsunami, la rupture ou l'effacement de barrage et la crue en estuaire. Pour chacun de ces types, elle spécifie une méthode « acceptable » de détermination de la hauteur d'eau qui doit servir de référence pour la détermination des protections. Cette hauteur est désignée comme la cote majorée de sécurité (CMS). La règle décrit également les principes de conception qui doivent permettre la sauvegarde des fonctions de sûreté de la centrale en cas d'inondation.

Nous nous intéresserons spécifiquement à l'aléa de crue en bord de mer par surcote. Dans ce cas, la CMS doit correspondre à « la conjonction de la marée astronomique maximale calculée (coefficient 120) et de la surcote marine millénale » :

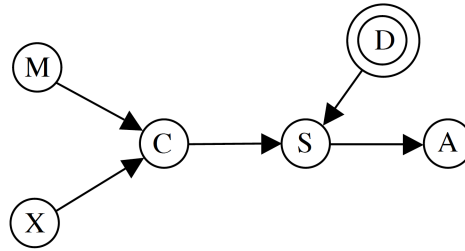
$$CMS = \text{marée de coefficient 120} + \text{surcote millénale} \quad (6.4.1)$$

La surcote est définie comme la différence entre le niveau observé de la mer et celui prévu par la marée astronomique²⁶. Elle est en soumise en particulier à un ensemble d'influences atmosphériques et climatiques appelé forçage atmosphérique. L'état de l'art de la modélisation en hydrologie la représente comme une fonction des champs de vent, de la pression atmosphérique, voire de la température et de la salinité de l'eau (DEFRA, 2004). L'estimation de ces modèles requiert toutefois des quantités importantes de données, qui sont généralement indisponibles si l'on s'intéresse à des périodes longues permettant d'observer des phénomènes extrêmes. Dans de tels cas, où les résultats des modèles sont affectés de fortes incertitudes, une approche alternative est de recourir à une analyse purement statistique de la surcote. Elle consiste à ajuster une loi de probabilités P à la chronique des surcotes observées en un lieu spécifique, puis d'en déduire la valeur x_m , dite de surcote millénale, telle que $P(X \geq x_m) = 0,001$.

C'est cette approche qui est préconisée par la RFS I.2.e pour l'évaluation des CMS. La définition (6.4.1) de la CMS découle donc d'une vision de la cote C comme somme d'une composante M déterminée par la marée astronomique et d'un terme résiduel X , qui n'a pas de signification

²⁶. Plus précisément, le terme surcote désigne cette différence lorsqu'elle est positive. Dans le cas contraire, on parle de décote.

FIGURE 6.4.1 – Risques d’inondation liée à une crue en bord de mer par surcote



physique propre, agrège l'effet d'un ensemble de variables latentes et constitue la part aléatoire de la représentation²⁷ :

$$C = M + X$$

Cette représentation est donc un exemple typique des relations structurelles décrites au chapitre 5. En la prolongeant par une représentation simplifiée de la sûreté d'une installation nucléaire, on obtient un réseau bayésien causal tel que celui de la figure 6.4.1, où S et A désignent comme auparavant l'état des systèmes importants pour la sûreté et l'accident grave, et D est une intervention de protection, telle que la construction d'une digue, calibrée conformément à la RFS. Le mode de détermination de la CMS correspond à une démarche principalement déterministe, puisque l'on retient la marée de plus fort coefficient²⁸ et que l'on choisit de cumuler les maxima des deux composantes de la cote, mais qui intègre une représentation probabiliste de la surcote.

Modalités d'évaluation de l'aléa

La RFS I.2.e contient des spécifications techniques détaillées à propos de la méthode d'évaluation de l'aléa de crue en bord de mer par surcote. Elle prescrit, sauf exception, l'usage de la méthode statistique dite du renouvellement. Celle-ci consiste à construire, à partir d'une suite d'observations à haute fréquence (généralement horaire), une sous-suite de valeurs en excès d'un seuil donné, que l'on considère comme les réalisations d'une variable aléatoire décrivant le niveau maximal atteint localement, puis d'estimer la loi de cette variable par les méthodes empiriques de la statistique classique (c'est-à-dire par le maximum de vraisemblance) (Miquel, 2006). Une fois estimées les paramètres de cette loi, il ne reste plus qu'à en dériver la valeur de la surcote millénale.

Outre l'incertitude inhérente à cette estimation, la méthode repose sur des hypothèses bien connues dont la validité ne peut être établie de façon certaine. La distribution de la variable sous-

27. L'estimation de la marée astronomique n'est pas totalement exempte d'incertitude, puisque la marée est la somme d'une multitude de composantes harmoniques de périodes, amplitudes et phases différentes. Son calcul est donc d'autant plus précis que le nombre de composantes prises en compte est important et que leurs caractéristiques sont bien connues. En pratique, toutefois, l'essentiel de l'incertitude porte sur le terme de surcote.

28. Le coefficient 120 ne correspond en fait pas toujours à la marée astronomique maximale, mais il semble qu'il s'agisse là d'une approximation dans la formulation de la règle.

jacente doit être stationnaire, ce que l'on teste généralement en vérifiant la stationnarité aux premier et second ordres de la distribution observée²⁹, et en redressant les observations si nécessaire. Ses réalisations doivent en outre être indépendantes. En pratique, on se contente généralement de vérifier que des maxima successifs n'ont pas été artificiellement engendrés partir du même événement, en leur imposant un critère de sélection simple. Enfin, l'échantillon d'observations doit être représentatif de la loi sous-jacente. Cette dernière hypothèse, qui est peut-être la plus hardie notamment en raison de l'inadéquation entre les périodes d'observation (de l'ordre du siècle) et les horizons de prévision (de l'ordre du millénaire), rencontre de grandes difficultés pratiques comme nous le verrons sous peu.

La RFS traite de la question des incertitudes. Elle commence par procéder à une classification quelque peu atypique de leurs formes : incertitudes relatives aux données, au choix du modèle statistique, au choix du modèle hydraulique, à la taille de l'échantillon, et enfin à sa représentativité. Elle estime que les trois premières formes se prêtent difficilement à la quantification et s'en remet en la matière au jugement de l'expert, pourvu qu'il soit motivé par un consensus scientifique ou des études de sensibilité. Pour tenir compte de la quatrième forme d'incertitude, elle exige que des intervalles de confiance soit établis pour l'estimation de la surcote millénale, et que la valeur de référence choisie soit la borne supérieure de l'intervalle à 70 %. Enfin, elle estime que la prise en compte de l'échantillon le plus large possible, le choix du meilleur modèle statistique et la marge introduite dans le calcul de la valeur millénale suffisent à compenser la dernière forme d'incertitude.

La règle contient une mise en garde générale au sujet de l'ampleur des incertitudes, qui rendrait « illusoire » toute considération d'événements plus que millénaux (de probabilité inférieure à 10^{-3}). Elle comprend également une clause de prudence selon laquelle des évolutions dans les connaissances ou les méthodes pourraient amener l'administration à demander aux exploitants la mise en place de dispositions complémentaires.

Cependant, par son approche, la règle elle-même contrevient à cet appel à la prise de distance vis-à-vis des méthodes employées. Le caractère détaillé de ses prescriptions, qui vont jusqu'à préciser la loi de probabilité « généralement » la plus appropriée pour rendre compte des observations de surcotes³⁰, la rend tributaire d'un état des connaissances et de certaines pratiques d'évaluation. Cette dépendance est aggravée par le fait que la règle propose de traiter de l'incertitude essentiellement par le biais d'un critère quantitatif fixe tel que la borne supérieure d'un intervalle de confiance. Par nature, l'incertitude dépend du contexte. Il ne peut être exclu qu'une observation amène à réviser simultanément à la hausse plusieurs de ses composantes considérées par la règle. En ne tenant pas compte de cette possibilité, la règle s'expose à ce qu'un seul événement démontre

29. C'est-à-dire la stabilité de ses moments d'ordre un (espérance) et deux (covariance).

30. « Généralement, les fonctions [de répartition] G ajustées aux données disponibles se rattachent à la loi dite de Weibull » : $G_{q_0}(q) = 1 - \exp(-(\frac{q-q_0}{\lambda})^k)$, où q_0 est le seuil retenu pour la construction de l'échantillon de surcotes et k et λ sont les paramètres à estimer.

qu'elle avait sous-estimé l'incertitude et la rende ainsi caduque. La sous-section suivante décrit un tel événement et ses conséquences.

6.4.2.2 Les insuffisances de la méthode

Le précédent du Blayais

En décembre 1999, la survenue des tempêtes Lothar et Martin se traduisit par des conditions climatiques exceptionnelles sur la côte Atlantique, et notamment par la concomitance d'un niveau d'eau très élevé et d'une forte houle. Dans la nuit du 27 au 28 décembre, des vagues remontant la Gironde submergèrent la digue protégeant la centrale nucléaire du Blayais et inondèrent la plateforme sur laquelle elle est située, en bordure d'estuaire. En raison de la tempête, la centrale avait déjà perdu les sources d'alimentation électrique auxiliaire 225 *kV* de ses quatre tranches, et les sources 400 *kV* de ses tranches 2 et 4. L'eau s'infiltra dans les parties basses des bâtiments et dans leurs galeries techniques, où elle atteint entre un et quatre mètres de hauteur. De nombreux systèmes de sûreté furent affectés par l'inondation, notamment les pompes d'injection à basse pression et d'aspersion dans l'enceinte des tranches 1 et 2 et celles des circuits d'eau brute secourue de l'une des deux voies de la tranche 1.

Les défaillances multiples causées par la tempête et l'inondation ne suffisaient pas à créer une situation d'accident grave, mais en augmentaient fortement la probabilité. En particulier, si le système d'alimentation en eau de secours des générateurs de vapeur était venu à défaillir, le réacteur 1 n'aurait disposé de capacités suffisantes de refroidissement que pendant une dizaine d'heures. Le 28 décembre, la DSIN déclencha le deuxième niveau du Plan d'urgence interne, ce qui entraîna, pour la première fois en France, la mobilisation d'une équipe de crise complète au centre technique de crise de l'IPSN. Mais l'incident ne connut pas de développement défavorable. Dans les jours suivants, 107 000 m^3 d'eau furent pompés depuis les installations et rejetés dans la Gironde.

L'incident du Blayais fut riche en enseignements.

Sur un plan général, il montra, onze ans avant la catastrophe de Fukushima, la vulnérabilité de la défense en profondeur aux événements naturels extrêmes, qui peuvent entraîner la perte simultanée des sources extérieures (alimentation électrique, source froide) et d'un vaste nombre de systèmes importants pour la sûreté. Dans les conclusions d'une analyse publiée quelques mois après l'incident, l'IPSN mettait notamment en exergue la possibilité qu'un aléa affecte plusieurs réacteurs d'un site, ce que de nombreuses sources ont récemment présenté comme l'une des leçons de Fukushima :

« L'inondation qui s'est produite sur le site du Blayais a mis en évidence un mode de dégradation possible du niveau de sûreté de l'ensemble des tranches d'un même site. »

(IPSN, 2000, p.12)

De façon plus particulière, l'épisode mit en lumière plusieurs déficiences de la RFS I.2.e. La CMS

du site avait été évaluée, au moment de la parution de la règle, à 5,02 m. Sur cette base, une digue de 5,20 m avait été construite en front de Gironde. En 1998, une nouvelle estimation du niveau de marée astronomique de coefficient 120 avait conduit à réviser la CMS à 5,46 m, et Électricité de France s'était engagée à rehausser la digue à 5,70 m. Durant l'inondation, l'eau avait franchi des obstacles situés jusqu'à 5,30 m. L'examen hydrologique de l'incident montra en outre le rôle important de la houle, qui n'était pas explicitement pris en compte dans la règle.

Le retour d'expérience du Blayais

L'incident du Blayais conduisit à une réévaluation de la démarche de protection des centrales nucléaires contre les risques d'inondation. Quatre phénomènes susceptibles de mener à une inondation, lorsqu'ils surviennent en conjonction avec un niveau d'eau élevé, furent ajoutés à la liste des aléas : la houle, le clapot³¹, l'intumescence³² et la remontée de nappe phréatique. La méthode de calcul des CMS, bien qu'elle soit restée fidèle aux prescriptions de la RFS, fut améliorée en certains points. Des travaux importants furent entrepris afin de renforcer des ouvrages existants de génie civil, et parfois d'en créer de nouveaux. Au Blayais, la hauteur de la digue en front de Gironde fut portée à 6,20 m et un mur pare-houle de 8,50 m fut construit. Des systèmes d'alerte furent mis en place et des procédures internes développées pour le pilotage de l'installation et, dans certains cas, la prévention des risques liés à la perte de ses sources externes. L'ensemble de ce processus, jusqu'à la validation de la nouvelle démarche par l'ASN, dura près de sept années.

La dernière étape du retour d'expérience de l'inondation du Blayais a été la rédaction d'un guide de l'ASN concernant la protection des installations nucléaires de base contre les inondations externes, appelé à se substituer à la RFS I.2.e. Une version préliminaire de ce guide a été soumise à une consultation publique en 2010 (ASN, 2010). Outre les avancées méthodologiques déjà intégrées dans le retour d'expérience, le projet de guide se distingue de la RFS par un plus grand souci d'exhaustivité et par la volonté d'indiquer, au-delà des instructions détaillées, une démarche générale. Il propose, par exemple, des « paramètres à surveiller » pour chaque situation de référence.

Au chapitre des incertitudes, le projet de guide reprend les grandes lignes de la RFS I.2.e, mais comporte quelques avancées. Tout en maintenant la référence principale à l'intervalle de confiance à 70 %, il précise ainsi que dans certains cas, le choix d'un intervalle plus large, comme celui à 95 %, peut permettre de tenir compte d'incertitudes relatives au modèle statistique employé. Le projet de guide reconnaît également que la représentativité de l'échantillon suppose la stationnarité de la distribution sous-jacente, ce qui « constitue une hypothèse forte pour les aléas influencés par les changements climatiques ou anthropiques ».

Enfin, au sujet de la CMS pour une crue en bord de mer par surcote, le projet de guide ne

31. Influence du vent à la surface d'un fleuve ou d'un canal.

32. Déformation d'un plan d'eau induite par une variation brutale de son écoulement.

cite plus la méthode du renouvellement, mais précise qu’au cas où des événements exceptionnels (« horsains ») auraient été observés dans la région concernée dans le passé, une majoration d’un mètre devrait être appliquée à la cote, et qu’une étude statistique complémentaire devrait établir le caractère adéquat de cette disposition.

Le phénomène des horsains

Les horsains³³ sont des points aberrants dans les chroniques de surcotes. Ils apparaissent dans la majorité des séries longues d’observations et, bien qu’ils aient été identifiés depuis longtemps, ils n’ont pas, à ce jour, reçu d’explication scientifique robuste. La figure 6.4.2 présente un cas de horsain dans un échantillon de surcotes construit en utilisant la méthode du renouvellement à partir des niveaux de mer relevés en port de Brest au cours de 148 années d’observation. La valeur observée le 15 octobre 1987 constituerait un événement dont la période de retour serait de 5 000 ans si l’on se fonde sur la valeur moyenne des paramètres estimés, de 2 000 ans si la référence est la borne supérieure de l’intervalle de confiance à 70 % de cette estimation (ce que recommande la RFS I.2.e) et encore de 1 000 ans si l’on retient l’intervalle à 95 %.

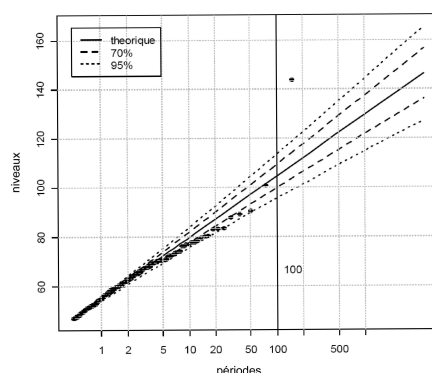
Pour quatre des cinq sites côtiers français accueillant des installations nucléaires, il existe une observation historique de surcote pouvant être qualifiée de horsain. Seuls deux cas correspondent au même épisode climatique, la tempête du 27 décembre 1999 responsable de l’inondation du Blayais. Alors que les périodes d’observations sont de l’ordre de quelques décennies, les périodes de retour attribuées à ces événements en suivant les recommandations de la RFS I.2.e sont de plusieurs siècles. De surcroît, il semblerait que la méthode de constitution de calcul des surcotes employée en France, tout en étant conforme à la RFS, conduise à une sous-estimation de certaines valeurs (IRSN, 2007, annexe 8, pp.10 et suivantes).

La prise en compte de ces observations peut difficilement être rendue compatible avec la méthodologie retenue. Un traitement statistique possible consiste à intégrer le horsain à l’échantillon en temps que donnée historique, en lui associant différentes périodes de retour que l’on estime raisonnables, et d’étudier l’effet de ces variations pour la distribution estimée. Cet effet est généralement négligeable : le horsain demeure « étranger » à la distribution estimée ; la surcote millénaire et ses intervalles de confiance ne varient que de quelques centimètres (IRSN, 2007, 2009b).

Cette stabilité a été interprétée par des exploitants comme le signe de la robustesse des valeurs millénales retenues, et donc des protections conçues sur la foi de ces valeurs, à des variations dans les hypothèses de modélisation. Elle nous semble, au contraire, devoir conduire à mettre en question les fondements de l’approche suivie. Le fait que des valeurs régulièrement observées sur des périodes de l’ordre de quelques décennies apparaissent comme des maxima probables sur des

33. Variante normande du terme français forain (*fors-ain*), le horsain désigne l’étranger, celui qui est venu d’ailleurs. Son équivalent dans la littérature technique en langue anglaise est *outlier*.

FIGURE 6.4.2 – Ajustement d’une loi de Weibull à l’échantillon de surcotes du port de Brest



Source : IRSN (2009b)

échelles temporelles beaucoup plus longues est une indication sérieuse du caractère inadapté de la méthode. Nous rejoignons donc sur ce point l’IRSN, qui a adopté à plusieurs occasions une position critique vis-à-vis de la RFS et des évaluations de CMS auxquelles elle mène :

« L’IRSN À cependant soulevé la difficulté de traiter par la méthode de traitement, un échantillon de valeurs de surcote parmi lesquelles figure une valeur nettement atypique. De façon générale, l’IRSN a formulé une réserve concernant l’application de la méthode de renouvellement préconisée par la RFS. En effet, l’IRSN constate que les méthodes probabilistes ne permettent pas de prendre en compte de manière satisfaisante les événements atypiques (« horsains ») qui apparaissent dans les échantillons de surcotes. L’IRSN a donc estimé que les valeurs de surcotes retenues par ÉdF et calculées en utilisant la méthode du renouvellement, couvriraient les hauteurs maximum historiques observées sans dégager de conservatisme significatif. » (IRSN, 2007, p.1)

Suivant les réserves de l’IRSN sur ce point, le projet de guide sur les inondations impose donc une majoration d’un mètre à la CMS en cas d’existence d’un horsain, ainsi que la conduite d’analyses statistiques spécifiques à ces phénomènes. La piste privilégiée, pour cela, consiste à bâtir, après ajustement statistique, une population de horsains observés dans l’ensemble d’une région et à tenter de les caractériser à partir de cette base élargie.

Cette disposition peut être qualifiée de progrès, puisqu’elle permettrait de tenir compte d’une information qui reste, à ce jour, largement ignorée.

6.4.2.3 Un cas d’incertitude non probabilisable

L’extrapolation des probabilités d’événements rares à partir d’un nombre limité d’observations est un problème bien connu des économètres. Les méthodes d’estimation les plus courantes d’équa-

tions linéaires telles que celle qui soutend la RFS I.2.e supposent vérifiées certaines conditions de régularité, notamment l'absence d'autocorrélation des termes d'erreur, qui sont ici les valeurs de surcote, et leur homoscedasticité (c'est-à-dire la constance dans le temps de leur variance). Les tests statistiques effectués dans l'application de la méthode du renouvellement visent à s'assurer de la validité de ces conditions.

Cependant, ces problèmes classiques ne sont, le plus souvent, que des manifestations d'une difficulté plus fondamentale, celle de parvenir à une spécification adéquate de la grandeur endogène, ici la cote. Une bonne spécification est caractérisée par des termes d'erreur aléatoirement distribués, stables et dont la contribution à la variable endogène est faible. Une mauvaise spécification, au contraire, peut se traduire, outre par l'autocorrélation ou l'hétéroscédasticité des termes d'erreur, par une performance médiocre dans les tests de vraisemblance habituels. Tous ces indicateurs, cependant, sont faussés lorsque le nombre d'observations est trop limité pour faire apparaître les problèmes de spécification.

Une approche alternative est celle de la statistique des extrêmes. Il s'agit là de s'intéresser spécifiquement aux queues de distributions de probabilités sans passer par les hypothèses de régularité communes à la plupart des lois habituelles, par exemple en étudiant à la distribution d'une variable conditionnellement au dépassement d'un certain seuil (méthode appelée *probability over threshold* ou POT). Ces méthodes, qui conduisent typiquement à des distributions empiriques sensiblement plus dispersées, ont connu des applications remarquées en hydrologie.

Considérons à présent le problème dans le cadre du modèle causal exposé plus haut.

Les problèmes de spécification sont la traduction du fait que l'ensemble des causes latentes agrégées dans le terme résiduel comporte une ou plusieurs forces dont l'action est à la fois irrégulière et importante ; lorsqu'elle est présente, elle engendre des valeurs de surcote anormalement élevées.

L'indication de l'existence d'un tel phénomène par la récurrence des horsains, conjuguée à notre incapacité, en l'état de nos connaissances, à l'identifier, forment une situation d'incertitude forte et difficilement probabilisable. Celle-ci affecte en effet l'ensemble de la distribution de la surcote, de sorte qu'il ne peut être exclu que les événements extrêmes soient à la fois plus probables et d'ampleur plus importante que ne le laissent à penser les observations courantes. La statistique des extrêmes n'offre qu'un remède partiel à ce mal en permettant d'étoffer les queues de distribution des probabilités. Le fait scientifique est que nous sommes ignorants des mécanismes fondamentaux à l'œuvre.

Dans une telle situation, l'attitude qui consiste à prendre acte de l'insuffisance des méthodes probabilistes semble être celle qui reflète le mieux l'état de nos connaissances relatives au problème de décision. Il reste à déterminer l'action à laquelle elle doit conduire. La démarche de la défense en

profondeur peut ici servir de guide. Elle préconise d'une part l'incorporation de marges de sûreté, comme le requiert le projet de guide sur les inondations, et d'autre part l'examen du comportement de l'installation sous l'hypothèse pénalisante de dépassement des valeurs de dimensionnement. Ce dernier volet vient d'être mis en place dans le cadre des ECS.

6.5 Conclusion

La catastrophe de Fukushima a montré des défaillances graves dans l'organisation de la sûreté nucléaire au Japon, et notamment dans l'équilibre des responsabilités et des pouvoirs au sein du triptyque autorité de sûreté-organisme de support technique-exploitant. Mais elle est également porteuse de messages plus généraux. Sans prétendre entrer dans un diagnostic technique précis des causes de l'accident, qui ne pourra être établi qu'au terme d'un long travail de collecte et d'analyse de données, nous avons essayé d'en tirer deux conclusions de portée générale. Premièrement, une attitude de précaution, qui se manifeste dans le cas nucléaire par la démarche de la défense en profondeur, impose l'identification et la prise en compte systématique des éléments d'incertitude susceptibles d'avoir un impact significatif sur l'issue d'une décision. Deuxièmement, la défense en profondeur a montré une très grande vulnérabilité dès lors que des événements dommageables dépassaient les hypothèses de dimensionnement.

Les évaluations complémentaires de sûreté semblent indiquer que ces conclusions étaient partagées, au moins en partie, par les acteurs de la sûreté nucléaire en France. Des décisions importantes semblent avoir été prises pour élargir les bases de la défense en profondeur et accroître sa robustesse face aux aléas naturels. L'analyse de la gestion des risques d'inondation dans les centrales nucléaires montre l'apport concret des ECS à la démarche de sûreté française.

La limite des ECS vient de ce que leur champ semble circonscrit aux types de phénomènes observés à Fukushima. L'élaboration des ECS elle-même a pourtant montré que la question était plus vaste : il existe, à la fois à l'intérieur du référentiel de sûreté et dans la façon dont il est mis en œuvre, des zones d'incertitude qui mériteraient d'être explorées afin que l'on puisse s'assurer, au cas par cas, de la robustesse de la sûreté. C'est à une pratique continue et compréhensive des tests de résistance qu'invite cette observation.

Il est vrai, comme l'indique cet épisode, que les retours d'expérience, que l'on assimile fréquemment à la seule prise en compte d'informations nouvelles, peuvent aussi conduire à une révision des décisions passées sans qu'une information réellement nouvelle n'ait été produite. Dans ce cas précis, l'insuffisante prise en compte des incertitudes liés aux aléas naturels, la faiblesse de la RFS I.2.e et la possibilité que plusieurs tranches d'un site soit affectés simultanément étaient connues. On ne peut donc pas parler, en la matière, de prise de conscience, mais éventuellement de changement d'équilibre dans un jeu à plusieurs joueurs, ou de changement des attentes sociales amenant les

décideurs publics à produire une nouvelle réponse. Ceci expliquerait les limites de l'exercice. Mais nous sortons là du domaine de la théorie de la décision pour entrer dans celles de la théorie des jeux ou de la rationalité limitée.

Quatrième partie

Une approche causale de la décision en incertitude

Chapitre 7

Introduction aux réseaux décisionnels : un cas d'école

*Some metaphysical foundations – some assumptions about
the application of probability to economics – are indeed unavoidable.*

(Hicks, 1931, p. 171)

7.1 Introduction

La gestion du risque social est souvent présentée comme une question relevant de la théorie de la décision. Les modèles traditionnels de décision dans le risque (von Neumann et Morgenstern) ou dans l'incertitude (Savage) sont pourtant peu adaptés à des problèmes de décision publique impliquant la survenue d'événements rares. Certes, les évolutions de la théorie de la décision au cours des trois dernières décennies ont permis de réels progrès dans la représentation de ces problèmes et ouvert la voie à des applications plus pertinentes¹. Mais la démarche théorique continue de laisser dans l'ombre un élément essentiel de la décision publique face au risque, à savoir l'évaluation de celui-ci.

Comme nous l'avons vu tout au long des chapitres précédents, c'est dans la phase d'évaluation du risque que sont identifiés les événements redoutés, leurs causes, leurs effets, leurs probabilités de réalisation ainsi que les choix d'action existants. C'est également à ce stade que se pose la question du traitement de l'incertitude non probabilisable. En se plaçant en aval de l'évaluation du risque, la représentation traditionnelle de la théorie de la décision ne peut donc pas aborder ces questions

1. Pour des applications de la nouvelle théorie de la décision à la gestion des risques dans le secteur nucléaire en France, on pourra notamment se reporter aux thèses de Yann Serquin (Serquin, 1998) et de Sébastien Destercke (Destercke, 2008).

de façon totalement satisfaisante.

Nous avons montré au chapitre 5 que l'évaluation d'un risque social était équivalente, au plan formel, à l'estimation d'un réseau bayésien causal (RBC). En particulier, l'évaluation du risque consiste à établir par voie empirique des relations fonctionnelles décrivant des mécanismes causaux. Les perturbations exogènes y traduisent l'action agrégée de causes *inobservables*, et leurs distributions de probabilités sont l'une des données primordiales d'une évaluation de risque, comme dans le cas d'un modèle causal (Pearl, 2000). On peut donc espérer qu'un modèle de décision fondé sur les RBC, tel que celui que nous exposons dans ce chapitre et le suivant, apportera un éclairage particulier à l'articulation entre évaluation et gestion des risques sociaux.

Cependant, nous avons également relevé que, dans des cas fréquents, le gestionnaire de risque ne disposait pas d'un modèle causal pré-établi et validé pour fonder ses décisions. Il nous semble donc nécessaire, au moment d'examiner la question de la décision, d'élargir le propos par rapport aux applications traditionnelles des RBC.

Le dénuement du décideur, dans une situation où l'évaluation « objective » du risque est absente ou incomplète, ne lui interdit pas de faire sa propre évaluation du risque, c'est-à-dire de se représenter la situation dans laquelle il est appelé à intervenir sous la forme d'un modèle causal et de former des croyances relatives à l'action de causes *non identifiées* dans cette situation. Nous adoptons ce point de vue subjectiviste², qui nous semble le plus adapté à notre objet, en soulignant qu'il élargit dans deux directions le point de vue traditionnel où un modèle causal complet est présenté au décideur *a priori*.

En premier lieu, nous permettons au décideur de considérer comme parent de X_i toute variable à laquelle il prête une influence causale directe sur X_i , qu'il soit en mesure d'en observer la valeur avant de faire son choix ou pas. Nous ferons appel à la version dynamique de notre modèle, où l'incertitude se résout séparément pour chaque variable, dans le premier cas, et à la version statique, où toute l'incertitude se lève de façon simultanée, dans le second. Il faut noter que ce problème est distinct de celui de la temporalité des causes et des effets. Les théories de la causalité considèrent généralement qu'une cause doit précéder ses effets sur l'axe temporel (Reichenbach, 1956; Good, 1961; Suppes, 1970), et nous les suivons sur cette voie. Mais l'incertitude d'un décideur au sujet d'une grandeur donnée subsiste par définition tant qu'il ne l'a pas observée, qu'elle soit déterminée ou non au sens physique du terme.

En second lieu, nous nous plaçons dans la tradition de la théorie subjectiviste de la décision, où le décideur peut se déterminer en fonction de ses croyances, qu'elles soient ou non fondées sur des probabilités pré-établies. Reflétant des croyances subjectives, les distributions de probabilités

2. Il s'agit bien sûr ici d'un parti-pris théorique, qui ne contredit en rien notre appel, à la fin du chapitre précédent, en faveur d'une objectivité scientifique des décisions publiques. Rappelons que la théorie bayésienne, loin d'autoriser le manque d'objectivité, impose au décideur d'intégrer de façon cohérente toute information nouvelle.

des perturbations Y_i ne peuvent donc pas être observées directement. Seul l'ensemble des valeurs possibles de chaque perturbation fait partie des données du modèle, et il est essentiel que cet ensemble soit suffisamment vaste a priori pour contenir toutes les croyances du décideur³.

Du fait de cet élargissement, nous pensons que le champ d'application du modèle proposé dépasse celui des risques sociaux. C'est un modèle de décision d'une grande généralité, dont les données primitives sont similaires à celles de Savage, augmentées d'un ensemble de relations causales.

Ce chapitre explore les principaux aspects d'un tel modèle à partir d'un problème simplifié de gestion du risque. Nous commençons, dans la première partie, par décrire le problème et construire un RBC selon la méthode préconisée au chapitre précédent. Nous étudions dans la deuxième partie les adaptations qu'il faut apporter au cadre de décision dans l'incertitude de Savage pour tirer parti des propriétés du RBC. La troisième et dernière section est consacrée à une discussion de l'apport de ce modèle à la compréhension de quelques problèmes classiques de la théorie de la décision dans l'incertitude.

Les visées du chapitre sont donc essentiellement illustratives, ce qui nous amènera à nous autoriser quelques approximations en matière de notation, de terminologie et même de logique. Nous laisserons le soin au chapitre suivant, consacré au développement formel du modèle, de corriger ces écarts, et de montrer qu'ils ne portaient pas à conséquence.

7.2 Un problème de gestion du risque et son traitement causal

Considérons le problème de décision suivant :

Au gré de son activité, une usine rejette dans l'environnement une substance cancérigène, qui contamine une région agricole où est située une grande ville. Le décideur social en charge de la gestion du risque sanitaire associé sait :

- que la consommation de produits contaminés au-delà d'un certain seuil peut entraîner le développement de cancers ;
- que l'activité engendre également certains bénéfices du point de vue de la société ;
- que la production agricole locale, si elle n'est pas consommée sur place, est soit exportée, soit perdue ;
- enfin que les conséquences sanitaires de la maladie sont d'autant plus graves que sa détection est tardive.

En revanche, le décideur n'a pas une connaissance déterministe des variables en jeu, c'est-à-dire qu'il ignore certains déterminants de chaque variable. À titre d'exemple, il ne sait pas ce qui influence le niveau d'activité (et donc de rejet) de l'usine ou, à niveaux de contamination et d'exposition donnés, le développement de la maladie.

3. Voir à ce propos la section 7.3.1.3

Il dispose de quatre leviers d'action : imposer à l'usine une limite réglementaire de rejets ; mener une campagne d'information visant à limiter la consommation ; favoriser l'exportation de la production agricole ; et mettre en place des outils de dépistage de la maladie.

7.2.1 Une représentation causale subjective

Nous cherchons, dans cette première partie, à décrire une méthode que le décideur peut mettre en œuvre pour organiser l'information dont il dispose avant d'effectuer ses choix. Cette méthode a les mêmes objectifs que l'évaluation du risque : estimer, aussi précisément que possible, les valeurs que pourraient prendre les principales grandeurs d'intérêt pour le décideur, ainsi que les probabilités de réalisation de ces valeurs, sous différents scénarios d'action.

7.2.1.1 Structure causale

Pour traduire ses problèmes par un RBC, notre décideur doit commencer par établir une structure causale, c'est-à-dire un système de variables liées par des relations de cause à effet. Ceci correspond à la phase d'identification du risque. En reliant ainsi les grandeurs considérées, il sera à même de discerner systématiquement les sources d'incertitude de son problème et de les représenter sous forme de facteurs exogènes ; il pourra ensuite s'interroger sur ses croyances relatives aux aspects incertains du problème ; enfin, il pourra rendre compte de la façon dont ses interventions modifieraient le cours des événements.

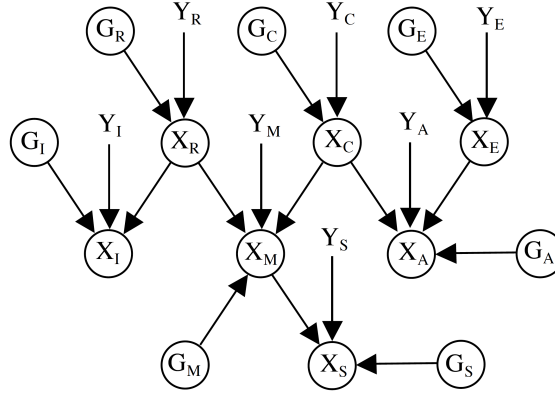
S'il suit les indications que nous avons rassemblées au chapitre précédent, le décideur doit en principe parvenir à une structure causale D proche du graphe de la figure 7.2.1, abstraction faite, pour le moment, des variables exogènes et des variables de décision, respectivement désignées par les lettres Y et G .

Le décideur a ici choisi de décrire son problème par le biais d'un système de variables $(X_R, X_I, X_C, X_E, X_A, X_M, X_S)$: X_R y désigne le niveau de rejets de l'usine, X_I l'activité industrielle, X_C la consommation locale des produits potentiellement contaminés, X_E la demande de ces produits à l'exportation, X_A la production agricole, X_M la prévalence de la maladie et X_S ses conséquences sanitaires (mesurées par exemple par un nombre de décès). Ces variables dépendent de liens causaux figurés par les flèches du graphe.

7.2.1.2 Modèle causal

L'étape suivante correspond à l'évaluation proprement dite du risque. Il s'agit, en premier lieu, de spécifier le modèle associé à la structure causale, c'est-à-dire d'estimer la forme fonctionnelle des relations causales, que nous appellerons mécanisme. Il faut alors compléter les déterminants de chaque variable par un terme additionnel résumant l'influence du monde extérieur sur elle. Ces

FIGURE 7.2.1 – Graphe causal du problème de décision



variables Y_i sont les perturbations exogènes dont le modèle tient sa nature probabiliste, ici exprimée par les croyances du décideur. Par exemple :

$$\begin{aligned}
 x_R &= f_R(y_R) = y_R \\
 x_I &= f_I(x_R, y_I) = x_R \cdot y_I \\
 x_C &= f_C(y_C) = \delta \cdot y_C \\
 x_E &= f_E(y_E) = \varepsilon \cdot y_E \\
 x_A &= f_A(x_C, x_E, y_A) = (x_C + x_E \cdot 1_{X_C > \alpha}) \cdot y_A \\
 x_M &= f_M(x_R, x_C, y_M) = \mu \cdot x_C \cdot 1_{X_R > \rho} \cdot y_M \\
 x_S &= f_S(x_M, y_S) = (2 - \sigma) \cdot x_M \cdot x_S
 \end{aligned}$$

où ρ désigne le seuil de toxicité des rejets, δ la part de la population concernée ayant le comportement alimentaire en question, ε les débouchés à l'exportation, α la production agricole concernée, μ la proportion de personnes exposées développant à terme un cancer, et σ la proportion de cancers décelés à un stade précoce. Par souci de simplicité, nous supposons ici que chaque variable Y_i ne peut prendre que deux valeurs, notées y_i et y'_i . Dans le cas de Y_R , nous imposerons en outre que $y_R = \rho < y'_R$.

7.2.1.3 Interventions

Il s'agit ensuite, pour le décideur, de tenir compte de ses possibilités d'intervention sur le problème, qui consistent en un ensemble de mécanismes qu'il peut substituer aux mécanismes en place. Les variables de décision de la figure 7.2.1 prennent ainsi pour valeur le mécanisme choisi par la décideur. G_R , G_C , G_E et G_S correspondent ici respectivement aux mesures de limitation des re-

jets, de prévention de l'exposition, de développement des exportations et de dépistage. G_I , G_A et G_M sont des variables dégénérées, l'ensemble des mécanismes que le décideur peut instaurer en ces nœuds se résumant à un singleton. À titre d'exemple, le décideur peut considérer qu'il dispose des options suivantes :

$$\begin{aligned}
x_R &= g_R(y_R) & g_R &\in \{f_R, f'_R\} \\
x_I &= g_I(y_I) & g_I &\in \{f_I\} \\
x_C &= g_C(y_C) & g_C &\in \{f_C, f'_C\} \\
x_E &= g_E(y_E) & g_E &\in \{f_E, f'_E\} \\
x_A &= g_A(y_A) & g_A &\in \{f_A\} \\
x_M &= g_M(y_M) & g_M &\in \{f_M\} \\
x_S &= g_S(x_M, y_S) & g_S &\in \{f_S, f'_S\}
\end{aligned}$$

où f'_R , f'_C , f'_E et f'_S désignent les interventions sur le niveau de rejets de l'usine, la consommation, les exportations et le dépistage du cancer :

$$\begin{aligned}
f'_R(y_R) &= \max(y_R, \rho) \\
f'_C(y_C) &= \delta' \cdot y_C \\
f'_E(y_E) &= \varepsilon' \cdot y_E \\
f'_S(x_M, y_S) &= s_0 + (2 - \sigma') \cdot x_M \cdot y_S
\end{aligned}$$

où $\delta' < \delta$, $\varepsilon' > \varepsilon$, $\sigma' > \sigma$ et où s_0 est un coût d'opportunité de la campagne de dépistage ($s_0 > 0$). Le décideur a ici fait le choix de négliger les coûts directs des trois premières mesures, c'est-à-dire les coûts qui ne résultent pas de la modification du niveau de rejets, de la consommation ou des exportations, mais de la mise en place des mesures elles-mêmes.

Pour que les variables de décision G_i puissent être intégrées au RBC comme dans la figure 7.2.1, il faut que les mécanismes décrits par les relations causales soient autonomes *aux yeux du décideur*, c'est-à-dire qu'il pense pouvoir modifier chacun d'entre eux sans que cela affecte les autres.

7.2.1.4 Réseau décisionnel

Il reste au décideur à construire la loi jointe P_g des variables du système sous chacune des combinaisons de mécanismes $g = (g_R, g_I, g_C, g_E, g_A, g_M, g_S)$ qu'il puisse envisager. Pour cela, il doit synthétiser sa connaissance des influences du monde extérieur sur le système sous la forme d'une distribution de probabilités P_Y sur les variables résiduelles $(Y_R, Y_I, Y_C, Y_E, Y_A, Y_M, Y_S)$. Il

en déduit facilement la distribution conditionnelle de chaque variable X_i par rapport à ses parents sous la combinaison de mécanismes considérée. À titre d'exemple :

$$\begin{aligned} P_g(X_R) &= P_Y(g_R(Y_R)) \\ P_g(X_I | X_R) &= P_Y(g_I(g_R(Y_R), Y_I) | Y_R) \end{aligned}$$

Si le décideur pense que les variables Y_i sont mutuellement indépendantes⁴, alors il sait qu'il peut considérer chaque variable X_i indépendante des variables autres que ses descendants dans D , conditionnellement à ses parents. De ce fait⁵, il peut facilement dériver la loi jointe P_g à partir des distributions conditionnelles des X_i selon la factorisation suivante :

$$\begin{aligned} P_g(X_R, X_I, X_C, X_E, X_A, X_M, X_S) &= P_g(X_R).P_g(X_I | X_R).P_g(X_C). \\ &P_g(X_E).P_g(X_A | X_C, X_E).P_g(X_M | X_R, X_C).P_g(X_S | X_M) \end{aligned}$$

Appelons P^* l'ensemble des distributions de probabilités P_g obtenues lorsque g parcourt l'ensemble $\mathcal{G}$ des vecteurs de mécanismes possibles. Comme la factorisation précédente est vraie pour toutes les distributions P_g , le décideur peut qualifier D de réseau bayésien causal⁶ compatible avec P^* . Dans le chapitre suivant, nous définirons formellement le trio formé par la structure D , les perturbations Y et l'ensemble $\mathcal{G}$ comme un réseau décisionnel.

Notons, pour clore cette section, que l'ordre de prise en compte des variables a été fixé par le décideur. Ce choix est en partie déterminé par la structure causale, mais en partie seulement. $(X_C, X_E, X_R, X_M, X_S, X_A, X_I)$, par exemple, est compatible avec la structure et conduit à la même distribution de probabilités.

7.2.2 La constitution d'un problème de décision causal

Nous cherchons à présent à retranscrire le problème causal énoncé ci-dessus dans le formalisme du modèle de Savage (1954), qui constitue l'un des cadres de référence de la théorie de la décision sous incertitude. Notre objectif est de décrire, sous la forme d'observations établies à partir du cas illustratif, les traits saillants du modèle causal de décision que nous exposerons au chapitre suivant.

Notre décideur doit donc résoudre un septuple (en réalité quadruple) problème de décision

4. Comme nous l'avons vu dans la sous-section 5.2.2.2 page 172, le décideur sait qu'il peut tenir les variables Y_i pour mutuellement indépendantes dès lors qu'il exclut l'existence d'une cause commune latente à plusieurs variables X_i et qu'il souscrit en outre à l'hypothèse dite de cause commune de Reichenbach.

5. Voir le corollaire 5.5 page 167.

6. Voir la définition 5.9 page 171 dans un contexte où nous considérons une classe plus restreinte d'interventions.

que l'on peut envisager de deux manières. Dans une séquence de choix dynamiques, le décideur connaît la valeur des parents de la variable au moment de déterminer son action. Autrement dit, les causes *identifiées* par le décideur sont également *observées* par lui préalablement à son choix. Ceci peut constituer une hypothèse forte. Par exemple, il est vraisemblable que notre décideur devra se déterminer au sujet de la campagne de dépistage avant de connaître l'ampleur de l'épidémie. Dans les cas où l'incertitude se lève de façon simultanée sur toutes les variables, le choix est statique, et il équivaut à celui d'une stratégie *a priori* ou d'un plan de contingence. Comme nous le verrons au chapitre suivant, le choix statique s'obtient comme un cas particulier du choix dynamique. Aussi, dans l'essentiel de cette section, nous nous placerons dans un contexte dynamique.

Nous commençons par introduire notre version du cadre de Savage pour un problème de décision isolé, où les relations causales ne tiennent qu'une place mineure, avant d'examiner un cas où leur rôle apparaît clairement. Nous finirons cette partie en évoquant la relation représentative des choix du décideur, qui fait partie des données primitives du modèle de Savage.

7.2.2.1 Un problème de décision isolé

Pour Savage, un problème de décision est représenté formellement à partir de trois notions primordiales :

- les *conséquences*, c'est-à-dire «tout ce qui peut arriver à la personne» à l'issue de sa décision ;
- les *états du monde*, c'est-à-dire «une description du monde qui ne laisse de côté aucun aspect pertinent» et dont la détermination «résoud toute incertitude» ;
- les *actes*, enfin, qui sont toutes les applications de l'ensemble des états du monde à celui des conséquences. ; chaque acte associe ainsi une conséquence à chaque état du monde.

Pour voir ce que recouvrent ces catégories dans notre cas, commençons par considérer une situation où toutes les décisions déterminantes pour le développement de la maladie auraient été prises, l'incertitude se serait levée et une certaine valeur $X_M = x_M$ serait réalisée et observée. Le problème du décideur se résumerait au choix d'engagement d'une campagne de dépistage. Les composantes savagiennes de ce problème de décision sont résumées par le tableau 7.1.

Les cellules du tableau représentent les conséquences, qui doivent être des descriptions complètes et dénuées d'incertitude de tous les enjeux de la décision pour le décideur, c'est-à-dire, ici, la valeur de X_S . Les quatre valeurs reportées dans le tableau représentent l'ensemble des conséquences que le décideur peut estimer compatibles avec la valeur x_M et ses options f_S et f'_S .

Les colonnes du tableau désignent les états du monde, ou plus précisément le regroupement de ces états en événements suffisamment détaillés pour rendre compte de l'incertitude. Les événements doivent être exhaustifs et mutuellement exclusifs, c'est-à-dire qu'ils doivent former une partition de l'ensemble des états du monde. Il s'agit, selon la description d'Arrow (1971), d'« une description

Table 7.1: Un problème de décision en S

Conditionnellement à $(X_M = x_M)$	$Y_S = y_S$	$Y_S = y'_S$
f_S : laisser -faire	$x_S = (2 - \sigma) \cdot x_M \cdot y_S$	$x_S = (2 - \sigma) \cdot x_M \cdot y'_S$
f'_S : dépistage	$x_S = s_0 + (2 - \sigma') \cdot x_M \cdot y_S$	$x_S = s_0 + (2 - \sigma') \cdot x_M \cdot y'_S$

du monde si complète que si elle était vraie et connue, alors les conséquences de chaque action serait connues » ; mais, devrait-on ajouter, pas davantage. Dans l'axiomatique de Savage, les états du monde doivent en effet comporter un nombre infini d'aspects, de sorte qu'aucun état ne soit entièrement descriptible et que tout événement, aussi spécifique soit-il, reste subdivisible⁷. Il en résulte une certaine confusion dans l'usage des concepts d'état du monde et d'événement dans la littérature. Nous essaierons d'y échapper en n'utilisant le premier terme que dans son sens rigoureusement savagien, et désignant par le second les événements associés aux conséquences plutôt que des sous-ensembles quelconques de l'ensemble des états, sauf mention explicite. Les événements qui apparaissent dans les colonnes du tableau 7.1 sont fondés sur les deux réalisations possibles de la variable exogène Y_S qui, lorsque la valeur des parents de X_S est fixée, représente la seule source d'incertitude dans la détermination des conséquences. Ces événements répondent aux conditions de l'axiomatique savagienne si l'on suppose que Y_S mesure un aspect particulier des états du monde, et que l'ensemble Ω de ces états possède la structure voulue :

$$Y_S : \Omega \rightarrow \mathcal{Y}_S$$

Les lignes du tableau, enfin, représentent des actes. Ce sont les objets du choix du décideur et des préférences que ce choix exprime. Les options relatives à X_S dont le décideur dispose effectivement sont f_S et f'_S . Cependant, la notion d'acte de Savage ne recouvre pas seulement les actions que le décideur peut réellement entreprendre, mais toutes les applications de l'ensemble des événements vers celui des conséquences, y compris celles qui mettent en rapport un événement avec une conséquence qui ne peut pas lui être associée dans la réalité⁸ (tels que $Y_S = y'_S$ et $X_S = (2 - \sigma) \cdot x_M \cdot y_S$). Par suite, le nombre total d'actes est de $4^2 = 16$.

Nous appellerons l'ensemble d'actes ainsi définis *le problème de décision en S* .

Observation 1 : Dans un problème de décision isolé, l'ensemble des états du monde est sous-jacent

7. Il s'agit spécifiquement de l'axiome P6.

8. Dans notre représentation, ces associations fictives sont celles qui contiennent aux mécanismes causaux.

à la variable exogène Y_i ; l'ensemble des conséquences est celui des valeurs de la variable X_i qui sont compatibles avec la valeur des parents de X_i et les actes réels à la disposition du décideur ; les actes sont toutes les fonctions du premier de ces ensembles vers le second.

Considérons à présent ce problème *a priori*, dans le cadre complet du RBC de la première partie. Le décideur n'a encore pris aucune décision et aucune incertitude n'a été levée. Mais il sait que lorsqu'il sera effectivement confronté à son choix, il aura déjà résolu ses problèmes en R , C et M et qu'il engagera son acte conditionnellement à une certaine valeur de ces variables. Il sait donc que, *conditionnellement à la valeur* x_M , son choix déterminera les conséquences sanitaires x_S de la maladie associé à une certaine réalisation de la variable résiduelle Y_S . Les actes qu'il peut effectivement envisager sont dérivées des interventions f_S et f'_S définies plus haut :

$$\begin{aligned} f_S(x_M, y_S) &= (2 - \sigma) \cdot x_M \cdot y_S \\ f'_S(x_M, y_S) &= s_0 + (2 - \sigma') \cdot x_M \cdot y_S \end{aligned}$$

Ces actes, dont ceux du tableau 7.1 font partie, sont des fonctions de Y_S paramétrées par la valeur de X_M . En toute rigueur nous ne devrions donc pas les noter de la même façon que les mécanismes, qui sont des fonctions de Y_S et X_M . Dans la suite, nous désignerons donc les mécanismes par la lettre g , et les actes par la lettre γ .

A priori, la variable X_M peut prendre dix valeurs différentes, qui définissent dix problèmes de décision distincts tels que celui que nous venons de décrire.

Considérons à présent le même problème de décision après la détermination de la valeur de X_R , par exemple $x_R = y_R = \rho$. La valeur de X_M n'est toujours pas déterminée, mais l'ensemble des valeurs compatibles avec x_R ne compte plus que deux éléments. Le nombre de problèmes de décision envisageables en X_S est, lui aussi, de deux.

Au moment du choix d'un acte, finalement, une valeur particulière de X_M s'est réalisée à la suite de la résolution de l'incertitude en amont. Le problème de décision est défini de façon univoque.

Le cas suivant nous permettra de détailler le processus de réduction qui a ainsi fini par engendrer les ensembles de conséquences, d'états du monde et d'actes constituant le problème de la décision en S .

Observation 2 : L'ensemble des valeurs possibles de X_i est réduit à chaque fois que l'un des ascendants de X_i est déterminé.

7.2.2.2 Un cas plus complexe

Considérons à présent le cas de la mesure de limitation des rejets, et supposons qu'il s'agisse du premier choix considéré par le décideur (de sorte que toutes les variables seraient indéterminées au

moment effectif de ce choix).

La démarche précédente peut être reconduite pour définir les événements à partir des réalisations de Y_R , considérer les conséquences des actes effectivement envisageables et, pour finir, compléter l'ensemble des actes. La différence essentielle avec le cas précédent concerne les conséquences, qui ne portent plus ici sur les seules valeurs de la variable X_R mais concernent également les variables, et plus précisément les problèmes de décision, en aval.

Supposons, en effet, que le décideur choisisse le laisser-faire (l'acte correspondant au choix du mécanisme f_R , et que nous n'avons pas encore caractérisé), que l'incertitude relative à Y_R se résolve et que X_R prenne la valeur $x_R > \rho$. Appelons cette réalisation une conséquence *directe* de l'acte.

En M , cinq problèmes de décision peuvent *a priori* être envisagés par le décideur, en fonction des valeurs des parents (X_R, X_C) . Il existe un seul mécanisme, f_M , auquel correspondent cinq actes en M . Chacun de ces actes associe deux valeurs de X_M aux deux événements $Y_M = y_M$ et $Y_M = y'_M$. Après la réalisation de x_R , l'un de ces problèmes est écarté, et deux des dix valeurs *a priori* possibles de X_M sont devenues incompatibles avec la trajectoire. Le problème de décision en M , que nous n'avons pas encore défini dans son intégralité, est donc dépendant de la valeur de X_R , et l'influence qu'il a subie peut être ainsi résumée : la réalisation de la valeur $(x_R > \rho)$ signifie que le seuil de toxicité est franchi, et qu'il y aura en tout état de cause une épidémie.

Le problème de décision en I , qui concerne également un enfant de X_R , subit une transformation similaire : la réalisation de la valeur $(x_R > \rho)$ signifie que l'activité industrielle sera de toute façon élevée.

Le resserrement de l'ensemble des valeurs possibles de X_M se répercute en outre sur le problème de décision en S , et constitue la première étape du processus de détermination de ce problème dont nous avons décrit l'issue : les conséquences sanitaires de la maladie seront, quoiqu'il arrive, ceux d'une épidémie.

Nous constatons donc, pour tous les descendants de X_R , une réduction de l'ensemble des problèmes de décision envisageables selon les termes de l'observation 2, c'est-à-dire par resserrement de l'ensemble des valeurs possibles de la variable (que nous avons appelé les conséquences directes).

Observation 2bis : En déterminant la valeur de X_i , un acte en i restreint les ensembles de valeurs possibles de ses descendants et, par là, les ensembles de problèmes de décision correspondants.

Nous parlerons alors de conséquences *indirectes*.

Mais alors, la réalisation de x_R modifie de surcroît les problèmes de décision en C , A et E .

En C , le problème de décision comporte des conséquences indirectes pour le problème de décision en M qui, nous l'avons vu, a commencé à être déterminé par la réalisation de x_R . La transformation du problème de décision en C reflète le fait suivant : *parce que* les situations simplement endémiques sont désormais exclues, tout niveau de consommation causera des cas de cancer, de sorte que des

mesures de prévention permettant de sauvegarder à la fois la santé publique et l'activité agricole ne seront plus envisageables.

Observation 3 : L'acte en i entraîne également une réduction des ensembles de problèmes de décision relatifs aux ascendants des descendants de X_i , cette fois par le biais d'un resserrement de leurs conséquences indirectes.

La réduction de l'ensemble des problèmes de décision en C change les conditions de détermination de X_C . Le décideur sachant que certaines versions du problème de décision en C ne se présenteront pas, il peut être en mesure d'en tirer déjà des conséquences relatives à l'ensemble des valeurs de X_C et celui des problèmes de décision envisageables en A . Cette répercussion, contrairement aux précédentes, passe par les actes que le décideur pense engager en C , et fait donc intervenir ses préférences : *parce qu'il sait qu'il doit choisir entre la sauvegarde de la santé publique et celle la production agricole, s'il préfère la première*, alors le décideur sait que la seconde correspondra, quoiqu'il arrive par ailleurs, à un faible niveau de consommation.

Observation 4 : En modifiant l'ensemble des arbitrages entre conséquences directes et indirectes que le décideur opère par ses actes, le changement du problème de décision en i affecte les problèmes de décision à tous les nœuds descendants de X_i .

Finalement, la modification du problème de décision en A se répercute sur celui en E selon la logique de l'observation 3 : *parce que* l'activité agricole sera pénalisée par le niveau de consommation locale, le développement des exportations aura, en tout état de cause, un impact important.

Au total, l'ensemble des problèmes de décision en aval de R est concerné par la détermination de cette variable.

Observation 5 : La réalisation d'une valeur de X_i change les problèmes de décision pour tous les descendants de X_i , tous les ascendants de ceux-ci, tous les descendants de ces derniers, et ainsi de suite.

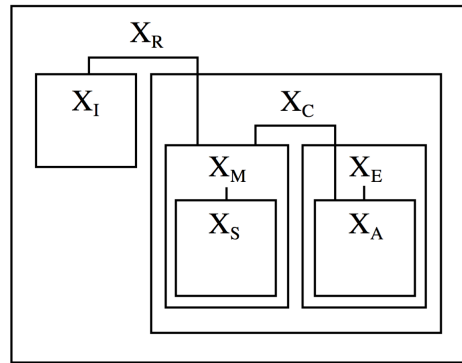
L'observation précédente en entraîne immédiatement une autre.

Observation 6 : Le problème de décision en i est dépendant de la valeur de toutes les variables dont X_i est le descendant, l'ascendant de l'un des descendants, etc.

Dans le chapitre suivant, nous définirons formellement l'ensemble des variables descendantes de X_i , ascendantes de l'un de ses descendants, etc., et nous donnerons à l'ensemble de problèmes correspondant le nom d'*ensemble des problèmes contingents à X_i* . En outre, nous définirons l'ensemble des variables dont le problème en i est contingent comme le *contexte en i* .

On peut distinguer deux sous-ensembles de problèmes de décision contingents à X_R : d'une part le problème en I , qui n'est contingent à aucun autre et dont aucun autre n'est contingent ; de l'autre, les problèmes en C , E , A , M et S , qui ont des relations de contingence emboîtées : les

FIGURE 7.2.2 – Relations parent-enfant et contingence des problèmes de décision



quatre derniers sont contingents à X_C , celui en S est contingent à X_M , et celui en A à X_E . On observe que la détermination d'une variable définit un tel sous-ensemble pour chacun de ses enfants (voir la figure 7.2.2, où les sous-ensembles sont représentés par des rectangles). Ceci est expliqué par le fait que les sous-ensembles de variables communiquant avec chacun des enfants d'une variable X_i sont d-séparés par cette variable (dans la mesure où ses enfants n'ont pas, entre eux, de lien parent-enfant), c'est-à-dire que tout élément de l'un est indépendant de tout élément d'un autre conditionnellement à X_i . Par exemple, une fois la valeur des rejets connue, l'activité industrielle n'est influencée par aucune des autres variables, et n'en influence aucune en retour. De même, le problème de décision relatif à la campagne de dépistage du cancer dépend de la valeur des rejets, de la consommation de produits contaminés et de l'ampleur de l'épidémie ; elle est en revanche indépendante de l'activité industrielle, des exportations et de la production agricole.

Il apparaît dès lors superflu de considérer explicitement tous les problèmes de décision contingents à X_i comme conséquences indirectes des actes en X_i . En se bornant à recenser les seuls problèmes de décision qui sont contingents à X_i mais ne sont contingents à aucun de leurs congénères, on assure la représentation de l'ensemble des problèmes de décision concernés :

Le problème de décision en R a pour conséquence indirecte

le problème de décision en I ; et

le problème de décision en C qui a pour conséquence indirecte

le problème de décision en E qui a pour conséquence indirecte

le problème de décision en A ; et

le problème de décision en M qui a pour conséquence indirecte

le problème de décision en S .

Observation 7 : Un acte en i a pour conséquence directe une valeur x_i et pour conséquences indirectes un sous-ensemble de problèmes de décision conditionnés par x_i . Cet ensemble, que nous noterons dans la suite $\hat{\Gamma}(x_i)$, porte sur tous les nœuds contingents à X_i et qui ne sont

Table 7.2: Le problème de décision en R

	$Y_R = y_R$	$Y_R = y'_R$
γ_R : laisser -faire	$\begin{matrix} x_R = y_R \\ \hat{\Gamma}(y_R) = (\Gamma_{I y_R}, \Gamma_{C y_R}) \end{matrix}$	$\begin{matrix} x_R = y'_R \\ \hat{\Gamma}(y'_R) = (\Gamma_{I y'_R}, \Gamma_{C y'_R}) \end{matrix}$
γ'_R : limitation des rejets	$\begin{matrix} x_R = y_R \\ \hat{\Gamma}(y_R) = (\Gamma_{I y_R}, \Gamma_{C y_R}) \end{matrix}$	$\begin{matrix} x_R = \rho \\ \hat{\Gamma}(\rho) = (\Gamma_{I \rho}, \Gamma_{C \rho}) \end{matrix}$

contingents à aucune variable elle-même contingente à X_i .

Notons $\Gamma_{I|x_R}$ le problème de décision en I engendré par la réalisation de x_R , c'est-à-dire l'ensemble des fonctions de $\{y_I, y'_I\}$ dans $\{x_R.y_I, x_R.y'_I\}$. Notons, de même $\Gamma_{C|x_R}$ le problème de décision en C conditionnellement à $(X_R = x_R)$, soit l'ensemble des fonctions de $\{y_C, y'_C\}$ dans l'ensemble des conséquences de la forme $(x_C, \Gamma_{M|x_R, x_C}, \Gamma_{E|x_R, x_C})$, où $x_C \in \{\delta.y_C, \delta.y'_C, \delta'.y_C, \delta'.y'_C\}$ et $\Gamma_{M|x_R, x_C}$ et $\Gamma_{E|x_R, x_C}$ sont les problèmes de décision en M et E conditionnellement à $(X_R = x_R, X_C = x_C)$, définis de façon similaire. Nous pouvons à présent réunir dans le tableau 7.2 tous les éléments du problème du décideur en X_R .

7.2.2.3 Les préférences et la propriété de Markov

En continuant d'appliquer le cadre de Savage à son problème, notre décideur est amené à exprimer des préférences entre actes. Nous le supposons ainsi doté de sept relations $\succeq_i$ guidant ses choix dans chacun des problèmes de décision auxquels il se trouve confronté. Il faut noter qu'aucune de ces relations n'est triviale, puisque même lorsque les options réelles du décideur se réduisent à un singleton, l'ensemble des actes, tel que nous l'avons défini, ne le fait pas. Or, par hypothèse, le décideur doit pouvoir comparer tout paire d'actes dans cet ensemble, fussent-ils virtuels.

Nous avons mis en évidence, dans les sections précédentes, la dépendance des différentes composantes d'un problème de décision en i à la *trajectoire*, nom par lequel nous désignerons les valeurs de variables X_j déjà observées. Il est légitime de s'interroger également sur celle de la relation $\succeq_i$ appliquée à ce problème. La dépendance des préférences vis-à-vis de la trajectoire est un thème classique de la théorie du choix, qui permet de rendre compte de phénomènes tels que les effets richesse ou la formation d'habitudes.

Nous avons fait, à ce sujet, une hypothèse cruciale qu'il convient de discuter à présent.

Savage caractérisait son concept de conséquences comme des «états de la personne», par opposition aux états du monde (Savage, 1954, p.14). Dans notre exemple, l'état du décideur comporterait ainsi sept composantes, décrites par les sept variables du réseau. On pourrait imaginer que ces variables soient adossées à un espace d'états similaire à l'ensemble des états du monde Ω que nous avons

esquissé, et dont les variables X_i mesurent certains aspects, précisément ceux qui sont concernés par le problème général du décideur.

La propriété de Markov, que nous avons supposé vérifiée dans la première section, nous impose alors des exigences particulières. Pour s'y conformer, le décideur doit en effet estimer qu'une fois observée la valeur des parents de X_i , celle des autres prédécesseurs de X_i ne constitue pas une information pertinente pour prévoir la valeur de X_i . L'hypothèse d'indépendance mutuelle des variables Y_i se traduit de façon similaire : le décideur doit estimer que les réalisations de Y_i sont indépendantes de celles de tous ses prédécesseurs. Or, dans le cadre subjectiviste de Savage, les croyances du décideur nous sont inaccessibles, et les seules conditions de cette nature que nous puissions édicter portent sur ses préférences en matière d'actes, c'est-à-dire à la fois sur ses croyances et sur son appréciation des conséquences. Considérons par exemple, l'hypothèse suivante : les préférences du décideur relatives aux actes en i ne dépendent pas de la trajectoire. Cela implique que le décideur ne s'appuie pas sur la connaissance de $(x_1, \dots, x_{i-1})$ pour évaluer les réalisations de Y_i , ce qui correspond à la condition recherchée, mais aussi que la valeur qu'il accorde à une conséquence en i ne dépend pas des conséquences qu'il a déjà observées. Que penser de cette clause supplémentaire ?

Il se trouve que les interventions (et donc les actes) en i définissent précisément les conséquences directes en i comme fonction de la valeur des parents de X_i (et de la variable résiduelle Y_i). Tout ce qui importe à l'aspect de l'état du décideur mesuré par X_i est donc pris en compte dans la détermination de X_i . Il en est de même pour toute autre variable, et donc aussi pour les conséquences indirectes en i . Dès lors, il semble naturel de supposer que l'appréciation par le décideur des conséquences en i ne dépend d'aucun des prédécesseurs de X_i ⁹. Il en résulte que l'hypothèse d'indépendance des préférences vis-à-vis de la trajectoire, qui implique la version subjective de la propriété de Markov, paraît être une condition raisonnable.

Considérons un exemple et un contre-exemple dans notre cas. En identifiant X_M comme seul parent de X_S , le décideur signifie que l'importance des conséquences sanitaires du cancer, *en ce qui le concerne*, ne dépend pas de la quantité de produits cancérigènes rejetés dans la nature, de la consommation d'aliments contaminés par ces produits ou d'autres aspects du problème dès lors que l'on connaît la prévalence de la maladie. À travers ses actes, il détermine précisément une évaluation des conséquences sanitaires qui reflète cet état de fait, étant une fonction de la seule prévalence (et de l'événement réalisé). À titre d'exemple, s'il estime que les coûts sociaux de la maladie sont plus ou moins (en fonction de l'état de la nature) proportionnels à sa prévalence lorsque l'épidémie est limitée, mais qu'ils deviennent inacceptables pour une épidémie de grande ampleur, cette convexité apparaît dans l'expression de ses actes. Ses préférences, et plus précisément son appréciation des

9. Voir aussi la discussion consacrée à la dépendance aux états (section 7.3.2.1).

conséquences en S , n'auront pas à être dépendantes de x_M . Il semble donc naturel de supposer que les préférences du décideur pour des actes en S , tels que le dépistage, ne dépendent pas de l'ampleur de l'épidémie ni, *a fortiori*, des rejets ou de la consommation de produits contaminés.

Si, par ailleurs, le décideur estimait que la société doit être d'autant plus soucieuse de la santé de ses membres que son niveau de revenu est élevé, et que sa décision d'engager une mesure coûteuse de dépistage dépendait, de ce fait, des bénéfices économiques liés aux valeurs de X_I et X_A , alors à ses yeux, le coût social effectif du cancer dépendrait également de ces variables, et *le graphe causal devrait être corrigé en conséquence*.

Observation 8 : En vertu des propriétés de la structure causale, il est naturel de supposer que la relation $\succeq_i$ est indépendante de la trajectoire. Cette indépendance implique que les croyances dérivées de $\succeq_i$ vérifient la propriété de Markov.

7.2.3 La résolution du problème

7.2.3.1 La représentation en utilité espérée

Si le décideur souscrit aux postulats de Savage pour une relation $\succeq_i$, alors il se comporte dans ses choix d'actes en i comme s'il maximisait en i une espérance d'utilité des conséquences. Des observateurs extérieurs pourraient dériver de ses préférences la distribution subjective de probabilités sur les événements et une fonction d'utilité relative aux conséquences qui représenteraient ce comportement.

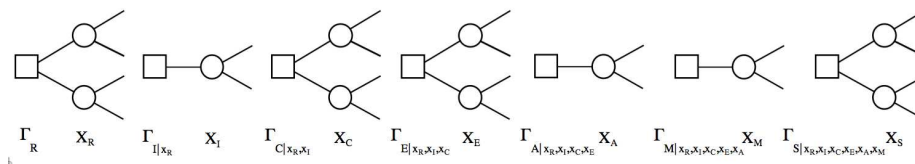
En imposant au décideur une condition supplémentaire de cohérence entre ses différentes relations de préférence, on peut conférer à cette représentation une propriété de récursivité, c'est-à-dire établir une correspondance entre les valeurs accordées par le décideur au problème de décision en j , d'une part comme conséquence indirecte des actes en i et de l'autre comme ensemble d'actes en j . Le décideur maximise alors en i l'espérance d'utilité attachée à la conséquence directe x_i et à l'espérance d'utilité des meilleurs actes à sa disposition en j conditionnellement à la valeur x_i . L'utilité en i apparaît comme une fonction d'agrégation de l'utilité de la conséquence directe x_i et de l'utilité espérée d'un acte ultérieur.

Les propriétés de cette fonction traduisent les préférences du décideur en matière de dynamique de résolution de l'incertitude, et en particulier son attitude face à l'ambiguïté.

En théorie bayésienne de la décision, ce modèle formel est généralement appliqué aux séquences temporelles de décision, où un acte en $X_i = X_t$ conduit à une conséquence directe x_t et à un problème de décision en $X_j = X_{t+1}$. Du fait de la récursivité, le décideur a en perspective, à travers X_{t+1} , l'ensemble des décisions qu'il lui reste à prendre jusqu'à une conséquence terminale x_T .

L'un des outils les plus communs de cette théorie pour illustrer les problèmes de décision séquentielle d'un tel décideur est l'arbre de décision, dont une forme synthétique est présentée dans

FIGURE 7.2.3 – Arbre de décision traditionnel (forme synthétique)



la figure 7.2.3. Nous y reportons les problèmes de décision de notre cas, dans l'ordre de traitement (R, I, C, E, A, M, S) , en oubliant momentanément la construction causale à laquelle nous venons de procéder.

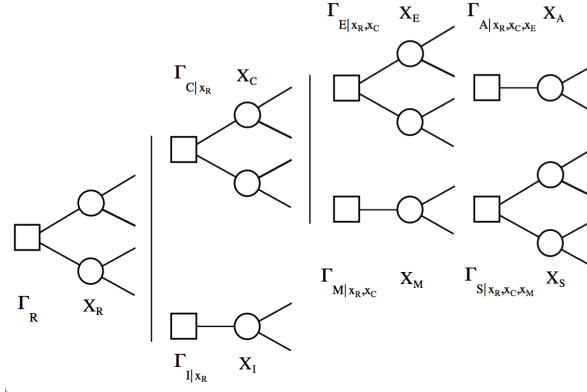
L'arbre de décision représente un jeu entre le décideur et la Nature. Les carrés du diagramme sont les problèmes de décision, dont proviennent des branches qui sont les choix (actes) du décideur. Ces dernières mènent à des nœuds de hasard, figurés par des cercles, où les conséquences de l'acte sont déterminées par le choix de la Nature (événement). Les branches sortantes de ces nœuds portent les probabilités subjectives associées aux événements et conduisent aux conséquences, c'est-à-dire ici les couples (conséquence directe, problème de décision suivant). Dans l'arborescence complète, chaque conséquence est reliée au nœud de décision du problème suivant, de sorte que l'arbre extensif correspondant à celui de la figure 7.2.3 comporte $4 \times 2 \times 4 \times 4 \times 2 \times 2 \times 4 = 2048$ terminaisons.

L'arbre décrit le comportement d'optimisation à rebours du décideur. Les terminaisons correspondent aux différentes valeurs x_S , et indiquent l'utilité associée par le décideur à ces valeurs ; en pondérant ces utilités par la probabilité de la branche qui y conduit pour toutes les branches du nœud de hasard X_S , le décideur calcule l'utilité espérée de l'acte $\gamma_{S|x_R,x_I,x_C,x_E,x_A,x_M}$ menant à ce nœud ; l'acte qui maximise cette valeur est celui que le décideur choisit s'il parvient à ce nœud de décision, et détermine l'utilité qu'il attache au problème $\Gamma_{S|x_R,x_I,x_C,x_E,x_A,x_M}$. Les branches du nœud de hasard précédent correspondent alors aux valeurs x_M et à l'utilité du problème $\Gamma_{S|x_R,x_I,x_C,x_E,x_A,x_M}$ associé, et indiquent l'utilité de ce couple. En poursuivant ce calcul jusqu'au nœud initial, le décideur connaît l'espérance d'utilité en R de toutes les séquences possibles de choix de sa part et de celle de la Nature, et peut choisir l'acte pour lequel cette espérance est la plus élevée.

7.2.3.2 La simplification causale

En tenant maintenant compte de notre traitement causal, nous pouvons considérablement simplifier ce schéma. Nous avons vu, en effet, que chaque acte a pour conséquences indirectes un sous-ensemble de problèmes de décision ultérieurs, et que chaque problème de décision ne dépend que de la valeur des variables auxquelles il est contingent. Une décision en i conduit ainsi à autant de séquences de décision indépendantes que X_i a d'enfants (non connectés entre eux).

FIGURE 7.2.4 – Arbre de décision causal (forme synthétique)



Dans notre cas, le choix du décideur en R doit maximiser l'espérance d'utilité qu'il attache à la conséquence directe x_R et à l'espérance d'utilité des meilleurs actes à sa disposition en I et C conditionnellement à la valeur x_R . À l'étape suivante, les deux problèmes d'optimisation dynamique qui découlent de la valeur de X_R , $\Gamma_{I|x_R}$ et $\Gamma_{C|x_R}$, sont *indépendants l'un de l'autre*.

Le jeu causal entre le décideur et la Nature peut être représenté par l'arbre de la figure 7.2.4, où les branches provenant du nœud de hasard X_R et menant à ces deux problèmes portent l'utilité des conséquences directes et indirectes de l'acte en R , c'est-à-dire du couple $(x_R, \hat{\Gamma}(x_R) = (\Gamma_{I|x_R}, \Gamma_{C|x_R}))$. Ces branches se prolongent *d'une part* par le nœud du problème $\Gamma_{I|x_R}$ et *d'autre part* par celui de $\Gamma_{C|x_R}$. Autrement dit, l'arbre de la figure représente une synthèse de l'arbre extensif associé à la figure 7.2.3, non pas seulement sur le plan symbolique, mais surtout parce qu'il en sélectionne uniquement les sous-arbres causalement pertinents.

La résolution de l'arbre se fait par récurrence arrière, comme dans le cas traditionnel : à l'avant-dernière étape de sa maximisation, le décideur résout $\Gamma_{I|x_R}$ et $\Gamma_{C|x_R}$ pour chaque valeur de X_R ; à la dernière, il résout Γ_R .

L'ensemble des arborescences ainsi résumées comporte $4 \times 4 \times 4 \times 2 + 4 \times 4 \times 2 \times 4 + 4 \times 2 = 264$ terminaisons.

7.2.3.3 Problèmes de décision statiques

Nous avons dit plus haut que la résolution du problème dynamique mettait en exergue la différenciation de l'attitude du décideur vis-à-vis des sources d'incertitude impliquées dans le problème. Nous pouvons à présent préciser cette affirmation.

Dans la procédure d'optimisation dynamique que nous venons de décrire, les branches issues d'un nœud de décision portent l'utilité d'un couple (conséquence directe, conséquences indirectes). Cette utilité joue le rôle d'un agrégateur entre celles de ses composantes. La théorie de l'utilité espérée

réursive (Kreps et Porteus, 1978; Chew et Epstein, 1991; Skiadas, 1998) a établi que la linéarité de cet agrégateur caractérisait l'attitude du décideur vis-à-vis de la dynamique de l'incertitude.

Lorsque le décideur est indifférent au degré de contingence des décisions¹⁰, son utilité pour le couple $(x_E, \Gamma_{A|x_R, x_C, x_E})$ est une fonction linéaire de l'utilité de x_E et de l'espérance d'utilité de l'acte choisi en A . Son espérance d'utilité en E peut alors être écrite comme utilité espérée du couple (x_E, x_A) . La propriété de linéarité se transmet ainsi de nœud en nœud en remontant l'arbre, jusqu'au premier problème de décision. Le problème de décision dynamique d'un tel décideur est alors équivalent à un problème statique, où il se placerait à un nœud de décision unique au début de l'arbre, déterminerait *a priori* des séquences complètes d'interventions, puis assisterait à la détermination simultanée de toutes les variables.

Si, en revanche, le décideur fait une distinction de nature entre une décision incertaine et une autre qui, de surcroît, est contingente à la première, alors cette équivalence est perdue. Le calcul d'optimisation peut bien sûr avoir lieu, mais l'utilité espérée en un nœud ne s'écrit plus comme combinaison linéaire des utilités des conséquences en aval.

Le problème de décision statique est donc résolu par le même processus d'optimisation à rebours que celui que nous venons de décrire, en y adjoignant l'hypothèse supplémentaire de linéarité de l'agrégateur.

7.3 Analyse de quelques problèmes traditionnels

Le cadre et les postulats de Savage, qui permettent de parvenir aux résultats que nous venons d'exposer, ont alimenté la recherche et les controverses depuis plus d'un demi-siècle, si bien qu'il semble même vain de vouloir proposer une synthèse de la littérature qu'ils ont suscitée. Nous nous contenterons, dans cette partie, de rappeler quelques-uns des problèmes traditionnels que pose la représentation savagienne d'un problème de décision en situation d'incertitude, et de montrer que notre interprétation causale permet de leur apporter un éclairage nouveau.

Les problèmes que nous évoquerons sont de deux ordres. D'une part, c'est la définition d'un problème de décision à partir de deux notions fondamentales, les conséquences et les événements, et du troisième, dérivé des deux premiers, d'actes qui ne semble pas correspondre, dans de nombreux cas, à l'expérience humaine de décision en situation d'incertitude. Si l'on s'en tient au cadre de Savage, cette construction peut donc paraître artificielle ou invraisemblable. D'autre part, les postulats de Savage, qui caractérisent les préférences du décideur à l'intérieur de ce cadre, imposent aussi des conditions particulières à la façon dont il définit les trois composantes de ce cadre. Or ces conditions rencontrent à leur tour certaines difficultés quand elles sont confrontées à la pratique.

Il faut noter que cet ensemble d'hypothèses, à première vue assez restrictif, est difficile à invalider

10. Des décisions contingentes sont des décisions conditionnées par des valeurs elles-mêmes incertaines *a priori*.

pour un observateur extérieur qui le mettrait à l'épreuve sur la base de cas concrets. En effet, la violation d'une hypothèse peut toujours, comme nous le verrons, être mise sur le compte d'un mauvais choix des conséquences, des événements ou des actes, puisque ces notions ne sont pas observables. Un observateur extérieur doit tenter de deviner les événements et conséquences par lesquels un décideur caractérise un problème donné¹¹. De ce fait, le cadre de Savage ne peut être considéré comme une théorie scientifique, dont l'expérimentation doit permettre d'établir la véracité selon les canons du positivisme logique, ou au moins de la rejeter selon le critère de falsifiabilité poppérien¹². Son examen se ramène souvent à des questions d'interprétation, et de ce fait, nous pouvons utilement mener cette discussion à partir de notre exemple précédent, sans avoir formellement développé notre modèle.

Nous commençons, dans la première section de cette partie, par analyser les difficultés liées à la définition même d'un problème de décision chez Savage, avant de consacrer la seconde section à trois implications particulières des postulats : l'indépendance des utilités par rapport aux états, l'absence d'aléa moral et le conséquentialisme.

7.3.1 La définition du problème de décision

Sans viser à l'exhaustivité, il est possible de distinguer quatre critiques fréquemment adressées à la façon dont le cadre de Savage rend compte d'un problème de décision :

- des situations relativement communes se traduisent par des problèmes de décision d'une grande complexité ;
- la distinction entre événements et conséquences est floue ;
- l'hypothèse de partition de l'ensemble des états par les événements est trop contraignante ;
- les actes virtuels sont absurdes.

Nous ne nous attarderons pas sur ce dernier argument, en notant simplement qu'en dépit d'une affirmation assez commune, le cadre de Savage n'exige pas que le décideur prenne effectivement les actes virtuels en considération. Le théorème de représentation nous dit simplement que le décideur se conduit *comme s'il* maximisait une utilité espérée, et la construction de Savage nous indique un moyen, impliquant notamment les actes virtuels, pour dériver la distribution de probabilités et une fonction d'utilité correspondant à cette conduite. Si l'existence d'actes virtuels entache la pertinence de la représentation de Savage, ce ne peut donc être sur le plan conceptuel mais tout au plus sur celui de la pratique, en l'occurrence des possibilités de dérivation des probabilités et utilités. À cette éventualité, Savage nous semble avoir répondu de façon aussi convaincante que possible

11. Dans les applications descriptives, qui nous intéressent ici moins que l'usage normatif, le modélisateur doit au préalable faire l'hypothèse quelque peu héroïque que le décideur use bien des *notions* d'événements et de conséquences pour se représenter un problème donné.

12. Gilboa et Schmeidler (2001) discutent de ce point dans le prologue de leur ouvrage, et proposent d'appeler une telle construction théorique un *cadre conceptuel*.

en affirmant que l'expérience mentale de l'association entre un événement et une conséquence inconciliables en pratique n'avait rien d'irréalisable, en tout cas dès lors que l'on admettait la possibilité du raisonnement contrefactuel¹³. Mais, bien sûr, il reste que les probabilités et utilités du décideur savagien doivent être élicités, et ne peuvent être révélés par ses choix spontanés en raison de l'existence d'actes virtuels.

7.3.1.1 La complexité des problèmes de décision

Revenons cette fois au point de départ de notre exemple : un décideur en charge de la politique de santé publique sait qu'une usine rejette des substances cancérigènes, peut-être à des doses toxiques, mais que son activité procure aussi des bénéfices ; que le population consomme les produits contaminés en quantité inconnue, et peut de ce fait s'exposer au développement du cancer, mais que cette consommation a également des avantages économiques ; que les produits non consommés pourraient être exportés, et que le cancer entraîne des conséquences sanitaires incertaines ; enfin que les actes qu'il peut engager sont de limiter les rejets de l'usine, d'inciter la population à réduire sa consommation, de favoriser les exportations et de mener une campagne de dépistage du cancer. Imaginons que le décideur essaie de représenter cette situation par un problème de décision unique, où tous les enjeux feraient partie du même ensemble de conséquences, où la description de chaque événement lèverait entièrement l'incertitude, et où les actes seraient reformulés comme des fonctions entre les ensembles d'états et de conséquences correspondants. La seule élaboration d'un tel problème, sans même parler de sa résolution, représenterait une tâche d'une complexité extrême.

S'appuyant sur un exemple similaire, Karni estime :

"in many instances, the state-space is too large and complex to be compatible with the limited cognitive ability of decision makers to grasp, let alone be invoked in the decision-making process. In such instances, as the following examples illustrate, the notion of states of the world as an uncertainty-resolving device seems unintuitive, non-compelling, and outright useless, for the purpose of modeling decision-making under uncertainty" (Karni, 2006, p.326).

Gilboa et Schmeidler proposent aussi deux cas de cette nature dans lesquels :

"[Expected Utility Theory] does not describe a plausible cognitive process. Should the decision maker attempt to think in the language of EUT, she would have to imagine all possible outcomes and all relevant states. Often the definition of states of the world would involve conditional statements, attaching outcomes to acts. Not only would the

13. 'To some - perhaps to you - it will seem grotesque if I say that I should not mind being hung so long as it be done without damage to my health or reputation, but I think it desirable to adopt such language so that the danger of being hung can be contemplated in this framework. [...] Any statement of the form, "I would do this if that," is somewhat mysterious philosophically; are the ones you emphasize so much worse than the others?' (Aumann et Savage, 1971).

number of states be huge, the states themselves would not be defined in an intuitive way" (Gilboa et Schmeidler, 2001, p.32).

Ces critiques paraissent excessives en ce qu'elles laissent à penser qu'un cadre conceptuel tel que celui de Savage devrait s'appliquer de façon uniforme à tous les problèmes, indépendamment de leur degré de complexité. Il semble naturel de penser que confronté à une telle situation, un décideur rationnel chercherait à distinguer les différentes sources d'incertitude, identifier les aspects de la situation qui lui importent, évaluer comment ses actions pourraient influencer ces derniers et dans quelle mesure leur issue est affectée par l'une ou l'autre des sources d'incertitude. Il chercherait ainsi à utiliser une partie de ses capacités cognitives à décomposer son problème en une succession de problèmes de décision interdépendants. Dans chacun de ceux-ci, le cadre de Savage pourrait apparaître comme un outil de décision pertinent. La méthode que nous avons exposée à partir d'un exemple relativement complexe suit précisément ce cheminement. Elle permet, comme nous l'avons vu, de réduire considérablement le nombre d'événements, de conséquences et d'actes à considérer par rapport à une approche séquentielle traditionnelle.

7.3.1.2 La distinction entre événements et conséquences

Nombre de situations d'incertitude se décrivent le plus naturellement à partir de l'ensemble d'actes effectifs à la disposition du décideur, ainsi que les conséquences de ces actes qu'il est possible d'entrevoir. En revanche, il est souvent malaisé de résumer l'incertitude portant sur ces conséquences par un ensemble d'événements exogènes. La démarche de Savage conduit alors à associer aux conséquences identifiées un ensemble *ad hoc* d'événements exhaustifs et mutuellement exclusifs, ce qu'Arrow (1971) a formalisé en définissant les événements comme des fonctions de l'ensemble des actes vers celui des conséquences¹⁴. L'espace des états du monde n'est plus alors qu'un artefact « endogène et construit » par le décideur, selon la formule de Machina (2003), ce qui met sérieusement en cause sa pertinence, au moins sur le plan descriptif.

Pour examiner cette question, considérons une variante de notre problème de décision en S et supposons même, dans un premier temps, que le décideur ignore tout des causes des cas de décès ; il connaît simplement un ensemble de valeurs possibles de X_S en présence et en absence d'action de sa part (respectivement sous γ_S et γ'_S), tel que $\{\text{négligeables, modérées, importantes, catastrophiques}\}$. La seule source d'incertitude qu'il puisse associer à son problème est alors un aléa général $\bar{Y}_S$ dont les valeurs sont définies par $x_S = \gamma_S(\bar{y}_S)$ et $x_S = \gamma'_S(\bar{y}'_S)$. Inversement, la seule grandeur à l'aide de laquelle les événements, c'est-à-dire les valeurs de $\bar{Y}_S$, puissent être décrites est X_S : l'événement sera *grave* lorsque les conséquences sanitaires associées, en l'absence de dépistage, seront *importantes*, etc.

14. Selon Gilboa (2009, p. 114), c'est Savage lui-même qui aurait le premier proposé cette définition.

Notons que notre décideur, disposant des actes γ_S et γ'_S sous forme fonctionnelle, peut calculer les valeurs $\bar{y}_S$ comme antécédents des valeurs x_S . Il lui est donc inutile de recourir à l'artifice dispendieux d'événements définis comme fonctions des actes vers les conséquences, qui l'encombrerait d'une grande quantité d'états virtuels¹⁵.

De telles situations sont fréquentes dans notre expérience quotidienne. Par exemple, quand, confrontés à un phénomène dont les sources nous échappent ou sont trop intriquées, nous disons : « si les choses continuent ainsi, alors... », la continuation des « choses » n'est qu'un événement abstrait associé aux conséquences observées, ce qui ne nous interdit pas de lui attribuer une probabilité. Des situations similaires se présentent fréquemment dans des problèmes de décision où les probabilités sont données, et où l'on peut donc les utiliser pour décrire les événements correspondants. Pour la gestion des risques liés à certaines catastrophes naturelles, à titre d'exemple, l'événement de référence est un aléa centennal, défini par ses conséquences et sa fréquence estimée, ce qui n'en fait aucunement un phénomène « exogène ».

Supposons à présent que notre décideur ait établi un lien causal entre le coût de la maladie et sa prévalence, considérée comme une variable exogène ($X_M = Y_M$). Ce lien change la nature de la représentation *dans la seule mesure* où la prévalence de la maladie est un phénomène distinct des conséquences sanitaires. Il devient alors possible de mettre en relation les conséquences avec des événements ayant une réalité propre, comme dans l'équation $x_S = \gamma_S(y_M)$ ou la proposition « Une maladie endémique a , en l'absence de mesure de dépistage, des conséquences modérées ». Nombre d'exemples classiques de la théorie de la décision correspondent à ce cas : l'œuf avarié, lorsque l'on choisit de le casser, est une cause autonome de l'omelette gâchée ; la foudre est une cause autonome de l'incendie de la maison en l'absence de paratonnerre, etc.

Cependant, *du fait même de l'existence autonome de la cause et de l'effet*, leur relation ne peut pas, en toute rigueur, être supposé déterministe, de sorte qu'elle n'est complète qu'en la présence d'une variable aléatoire supplémentaire : $x_S = \gamma_S(x_M, y_S)$. On peut, naturellement, choisir de négliger cette variable, ce dont nous discuterons dans le point suivant.

En résumé, les événements constitutifs d'un problème de décision n'apparaissent comme des artefacts que lorsqu'ils ne sont pas fondés sur les réalisations d'une cause identifiée. Dans ce cas, la description des événements ne peut reposer que sur les conséquences, sans toutefois que cela constitue un obstacle à la conduite du raisonnement savagien.

15. Contrairement aux actes virtuels, ces événements virtuels sont totalement inutiles dans la dérivation des probabilités et utilités du décideur, mais, dans un cadre savagien où la partition des états du monde viendrait à manquer et devrait être construite, ils seraient impossibles à distinguer des événements réels.

7.3.1.3 L'exhaustivité des événements

L'hypothèse selon laquelle le décideur pourrait, en toute circonstance, se donner une partition de l'ensemble des états du monde constitue une troisième source de difficultés. Comme une partition est exhaustive, elle signifie en effet que le décideur attribue une probabilité de 1 à l'union des événements répertoriés. Karni en déduit que le décideur doit estimer connaître le monde dans sa totalité et ignorer la simple possibilité d'une surprise (Karni, 2006, p.326). Cette conclusion mérite toutefois d'être pondérée, puisque l'hypothèse d'exhaustivité laisse le décideur libre d'inclure dans sa partition des événements destinés en quelque sorte à le prémunir contre les surprises. Machina reconnaît cette possibilité, mais considère qu'il s'agirait alors d'un événement résiduel « attrape-tout », dont les conséquences seraient nécessairement très vagues (Machina, 2003, p.31).

À ce point, il peut être utile de revenir à notre exemple. Nous considérons à présent à un problème statique où le décideur ne s'intéresserait qu'à la seule prévalence X_M de la maladie, qu'il saurait liée aux rejets X_R , à la consommation de produits contaminés X_C et à un ensemble d'autres facteurs Y_M . Nous supposons qu'il n'a d'autre acte à sa disposition que la seule observation de X_M (disons qu'il s'agit d'un problème de prévision). Supposons de surcroît que le décideur ne soit pas certain de la valeur du paramètre μ , et qu'en conséquence il le considère comme une variable $X_\mu = f_\mu(Y_\mu) = Y_\mu$, qui devient un parent supplémentaire de X_M dans le graphe causal. Les événements en X_M correspondent alors à des quadruplets (y_R, y_C, y_μ, y_M) , dont chacun est associé à une conséquence :

$$\begin{aligned} x_M &= \gamma_M(x_R, x_C, x_\mu, y_M) \\ &= \gamma_M(\gamma_R(y_R), \gamma_C(y_C), \gamma_\mu(y_\mu), y_M) \\ &= \delta \cdot y_\mu \cdot y_C \cdot 1_{y_R > \rho} + y_M \end{aligned}$$

Cela suffit à notre décideur pour parer à toute éventualité, depuis l'absence de la maladie jusqu'à son extension à l'ensemble de la population, et quelle que soit la combinaison de causes identifiées (X_R , X_C et X_μ) et non identifiées (Y_M) à son origine. Constatons que la conséquence d'aucun événement n'est définie de façon vague.

Parer à toute éventualité, dans ce contexte, signifie assigner une probabilité *a priori* non nulle à tout événement envisageable, c'est-à-dire admettre la possibilité de surprises, pour peu que celles-ci soient imaginables. Les surprises, nous le savons, sont un sujet épineux pour la théorie bayésienne, puisqu'aucune observation ne permet de corriger la probabilité d'un événement selon la règle de révision de Bayes lorsque sa probabilité *a priori* a été estimée nulle. De ce fait, un décideur bayésien qui pense connaître la totalité du monde ne peut jamais revenir sur cette opinion. *A contrario*, un décideur dont la partition initiale de l'ensemble des états du monde est suffisamment riche pourra

toujours réviser ses croyances si celles-ci sont infirmées par l'expérience.

Pour construire cette partition, le décideur sera amené, comme au point précédent, à engendrer des événements à partir des conséquences, c'est-à-dire dans un premier temps à identifier (x_R, x_C, x_μ, y_M) tels que $x_M = f_M(x_R, x_C, x_\mu, y_M)$, puis y_R tel que $x_R = f_R(y_R)$, etc. Cette démarche d'inférence des effets aux causes sera, à nouveau, plus économique que celle consistant à définir les événements comme toutes les fonctions des actes vers les conséquences. Gilboa constate que la résolution du problème de Newcombe passe par un tel « élargissement contrôlé » de l'ensemble d'événements. Il en tire l'enseignement suivant :

"in order to avoid surprises, one should take into account all possible "causal" relationships in the formulation of the states of the world" (Gilboa, 2003, p.12).

mais il n'indique pas comment y parvenir dans le cas général en évitant une « explosion de la cardinalité » des ensembles d'actes et d'états du monde.

Notre représentation permet de mettre en œuvre une version légèrement amendée de cette recommandation : il n'est pas possible de tenir compte de toutes les relations causales, au moins dans la vision de la causalité que nous défendons ici ; mais il est possible d'estimer leurs effets agrégés, et de construire le nombre strictement nécessaire d'événements sur cette base.

Notre réponse à la critique initiale de Karni et Machina est certes incomplète. En premier lieu, elle repose à ce stade sur l'hypothèse que le décideur connaît une structure causale adaptée à son problème. Nous remettons l'analyse de situations où il ne peut pas tenir une telle croyance au chapitre suivant¹⁶. Par ailleurs, le fait même que les événements considérés doivent avoir une probabilité *a priori* non nulle signifie que leur nombre doit être fini. Dans toute application réaliste, il ne pourra même être que très limité.

Pour autant, il semble excessif d'en déduire que le décideur doit croire connaître le monde ; il doit plutôt penser connaître les limites de ce qu'un événement inattendu pourrait signifier pour son état. Ce faisant, naturellement, il s'exposera à quelques surprises.

7.3.2 Les conditions portant sur les préférences du décideur

Les postulats de Savage requièrent du décideur :¹⁷

- qu'il accorde à chaque conséquence une valeur intrinsèque, indépendante de l'état du monde et de l'acte sous lesquels elle se réalise (postulats *P3* et *P4*) ;
- qu'il attribue à chaque événement une probabilité subjective, quel que soit l'acte choisi et la conséquence associée à sa réalisation (postulat *P4*) ;

16. Voir la discussion sur l'incertitude causale (section 8.4.2.2)

17. Nous laissons ici de côté la question de la représentation de l'incertitude elle-même par une distribution de probabilités, qui fera l'objet d'un traitement plus détaillé dans la dernière partie du chapitre suivant (voir en particulier la section 8.4.1).

- qu'il n'évalue un acte que sur la base de ses conséquences et des événements auxquels il les associe (postulat $P2$).

Nous discutons à présent de situations qui mettent en question la pertinence de ces conditions.

7.3.2.1 La dépendance aux états

Supposons que nous formulions le problème de décision dans les termes du point précédent, où le décideur ne semble s'intéresser qu'à la prévalence de la maladie X_M , mais qu'il soit en outre sensible à la qualité de l'offre de soins. Toutefois, ne voyant pas comment lier causalement cette variable à X_M , le décideur aurait choisi de la considérer comme une composante des facteurs exogènes Y_M (il va de soi que cette construction est erronée). Il se retrouverait alors dans une situation où il apprécierait une conséquence x_M différemment selon l'événement qui s'est réalisé, en d'autres termes, où sa fonction d'utilité serait dépendante des états. De telles préférences contreviennent aux postulats $P3$ et $P4$ de Savage, et interdisent une représentation en utilité espérée avec une distribution de probabilités unique sur l'ensemble des états du monde (Drèze, 1961). Supposons cependant qu'après mûre réflexion, le décideur trouve l'origine de cette anomalie : son état personnel n'est en fait pas déterminé par la prévalence de la maladie mais par son coût social, dont la prévalence et la qualité de l'offre de soins sont des causes. Il peut alors modifier son RBC, en y intégrant deux variables supplémentaires X_O et X_S , et reprendre la définition de son problème de décision.

Imaginons à présent que dans la même situation, le décideur se montre en outre sensible au fait que lui-même pourrait être affecté par la maladie. N'étant pas un pur altruiste, son appréciation des conséquences sanitaires de l'épidémie serait très différente selon que cet événement se produirait ou pas, et il serait donc confronté à un nouveau problème de dépendance aux états. Comme dans le cas précédent, celui-ci résulterait d'une sous-spécification de sa représentation causale, mais cette fois d'un type différent. La variable manquante dans le cas précédent représentait une cause identifiée d'un aspect de la situation important aux yeux du décideur. Ici, c'est la variable représentant un tel aspect qui fait défaut et dont la prise en compte résoud le problème de dépendance.

Ces deux cas de reformulation du problème de décision correspondent, en substance, à ceux proposés par Savage en réponse au problème de dépendance aux états dans le problème du choix entre maillot de bain et raquette de tennis (Savage, 1954, p.25) et, de façon peut-être moins claire, dans celui de l'opération chirurgicale de l'épouse de M. Jones (Aumann et Savage, 1971, p. 309). Naturellement, l'argument de Savage ne reposait pas sur une interprétation causale. Mais celle que nous proposons permet à la fois de clarifier le débat et d'étayer son raisonnement.

Dans notre modèle, la dépendance aux états ne peut survenir que si la transposition d'une situation incertaine en RBC a été incomplète. L'utilité d'une conséquence donnée ne peut dépendre d'un élément extérieur que dans deux cas de figure : si celui-ci en est une cause, directe ou indirecte

(c'est-à-dire un ascendant markovien), *qui est connue du décideur* (puisque l'utilité de celui-ci ne peut pas dépendre *a priori* d'aspects du monde qu'il n'a pas identifiés comme causes du problème considéré) ; ou si celui-ci est un aspect du problème qui détermine l'état du décideur et, de surcroît, un aspect dont l'utilité est inséparable de celle de la conséquence considérée. Le problème de la maladie personnelle du décideur (ou de l'opération chirurgicale de sa femme dans le cas de M. Jones) est parfois considérée comme un cas d'inapplicabilité de la théorie de la décision en raison de la modification de la capacité même du décideur à apprécier des conséquences (ce que Gilboa appelle son « moteur hédonique » (Gilboa, 2009, p. 130)). Notre représentation ne permet pas de répondre à la question de validité du raisonnement contrefactuel dans un cas où le monde contrefactuel est radicalement différent du monde actuel. Mais elle permet de distinguer ce problème de celui de séparation de conséquences dont les utilités sont inséparables, ce qui ne nous semble pas avoir été toujours le cas dans les contributions passées à cette question.

L'analyse causale affine aussi notre compréhension d'un autre aspect épineux du cadre de Savage : celui des limites du « petit monde » représentant un problème de décision, et plus précisément des conséquences ultimes à considérer. Dans la mesure où le décideur verrait dans chaque variable qu'il prendrait en compte une influence causale vers une autre variable qui lui importerait, sa représentation pourrait s'étendre à l'infini. Comment lui imposer des limites adéquates pour une situation de décision donnée ? Savage avouait être lui-même incapable de formuler un critère pour cela (Savage, 1954, p. 16). Le décideur peut effectivement arrêter son exploration prospective lorsque l'utilité d'une conséquence lui fournit une mesure satisfaisante de l'utilité des problèmes de décision qui en découlent, comme lorsque l'utilité d'une somme monétaire est considérée approximer l'utilité espérée des usages futurs qui en seront faits. Dans notre modèle, cet « élagage » s'avère toutefois problématique si les problèmes de décision éliminés font parties des conséquences indirectes d'autres problèmes de décision qui sont, eux, conservés. Il apparaît ainsi que le décideur peut éviter de développer une ramification du réseau si ses choix peuvent être représentés par une espérance d'utilité récursive et si aucun problème de décision à venir ne mène à la branche élaguée, c'est-à-dire si celle-ci intègre tous les problèmes de décision contingents au dernier problème considéré.

7.3.2.2 L'aléa moral

Imaginons qu'un observateur ait pour but de dériver les probabilités et utilités du décideur à partir de ses choix dans le problème de décision en X_S . Supposons que cet observateur connaisse le modèle causal que le décideur a en tête et les interventions qu'il peut effectuer, sauf au nœud X_C où il ignore la possibilité pour le décideur d'influer sur le niveau de consommation par le biais du mécanisme f'_c . De ce fait, l'observateur peut incorrectement attribuer l'action du décideur sur X_C à la variable exogène Y_C . Supposons maintenant que dans une situation où les rejets sont im-

portants, le choix optimal du décideur soit de limiter la consommation et, se trouvant dans une situation simplement endémique, de choisir le laisser-faire en X_S ; mais qu'en revanche, lorsque les rejets sont moindres, la meilleure option consiste à ne pas agir sur le niveau de consommation, et de réduire les conséquences sanitaires d'une épidémie plus forte par le biais d'une campagne de dépistage en X_S . Du point de vue de l'observateur, les préférences du décideur en X_S présentent alors une incohérence : le dépistage paraît toujours plus avantageux que le laisser-faire sous l'événement ($Y_S = y'_S$) et moins avantageux sous l'événement ($Y_S = y_S$), mais le choix du décideur semble dépendre de l'ampleur des conséquences attendues. En d'autres termes, la probabilité subjective attribuée à ($Y_S = y_S$) semble changer avec les conséquences. L'observateur est amené à conclure que les préférences du décideur ne se conforment pas au postulat $P4$ de Savage, et ne peuvent de ce fait être représentées par une espérance d'utilité. Le problème, cependant, réside non pas dans le comportement du décideur, mais dans l'interprétation de l'observateur. Il a été qualifié dans la littérature d'aléa moral (Drèze, 1961 ; Arrow, 1963), par référence au problème que rencontre un assureur quand l'existence d'une couverture assurantielle conduit à un changement de comportement de l'assuré dans le sens d'une dégradation du risque. Dans une situation d'aléa moral, la séparation entre états du monde, conséquences et actes est donc entravée par la capacité du décideur à influencer les événements par une action extérieure.

Considérons à présent une situation où le décideur choisit de traiter en premier lieu le problème des débouchés à l'exportation. Si le résultat est positif, il peut laisser l'usine rejeter autant de substance que le nécessite son activité et mener une campagne nationale contre la consommation de produits contaminés. Dans le cas contraire, il choisit de limiter les rejets à la source. L'observateur ignore cette fois l'action en X_E , dont il attribue les effets à Y_E , et il conclut que les préférences du décideur en X_R dépendent non pas des conséquences directes, mais de conséquences assez lointaines. Cet exemple montre qu'il y a aléa moral non seulement lorsque le décideur peut influencer le problème de décision considéré par une action sur ses causes, mais plus généralement lorsqu'il peut en modifier le contexte au sens de l'observation 6.

Notre représentation caractérise le contexte causal d'une décision, et considère que le décideur et la Nature jouent des rôles symétriques dans son établissement. Comme la dépendance aux états, l'aléa moral ne peut y apparaître qu'en cas de sous-spécification du modèle. L'oubli d'une intervention sur une variable X_i affecte alors l'observation de tous les problèmes de décision contingents à X_i .

7.3.2.3 Le conséquentialisme

Dans la théorie de la décision dite « objective », c'est-à-dire celle qui considère des problèmes de décision où les probabilités des événements sont pré-établies, un axiome d'indépendance requiert

que les préférences du décideur entre deux actes soient préservées lorsque chacun de ces actes est « mélangé » par combinaison linéaire avec un troisième. L'un des corollaires de cet axiome, lorsque l'opération de mélange est elle-même interprétée comme une loterie, est une propriété que l'on a convenu de nommer *conséquentialisme* (Hammond, 1988a ; Karni and Schmeidler, 1991). Dans un problème de décision séquentielle, les choix du décideur conséquentialiste à un nœud donné ne dépendent pas de la trajectoire qui l'a amené jusque là, ni des alternatives qui ne se sont pas réalisées. Le principe est également à l'œuvre de façon sous-jacente dans le cadre de Savage, où il découle du postulat *P2* (principe de la chose sûre), et apparaît clairement dès lors que l'on considère un problème de décision dynamique sans le réduire au choix *ex ante* d'une stratégie (Hammond, 1988b). Un décideur qui ne se conforme pas à ce principe peut faire des choix dynamiques incohérents qui l'exposent aux paris hollandais.

En imposant au décideur de ne se préoccuper que des conséquences de ses choix à venir, le conséquentialisme véhicule effectivement un certain nombre de présupposés de la théorie qui s'opposent à une conduite fondée sur des principes déontologiques¹⁸. Le plus important de ces présupposés est peut-être l'absence de distinction entre des conséquences qui relèvent de l'action du décideur et d'autres, par ailleurs équivalentes, qui résultent des « choix de la Nature ». Une des ses formes particulièrement expressives est l'indifférence entre un bien (ou un mal) que l'on a causé et un autre auquel on a permis de se produire (Howard-Snyder, 2007). Machina (1989) décrit le dilemme d'une mère qui préfère voir le sort choisir lequel de ses enfants elle doit privilégier plutôt que d'avoir à en décider elle-même. Il qualifie ses préférences de « non séparables entre événements », mais ce faisant, il omet, nous semble-t-il d'indiquer la véritable origine du problème.

Pour étudier un problème équivalent dans notre modèle, considérons la situation de notre gestionnaire de santé publique lorsqu'il évalue ses actes en X_R . Il a alors la perspective d'obtenir le même niveau de rejets ρ s'il choisit de les limiter (f'_R) que s'il choisit le laisser-faire (f_R) et que la Nature choisisse de limiter l'activité de l'usine ($Y_R = \rho$). Le décideur peut alors souscrire à un principe déontologique selon lequel une conséquence (la faiblesse des rejets) n'aurait pas la même valeur, dans ses aspects positifs (l'absence de toxicité) comme négatifs (la faiblesse de l'activité), selon qu'il l'aurait provoquée ou qu'elle résulterait du cours des choses. Il paraît artificiel d'interpréter cette position comme une dépendance de l'utilité des conséquences à la valeur de la conséquence non réalisée, et d'attribuer l'exclusion de telles préférences au principe de conséquentialisme. Ce que le décideur signifierait est plutôt une appréciation différente de ses actes selon qu'ils respectent le mécanisme causal préexistant (f_R) ou qu'ils le changent (f'_R).

En considérant de la même façon le laisser-faire et les interventions actives du décideur, nous avons donc implicitement souscrit à un principe conséquentialiste important. Cette hypothèse a

18. Pour une discussion générale du conséquentialisme, on pourra se reporter à Sinnott-Armstrong (2006).

été faite lors de la définition de nos concepts initiaux, et non dans le cadre axiomatique, et il en est de même pour la théorie de la décision traditionnelle. Pour en changer, il nous suffit de permettre au décideur d'inclure dans la structuration de son problème une variable supplémentaire de gratification éthique, et d'ajouter des flèches entre tous les nœuds de décision pertinents et le nœud de cette variable.

L'hypothèse de conséquentialisme renvoie donc, de façon ultime, à la définition des conséquences. Si les conséquences d'un acte en X_i tiennent compte de tout ce qui détermine l'état du décideur, alors il est naturel de supposer que les préférences de ce dernier ne dépendent que des conséquences (Hammond, 1989). Comme nous en avons discuté plus haut¹⁹, ceci est assuré dans notre cadre par l'intégration au RBC de toutes les causes de X_i connues du décideur. Si cette condition est respectée, alors il semble y avoir peu d'arguments à opposer à l'hypothèse de conséquentialisme telle que formulée habituellement. Si ce n'est pas le cas, alors nous pouvons conclure, une nouvelle fois, que la structure causale proposée est incomplète.

7.4 Conclusion

À partir d'un exemple fictif, nous avons exposé dans ce chapitre les grandes lignes et les principaux résultats d'un modèle de décision fondé sur les RBC. Nous avons montré comment le dispositif d'évaluation du risque proposé au chapitre précédent pouvait être adapté à un contexte de décision. Par son intermédiaire, nous avons proposé une relecture du cadre de Savage, qui caractérise un problème de décision par des ensembles de conséquences, d'événements et d'actes. Nous avons montré que la prise en compte de ce dispositif causal accroissait très sensiblement la capacité de la théorie de la décision à traiter des problèmes complexes. Nous avons finalement survolé quelques-uns des problèmes classiques d'interprétation que pose la définition d'un problème de décision.

À la lumière de cette dernière discussion, il ne semble pas exagéré de dire que les rapports causaux entre états du monde, actes et conséquences constituent la part refoulée de la théorie bayésienne de la décision, notamment chez Savage. L'existence de tels rapports est évidente dans tous les exemples classiques où le cadre savagien semble bien adapté : l'état de l'œuf et l'action de le casser causent la qualité de l'omelette, la survenue de la foudre et l'absence de paratonnerre causent un incendie, etc. En contrepoint, la théorie doit faire des détours et accepter certaines ambiguïtés afin de ne pas avoir à rendre explicites ses soubassements causaux. Les problèmes que nous venons d'évoquer sont, à cet égard, à l'image de symptômes qui s'estompent largement dès que le non-dit est mis en lumière.

19. Voir la sous-section 7.2.2.3

Chapitre 8

Réseaux décisionnels : cadre et résultats théoriques

*either we have no idea of force or energy,
and these words are altogether insignificant,
or they can mean nothing but that determination of the thought,
acquir'd by habit, to pass from the cause to its usual effect.*

Hume (1739)

8.1 Introduction

La théorie de la décision moderne s'est développée à partir des travaux fondateurs de Ramsey (1926) et de Finetti (1937), pour se structurer autour de deux modèles principaux : celui de décision dans le risque, caractérisé par des probabilités objectives, de von Neumann et Morgenstern (1944), et ceux de décision dans l'incertain de Savage (1954) et d'Anscombe et Aumann (1963), qui représentent l'approche bayésienne. Pour celle-ci, le décideur engendre, face à l'incertitude, sa propre distribution de probabilités (dites subjectives), que des paris bien choisis peuvent lui faire révéler. Nous nous placerons dans ce cadre, qui est le plus général, et nous considérerons comme un cas particulier les situations où le décideur dispose d'une distribution de probabilités établie à l'avance.

Dans ses prémisses, la théorie bayésienne de la décision considère déterminés l'ensemble des états du monde auxquels une situation incertaine peut mener, les actions envisageables par le décideur, ainsi que les conséquences qu'elles produiraient dans chaque état du monde. Elle ignore ainsi les relations de cause à effet qui peuvent exister entre ces grandeurs. Au chapitre précédent, nous avons

indiqué que ceci était la cause de certaines faiblesses et ambiguïtés bien connues du cadre de Savage.

Jeffrey (1965) a cherché, dans sa théorie « évidentielle » de la décision, à combler cette lacune en traitant les trois grandeurs comme des propositions logiques. Ceci permet au décideur d'évaluer une proposition conditionnellement à une autre qu'il est en son pouvoir de rendre vraie, c'est-à-dire à une action. Jeffrey propose ainsi d'utiliser les croyances du décideur relatives à des propositions non causales comme révélateurs des relations causales sur lesquelles il s'appuie. Mais précisément parce qu'elle substitue des dépendances probabilistes aux rapports causaux, la théorie évidentielle se prête à de multiples paradoxes (Skyrms, 1980, p. 128-31).

Par contraste, la théorie « causale » de la décision (Stalnaker, 1968 ; Gibbard and Harper, 1978) représente explicitement l'influence causale d'une action x sur des conséquences y , notée $x \square \rightarrow y$, qu'elle évalue à partir de la notion de monde le plus proche (Lewis, 1973). Pearl a souligné les points de similarité entre cette démarche et celle fondée sur les interventions de type $do(x)$ (Pearl, 2000, pp. 238 et suivantes).

Les réseaux bayésiens ont été, de leur côté, fréquemment associés à la théorie de la décision, particulièrement dans sa version « objective » . Les diagrammes d'influence, qui en étaient l'une des premières esquisses, étaient d'ailleurs conçus par leurs auteurs comme une généralisation des arbres de décision (Howard and Matheson, 1981 ; Shachter, 1986). Leur grande efficacité pour représenter et structurer des problèmes de décision complexes en fait désormais des outils d'application courants de la théorie (Clemen, 1996).

Enfin, l'interface entre les modèles économétriques, dont les RBC constituent une généralisation, et la théorie de la décision a fait l'objet d'une attention renouvelée ces dernières années, après les réflexions initiales de Friedman (1953) et Brainard (1967). Un courant de cette littérature s'est inspiré des modèles axiomatiques de la théorie de la décision en ambiguïté¹ pour en dériver une approche « robuste » de l'usage des modèles économétriques mettant en œuvre des méthodes de la théorie du contrôle (Hansen et Sargent, 2000 ; Hansen et Sargent, 2001 ; Barillas, Hansen et Sargent, 2009). Un autre courant a cherché à dériver, à partir de différentes familles de modèles économétriques, une distribution de probabilités a priori et une mesure de l'incertitude relative à cette distribution (Chamberlain, 2000 ; Chamberlain, 2007), pour leur appliquer les critères de choix de la « théorie statistique de la décision » (Berger, 1980).

Le modèle de réseaux décisionnels que nous développons formellement dans ce chapitre, après en avoir exposé les grands traits au précédent, est à l'intersection de ces champs de recherche. Du fait qu'il s'interprète naturellement comme un modèle de décision dynamique, il rejoint en outre la littérature sur la décision en ambiguïté, qui a été au centre du renouveau de la théorie de la décision depuis trois décennies.

1. Dans tout ce chapitre, nous appellerons ambiguïté toute incertitude qui ne peut être exprimée de façon satisfaisante par une simple distribution de probabilités.

Les définitions, l'axiomatique et les principaux résultats du modèle sont exposés dans la première partie de ce chapitre. La partie suivante montre que le modèle fournit une version causale des principaux modèles de décision statique et dynamique dans le risque, et de décision statique dans l'incertain. Comme nous l'expliquerons, le modèle de décision dynamique dans l'incertain présente des complications supplémentaires que nous n'avons pas résolues à ce jour. La dernière partie aborde les problèmes de décision en situation d'ambiguïté.

Les concepts formels développés dans ce chapitre sont plus ardues que dans les précédents, et les notations sont parfois lourdes. Mais ces inconvénients sont fréquents dans les modèles de décision dynamiques, dont nous devons reproduire la démarche afin de définir des actes causalement liés. Une fois cette définition posée, notre cadre permet, comme nous le verrons, de simplifier et de préciser certains aspects des modèles existants.

8.2 Un modèle de décision causale

Reprenons le cadre formel proposé dans notre définition du risque². Nous nous intéressons à présent au problème de détermination des variables $G_i, i = 1, \dots, n$ dans la structure causale D , c'est-à-dire au choix d'un ensemble d'actes causalement liés. L'ensemble χ représente celui des états d'un décideur, dont chaque variable X_i mesure un aspect. Les variables Y_i , dont chacune représente l'incertitude relative à la réalisation d'une variable X_i , sont désormais des fonctions mesurables définies sur un ensemble d'états du monde Ω .

Pour une combinaison d'actes donnée, l'état $(x_1, \dots, x_n)$ du décideur est entièrement déterminé dès lors qu'un élément ω de Ω survient. De ce fait, l'aspect dynamique de notre représentation repose sur une décomposition de cette survenue. Si le décideur observe directement ω , alors nous ne pouvons considérer que deux états : celui où il détermine ses choix et celui où il en découvre les conséquences ; le problème de décision est statique. Si, en revanche, il observe progressivement des aspects de ω mesurés par les variables Y_i , alors il doit tenir compte des conséquences d'un acte avant d'engager le suivant, et le problème de décision est dynamique. Nous retenons cette dernière situation comme référence dans ce chapitre, et nous traitons de la décision statique comme cas particulier³. Nous supposons donc que toutes les variables X_i sont initialement indéterminées, et que l'incertitude se lève de façon graduelle, dictant au décideur un ordre de traitement de ses problèmes. Lorsqu'il envisage une mesure relative à X_i , le décideur a donc déjà agi sur les variables précédentes, et il en connaît les valeurs ainsi que les limites qu'elles imposent à son choix. Par ailleurs, il doit tenir compte des conséquences directes de sa décision pour X_i et de l'incertitude

2. Voir la partie 5.3 page 181.

3. Naturellement, les cas où certains aspects de l'état du monde (c'est-à-dire les valeurs de certaines variables exogènes) sont observés au même moment, mais pas toutes, se traitent aussi comme des cas particuliers du problème dynamique.

spécifique qui les affecte, mais aussi de ses effets sur le reste de la structure.

La dynamique de l'incertitude, qui se caractérise donc par l'ordre dans lequel le décideur observe la réalisation des variables Y_i , doit respecter la structure causale, de sorte que les parents soient toujours déterminés avant les enfants. Cette hypothèse est la simple traduction du principe commun selon lequel une cause doit précéder temporellement ses effets (Reichenbach, 1956; Good, 1961). Observons qu'elle n'interdit pas que le décideur doive agir sur les effets avant d'avoir connaissance des causes, ce qui constitue un simple cas de décision statique. En revanche, notre représentation ne permet pas de rendre compte de situations où le décideur observe des effets mais reste incertain au sujet de leurs causes. Ces cas relèvent de l'incertitude causale, que nous aborderons dans la dernière partie de ce chapitre. Il faut enfin noter que l'ordre $1, \dots, n$, que nous estimons fixé dans la suite, n'est que partiellement imposé par la structure causale.

8.2.1 Des modèles causaux à la théorie de la décision : discussion préalable

8.2.1.1 États du monde et causes inobservables

Lorsque le décideur considère le choix d'un acte en i , la valeur $x_{Pa(i)}$ des parents de X_i est donc déterminée. Conditionnellement à cette valeur, chaque réalisation ($x_i \in \chi_i$) se trouve alors associée, pour un acte donné, à une réalisation de la variable Y_i . Dans notre représentation, cette dernière est l'image d'un événement dans un espace des états savagien $(\Omega, \mathcal{A})$. En particulier, notre version du postulat $P6$ de Savage requiert que la σ -algèbre $\mathcal{A}$ ne comporte aucun atome (c'est-à-dire aucun événement qui ne puisse être divisé en sous-événements). Cependant, l'approche causale apporte à cette hypothèse une justification qui peut sembler plus intuitive, plus convaincante et surtout plus satisfaisante sur le plan normatif que celle proposée par Savage dans un cadre probabiliste.⁴ Cette justification est fondée sur le lien entre l'ensemble Ω et la variable résiduelle Y_i .

Comme nous l'avons mentionné plus haut, dans une interprétation causale de la relation $X_i = g_i(X_{Pa(i)}, Y_i)$, Y_i représente l'agrégation de toutes les causes inobservables de X_i . Une transposition fidèle de cette relation dans notre nouveau cadre aurait la forme :

$$X_i = g_i(x_{Pa(i)}, Y_i(\omega))$$

Ω est donc l'ensemble de définition de toutes les causes inobservables de chaque X_i .

Les événements en i sont dès lors subdivisibles tant que le décideur peut imaginer différentes

4. La justification usuelle du postulat $P6$, qui a pour corollaire l'absence d'atome dans la σ -algèbre $\mathcal{A}$, est qu'un événement peut toujours être subdivisé en considérant une source d'incertitude extérieure tel qu'un lancer de pièce. Ainsi, une séquence infinie de lancers permet d'obtenir des événements aussi petits qu'on le souhaite (Savage, 1954, p.38). L'explication peine toutefois à convaincre, sur les plans normatif comme positif, car rien ne lie cette source d'incertitude au problème effectif du décideur, et l'on ne voit pas pourquoi ce dernier la prendrait en considération.

combinaisons de causes inobservables conduisant à leur réalisation. Si l'on considère que, dans le cas général, les variables susceptibles d'exercer une influence causale sur X_i sont en nombre infini, il en résulte que l'algèbre $\mathcal{A}$ des événements n'a pas d'atome. En d'autres termes, rien ne permet à un décideur de décrire de façon complète un état du monde en *i a priori* : quel que soit le nombre de causes potentielles qu'il identifie, il peut toujours imaginer qu'un facteur supplémentaire vienne influencer la détermination de la conséquence.

8.2.1.2 Le passage aux actes

La théorie des réseaux bayésiens causaux (RBC) s'intéresse avant tout aux rapports entre *variables*. Les interventions y servent principalement de révélateurs de ces rapports. Ceci explique d'une part que la théorie se soit le plus souvent limitée à la classe restreinte des interventions $do(X_i = x_i)$, qui représente des décisions sans incertitude (ou, dans la terminologie de la théorie de la décision, des actes constants), et d'autre part qu'elle ne se soit guère tournée vers l'analyse dynamique qui consiste à s'interroger sur les conséquences d'une intervention pour le choix d'interventions ultérieures.

En théorie de la décision, au contraire, l'acte est l'objet central, qui doit être traité de la manière la plus générale et la plus systématique possible.

Pour accommoder les RBC à ce changement de perspective, nous avons commencé par étendre la notion d'intervention, en passant des interventions constantes $do(X_i = x_i)$ à toutes les fonctions reliant la valeur de X_i à l'ensemble de ses causes, comme le proposaient déjà Marschak (1950) ou Simon (1953)⁵. Pour éviter toute confusion, nous continuerons, comme au chapitre précédent, à éviter le terme d'intervention, et nous désignerons les fonctions g_i du nom de *mécanismes* afin de marquer la distinction entre leur rôle et celui des actes.

En deuxième lieu, comme esquissé au chapitre précédent, nous inversons l'usage des relations causales, en utilisant les variables pour mettre en évidence les rapports entre actes. Pour clarifier ce point, notons que tout problème de décision en i est conditionné par une valeur de $x_{Pa(i)}$. Ainsi, lorsqu'il considère son problème en i , le décideur trouve à sa disposition un ensemble de mécanismes g_i déterminé par la valeur des parents de X_i , qui découle elle-même d'actes engagés en amont de i . Cet ensemble est en revanche indépendant de tout autre antécédent de X_i . Chacun des mécanismes g_i n'est fonction que du seul état du monde ω , par le biais de la variable Y_i représentant l'effet agrégé des causes inobservables de X_i .

Au nombre de ces fonctions, il peut y avoir, le cas échéant, un « mécanisme » de détermination de la variable X_i existant en dehors de toute intervention du décideur, au sens de l'approche fonctionnelle de la causalité. Ceci correspondrait à une situation que le décideur pourrait qualifier

5. Voir la sous-section 5.3.1

de *laisser-faire*, où il estimerait n'exercer aucune action directe sur X_i . Il est toutefois important de noter que la relation entre une variable et ses parents markoviens, ainsi que l'ensemble initial de mécanismes, sont ici établis subjectivement par le décideur, et qu'ils ne sont pas nécessairement déduits de l'observation d'un mécanisme physique autonome. De même que l'approche personnaliste des probabilités généralise l'approche fréquentiste, nous proposons donc ici une généralisation de l'approche fonctionnelle de la causalité.

Sur cette base et en troisième lieu, nous allons chercher à caractériser l'ensemble des conséquences d'un acte γ_i s'appuyant sur le mécanisme g_i .

Dans la tradition de la théorie de la décision sous incertitude (Savage, 1954; Anscombe et Aumann, 1963), nous considérerons comme acte toute fonction de l'ensemble des états du monde à celui des conséquences. Ceci inclut des actes réels, que le décideur pourra effectivement entreprendre, et des actes virtuels, introduits pour les besoins de dérivation d'une distribution de probabilités et d'une fonction d'utilité. Les actes réels, dans notre approche, coïncident avec les mécanismes disponibles en i .

8.2.2 Définition d'un réseau décisionnel bayésien

Soit Ω un ensemble et $\mathcal{E} = 2^\Omega$ l'ensemble de toutes ses parties. Soit $X = (X_1, \dots, X_n)$ un vecteur de fonctions mesurables (ou variables) définies sur $(\Omega, \mathcal{E})$ et à valeurs dans des ensembles χ_i munis d'algèbres $\mathcal{A}_i$. L'ensemble Ω est celui des *états du monde*, et chacun de ses sous-ensembles non vides est un *événement*. L'ensemble $\chi = \chi_1 \times \dots \times \chi_n$ est celui des *états du décideur*.

8.2.2.1 Modèles décisionnels

Définition 8.1. Une structure causale acyclique pour X est une relation d^X entre les composantes de X vérifiant les propriétés suivantes :

1. d est irréflexive : $\forall i \in \{1, \dots, n\} \quad X_i \notin d^X(X_i)$
2. d n'admet pas de cycle : $\forall \{i_1, \dots, i_k\} \subset \{1, \dots, n\} \quad X_{i_h} \in d^X(X_{i_{h+1}}) \forall h = 1, \dots, k-1 \Rightarrow X_{i_k} \notin d^X(X_{i_1})$
3. Le graphe de d est connecté : $\forall i \in \{1, \dots, n\}, \exists j \in \{1, \dots, n\} \quad X_i \in d^X(X_j) \text{ ou } X_j \in d^X(X_i)$

Nous reprenons la terminologie et les notations introduites au chapitre 5. Les antécédents $X_{j_1}, \dots, X_{j_i}$ d'une variable X_i par d sont appelés ses parents. Pour chaque variable X_i , nous définissons trois ensembles d'indices: $Pa(i)$ pour ses parents; $En(i)$ pour ses enfants; et $Pr(i)$ pour ses prédécesseurs $X_1, \dots, X_{i-1}$. Nous évoquerons également les ascendants d'une variable, c'est-à-dire ses parents, les parents de ses parents, etc., et, de façon similaire, ses descendants. Dans la suite, les variables (ou valeurs) indexées par un ensemble désignent le vecteur des variables (ou valeurs)

correspondantes, et les ensembles indexés par un ensemble désignent le produit des ensembles correspondants. Ainsi $X_{Pr(i)} = (X_1, \dots, X_{i-1})$, $\chi_{Pr(i)} = \bigotimes_{j=1}^{i-1} \chi_j$, et $\chi_{Pa(i)} = \bigotimes_{j \in Pa(i)} \chi_j$. Dans tous les cas où un ensemble d'indices est vide, la variable vectorielle est définie, par convention, comme une constante convenablement choisie, et l'ensemble produit comme le singleton correspondant. Nous aurons, par exemple : $X_{Pr(1)} = x_1^0$; $X_{Pa(i)} = x_i^0$ lorsque X_i est une *racine* de la structure causale, c'est-à-dire lorsque $d^{-1}(X_i) = \emptyset$; et $X_{En(i)} = x_i^0$ lorsque X_i est une *feuille*, soit lorsque $d(X_i) = \emptyset$. Nous appellerons en outre *trajectoire en i* la valeur $x_{Pr(i)}$ du vecteur des prédécesseurs de X_i .

Lorsque nous situerons notre point de vue en i , nous parlerons d'*amont* pour des grandeurs en $j < i$ et d'*aval* pour celles en $j > i$.

Soit $Y = (Y_1, \dots, Y_n)$ un vecteur de fonctions mesurables, que l'on appellera *aléas*, définies sur $(\Omega, \mathcal{E})$ et à valeurs dans des ensembles $\mathcal{Y}_i$ munis d'algèbres $\mathcal{B}_i$.

Définition 8.2. Un mécanisme en i est une fonction de $\chi_{Pa(i)} \times \mathcal{Y}_i$ dans χ_i .

Dans la suite, nous supposons le décideur muni d'un *modèle décisionnel* de X consistant en une structure d^X , un vecteur d'aléas Y et, pour chaque i , d'un ensemble $\mathcal{G}_i^X$ de mécanismes en i . En d'autres termes, nous lui prêtons la croyance que s'il choisit en i un mécanisme g_i , alors la valeur de X_i sera donnée par :

$$X_i(\omega) = g_i(X_{Pa(i)}(\omega), Y_i(\omega))$$

lorsque $\omega \in \Omega$ sera déterminé. Nous imposons que les fonctions g_i soient définies en tout point de $\chi_{Pa(i)} \times \mathcal{Y}_i$, de sorte que la valeur de X_i soit définie quelle que soit la trajectoire en i . Nous notons $\mathcal{G}^X = \mathcal{G}_1^X \times \dots \times \mathcal{G}_n^X$ l'ensemble des vecteurs de mécanismes dont dispose le décideur, et $\mathcal{D}^X$ son modèle décisionnel :

$$\mathcal{D}^X = \langle d^X, Y, \mathcal{G}^X \rangle$$

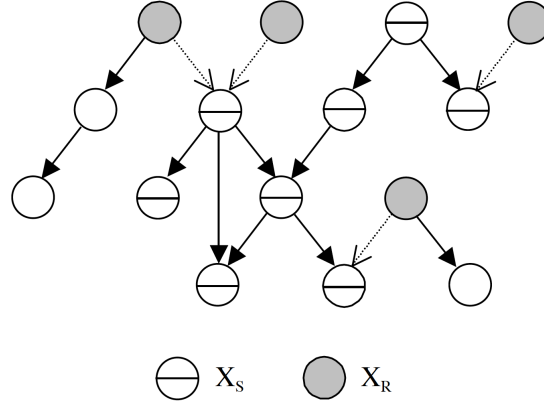
Le vecteur X et le modèle décisionnel $\mathcal{D}^X$ qui lui est associé constituent un *réseau décisionnel*. C'est, comme nous le verrons, l'équivalent d'un petit monde de Savage que l'on aurait doté d'une structure causale.

Pour choisir le modèle causal le mieux approprié à sa situation, nous supposons que le décideur dispose d'une connaissance sous-jacente des variables X , que l'on notera $\mathcal{K}^X$, et d'une relation de préférence $\succeq^*$ (c'est-à-dire une relation transitive et complète) sur l'ensemble des paires (modèle décisionnel, connaissance sous-jacente). $\succeq^*$ se décompose en une relation symétrique $\sim^*$ et une relation asymétrique $\succ^*$.

Définition 8.3. On dit que la restriction d^{X_S} de d^X à un sous-vecteur X_S de X est une sous-structure causale si :

1. d^{X_S} est connectée ;

FIGURE 8.2.1 – Séparation de sous-réseaux décisionnels



2. Tout enfant d'une composante de X_S selon d^X est une composante X_S .
 - Sous ces conditions, on dit que X_S est c -séparé par un sous-vecteur X_R de X si ;
 3. X_R et X_S n'ont aucune composante commune ;
 4. Tout parent d'une composante de X_S qui n'est pas lui-même une composante de X_S est une composante de X_R .
- Enfin, si les conditions 1 à 4 sont remplies, alors on définit $\mathcal{G}_{X_R=x_R}^{X_S}$ comme la restriction de $\mathcal{G}^X$ à X_S dans laquelle la valeur des parents de composantes de X_S extérieurs à X_S serait fixée par $X_R = x_R$. On dit que le modèle décisionnel $\langle d^{X_S}, Y_S, \mathcal{G}_{X_R=x_R}^{X_S} \rangle$ est un sous-modèle de $\mathcal{D}^X$, que l'on note $\langle \mathcal{D}^{X_S} \mid \mathcal{D}_{X_R=x_R}^X \rangle$.

La notion de c -séparation d'un groupe de variables à l'intérieur d'un réseau décisionnel est à comparer à celle de d -séparation à l'intérieur d'un réseau bayésien⁶. Un sous-réseau décisionnel peut ainsi être distingué à l'intérieur d'un réseau principal en affectant des valeurs constantes à certains nœuds, en sectionnant les arcs descendant de ces nœuds et en conservant tous les arcs reliant le groupe de variables ainsi c -séparé. Le diagramme de la figure 8.2.1 en fournit une illustration.

Intuitivement, la c -séparation du groupe de variables X_S signifie que les rapports causaux entre les problèmes de décision correspondants et ceux portés par le reste du réseau font intervenir les variables de X_R . Une fois déterminée la valeur de ces variables, il n'existe plus d'interaction causale entre les deux groupes de problèmes. Un version causale du principe de séparabilité voudrait alors que le décideur, lorsqu'il traite des problèmes de décision à l'intérieur du sous-réseau, soit indifférent aux événements causalement indépendants. L'axiome suivant formule cette condition.

Axiome D1 (séparabilité causale) : Si X_R et X_S sont deux sous-vecteurs de X tels que X_R

6. Voir la sous-section 5.2.1.2

c -sépare X_S , et $\mathcal{D}^{X_S}$ est un modèle décisionnel de X_S alors :

$$(\prec \mathcal{D}^{X_S} \mid \mathcal{D}_{X_R=x_R}^X, K^X) \sim^* (\mathcal{D}^{X_S}, (K^{X_S}, X_R = x_R))$$

8.2.2.2 Problèmes de décision

Par comparaison avec un problème de décision classique, le décideur doit ici apprécier un acte γ_i à l'aune de ses conséquences non seulement pour X_i , que nous appellerons ses *conséquences directes*, mais aussi, via la détermination d'une valeur de X_i , pour les autres variables encore indéterminées ($X_j, j > i$). Ces *conséquences indirectes* consisteront en une modification de certains problèmes de décision en aval de i . Nous nous devons donc de suivre l'évolution d'un problème de décision à mesure que se détermine une trajectoire $x_{Pr(i)}$.

Appelons $\Gamma_{i|x_{Pr(i)}}$ le problème de décision en i et $\mathbf{\Gamma}_{j|x_{Pr(i)}}$ l'ensemble des problèmes de décision en j envisageables par le décideur depuis i ($j > i$). Nous pouvons écrire $\Gamma_{i|x_{Pr(i)}}$ sous la forme générale :

$$\begin{aligned} \Gamma_{i|x_{Pr(i)}} &= (\chi_{i|x_{Pr(i)}} \times \bigotimes_{j \in In(i)} \mathbf{\Gamma}_{j|x_{Pr(i)}})^{\Omega} \\ \mathbf{\Gamma}_{j|x_{Pr(i)}} &= \left\{ (\chi_{j|x_{Pr(i)}, x_i} \times \bigotimes_{k \in In(j)} \mathbf{\Gamma}_{k|x_{Pr(i)}, x_i})^{\Omega}, x_i \in \chi_{i|x_{Pr(i)}} \right\} \end{aligned}$$

où $\chi_{j|x_{Pr(i)}}$ est l'ensemble des valeurs de X_j que le décideur estime compatibles avec la trajectoire $x_{Pr(i)}$, et $In(i)$ est l'ensemble d'indices des problèmes de décision qui résument les conséquences d'un acte en aval de i .⁷

Un acte en i est donc une fonction qui associe à chaque état du monde ω une valeur de X_i compatible avec la trajectoire $x_{Pr(i)}$ et un ensemble de problèmes de décision en certains nœud en aval de i . Un acte en j envisageable depuis i est une fonction qui associe à chaque état du monde ω une valeur de X_j compatible avec la trajectoire $x_{Pr(i)}$ et un ensemble de problèmes de décision envisageables depuis i en certains nœud en aval de j .

Pour que cette définition soit complète, nous devons clarifier la notion de compatibilité et caractériser l'ensemble $In(i)$. Nous commençons la compatibilité, et par une définition préalable.

Notation : Pour tous $\mathcal{E}_i \subset \Omega$ et $\gamma_i, \gamma'_i \in \Gamma_{i|x_{Pr(i)}}$, l'acte qui coïncide avec γ_i sur $\mathcal{E}_i$ et avec γ'_i sur $\mathcal{E}_i^c$ est noté $(\gamma_i, \mathcal{E}_i; \gamma'_i, \mathcal{E}_i^c)$.

Définition 8.4. Un événement $\mathcal{E}_i$ est dit nul si, pour tout $x_{Pr(i)}$ et tous $\gamma_i, \gamma'_i, \gamma''_i \in \Gamma_{i|x_{Pr(i)}}$:

$$(\gamma_i, \mathcal{E}_i; \gamma''_i, \mathcal{E}_i^c) \sim_i (\gamma'_i, \mathcal{E}_i; \gamma''_i, \mathcal{E}_i^c)$$

7. Nous rappelons ici notre convention de notation concernant les ensembles d'indices, notamment pour les feuilles de la structure causale ($En(i) = \emptyset \Rightarrow In(i) = \emptyset$).

La notion de compatibilité traduit le fondement causal de notre approche, selon lequel les valeurs possibles de X_i sont celles que le décideur estime pouvoir associer, par l'intermédiaire de l'un des mécanismes dont il dispose, à une valeur possible des parents de X_i et à un événement possible.

Définition 8.5. Nous disons qu'une valeur x_i est compatible avec une trajectoire $x_{Pr(h)}$ si l'une des conditions suivantes est remplie :

1. $i < h$ et x_i est une composante de $x_{Pr(h)}$;
2. $i \geq h$ et il existe un mécanisme $g_i \in \mathcal{G}_i$ et un événement non nul ($Y_i(\omega) = y_i$) tels que :

$$x_i = g_i(x_{Pa(i)}, y_i)$$

où chaque composante de $x_{Pa(i)}$ est compatible avec $x_{Pr(h)}$.

Nous disons qu'une trajectoire $x_{Pr(i)}$ est compatible avec $x_{Pr(h)}$ si $x_{Pr(i)} = (x_{Pr(h)}, x_h, \dots, x_{i-1})$ où pour tout $k = h, \dots, i-1$, x_k est compatible avec $(x_{Pr(h)}, \dots, x_{k-1})$.

Afin d'identifier l'ensemble $In(i)$ des conséquences indirectes des actes en i , nous devons aussi passer par une définition préalable.

Définition 8.6. Nous disons qu'un problème de décision en j est contingent à une variable X_i (avec $i < j$) s'il existe un sentier de X_j à X_i passant uniquement par des nœuds d'indices supérieurs à i . Nous appelons contexte du problème de décision en i la valeur de toutes les variables dont le problème de décision en i est contingent. Nous notons les ensembles d'indices correspondants $Cg(i)$ et $Cx(i)$:

$$\begin{aligned} Cg(i) &= \{j > i \mid \exists k_1, \dots, k_s \geq i, k_1 = i, k_s = j, \\ &\quad \forall h = 2, \dots, s \quad k_h \in Pa(k_{h-1}) \cup En(k_{h-1})\} \\ Cx(i) &= \{h < i \mid i \in Cg(h)\} \end{aligned}$$

Les variables contingentes à X_i ⁸ sont donc déterminées après X_i , et lui sont reliées uniquement par l'intermédiaire d'autres variables contingentes. Il en résulte que chaque variable contingente à X_i lui est reliée par au moins un sentier qui passe par l'un de ses enfants.

Théorème 8.7. *Sous l'axiome de séparabilité, les conséquences indirectes d'un acte en i sont constituées des problèmes de décision contingents à X_i , et qui ne sont contingents à aucune variable elle-même contingente à X_i :*

$$In(i) = \{j \mid j \in Cg(i) \text{ et } \forall k \in Cg(i), j \notin Cg(k)\}$$

8. Par abus de langage, nous appellerons ainsi les variables dont les problèmes de décision sont contingents à X_i .

Démonstration. Le groupe de variables constitué de X_i et des variables qui lui sont contingentes est c -séparé par des variables de rang inférieur à i . En i , ces variables sont toutes déterminées et leurs valeurs font partie de la trajectoire $x_{Pr(i)}$. Sous l'axiome de séparabilité, le décideur est donc indifférent entre le sous-réseau décisionnel constitué par ce groupe de variables et un réseau décisionnel distinct, calibré de façon appropriée par certaines composantes de $x_{Pr(i)}$. Il en résulte que du point de vue du décideur, les conséquences d'un acte en i sont restreintes à ce sous-réseau, c'est-à-dire à la valeur de X_i et aux problèmes de décision qui lui sont contingents. Les conséquences indirectes des actes en i sont donc des problèmes de décision contingents à X_i .

En outre, le groupe de variables contingents à X_i admet une partition en sous-groupes c -séparés par des variables de rang inférieur ou égal à i . À l'intérieur de chacun de ces sous-groupes, il existe une variable de rang minimal, dont toutes les autres variables sont contingentes. Définissons $In(i)$ comme l'ensemble de ces variables, et poursuivons le fractionnement par c -séparation jusqu'à obtenir des sous-groupes réduits à une variable. Considérons à présent $j \in Cg(i)$ tel que $j \notin In(j)$. D'après notre construction, il existe une suite croissante $k_1, \dots, k_s$ d'indices tels que $k_1 \in In(i)$, $\forall h = 2, \dots, s \quad k_h \in In(k_{h-1})$ et $j \in In(k_s)$. Les conséquences de l'acte en i pour le problème de décision en j font donc partie de celles pour le problème de décision en k_s , qui font elles-mêmes partie de... celles pour le problème de décision en k_1 . $In(i)$ intègre donc bien l'ensemble des effets indirects de l'acte en i . $\square$

Considérons un enfant X_j de X_i . En i , l'ensemble des valeurs de ses parents compatibles avec la trajectoire est :

$$\chi_{Pa(j)|x_{Pr(i)}} = \bigotimes_{k \in Pa(j), k < i} \{x_k\} \times \bigotimes_{k \in Pa(j), k \geq i} \chi_{k|x_{Pr(i)}}$$

La réalisation d'une valeur x_i de X_i entraîne une réduction de cet ensemble à :

$$\chi_{Pa(j)|x_{Pr(i+1)}} = \bigotimes_{k \in Pa(j), k \leq i} \{x_k\} \times \bigotimes_{k \in Pa(j), k > i} \chi_{k|x_{Pr(i)}}$$

La réduction de $\chi_{Pa(j)|x_{Pr(i)}}$ implique à son tour celle de $\chi_{j|x_{Pr(i)}}$ et de $\Gamma_{j|x_{Pr(i)}}$. Un argument similaire montre que les enfants de X_j , puis leurs propres enfants, sont affectés de la même façon. Pour tous les problèmes de décision au nœuds descendants de X_i , la contingence signifie donc une réduction de l'ensemble des conséquences directes compatibles avec la trajectoire.

Considérons à présent un parent X_k de X_j (avec $k > i$). Alors le problème de décision en j est contingent à X_k . Sa modification par la réalisation de x_i se répercute donc sur les conséquences indirectes des actes en k , et réduit de ce fait l'ensemble des problèmes de décision envisageables en k . Le même raisonnement peut être tenu pour tous les ascendants de X_j , et plus généralement pour tous les ascendants d'un descendant de X_i . Pour les problèmes de décision aux nœuds ascendants

d'un descendant de X_i , la contingence signifie donc une réduction des conséquences indirectes compatibles avec la trajectoire.

Considérons finalement un enfant X_h de X_k . Alors la modification du problème de décision en k consécutive à la réalisation de x_i peut entraîner un changement du choix optimal du décideur en k . Le problème en h , qui est contingent à X_k , en serait affecté. Pour les problèmes de décision aux nœuds descendants de l'ascendant d'un descendant de X_i , la contingence signifie donc la réduction possible de l'ensemble des conséquences directes compatibles avec la trajectoire à la suite de choix du décideur dans des problèmes de décision contingents à X_i .

Ainsi, par répercussions successives, tous les problèmes de décision contingents à X_i sont affectés par sa détermination. Le théorème montre cependant que le décideur peut considérer ces effets indirects à travers un petit sous-ensemble de problème (un pour chaque sous-réseau décisionnel séparé par X_i).

Corollaire 8.8. *Le problème de décision en i ne dépend de la trajectoire que par le biais du contexte en i .*

Démonstration. Le théorème montre que les seuls actes qui aient des conséquences pour le problème de décision en i correspondent aux variables dont il est contingent. Par définition, le contexte en i est constitué de la valeur de ces variables.

□

On peut donc réindicer toutes les composantes du problème de décision en i par le contexte $x_{Cx(i)}$ en lieu de la trajectoire $x_{Pr(i)}$.

Corollaire 8.9. *Pour tout $j \in In(i)$, le problème de décision en j est entièrement déterminé avec la réalisation d'une valeur de X_i .*

Démonstration. Le problème de décision en $j \in In(i)$ n'est contingent à aucune variable autre que X_i (sinon le problème de décision correspondant serait contingent à X_i , et il y aurait une contradiction avec la définition de $In(i)$), de sorte qu'il est invariable à la résolution de l'incertitude aux nœuds compris entre i et j :

$$\forall j \in In(i) \quad \Gamma_{j|x_{Cx(j)}} = \Gamma_{j|x_{Cx(i)}, x_i}$$

□

Les problèmes de décision inclus dans $In(i)$ sont donc des grandeurs déterministes, comme il se doit pour toute conséquence dans le cadre de Savage. Comme chaque valeur de X_i détermine un problème de décision unique en chacun des nœuds de $In(i)$, il est possible de définir une fonction

$\hat{\Gamma}_{In(i)|x_{Cx(i)}}$ telle que :

$$\begin{aligned} \hat{\Gamma}_{In(i)|x_{Cx(i)}} : \chi_{i|x_{Cx(i)}} &\rightarrow \bigotimes_{j \in In(i)} \Gamma_{j|x_{Cx(i)}} \\ x_i &\mapsto \bigotimes_{j \in In(i)} \Gamma_{j|x_{Cx(i)}, x_i} \end{aligned}$$

8.2.2.3 Relation binaire sur l'ensemble des actes

Nous supposons que les choix du décideur à l'étape i se conforment à une relation $\succeq_i$ sur l'ensemble des actes en i . $\succeq_i$ se décompose en une relation symétrique $\sim_i$ et une relation asymétrique $\succ_i$.

Nous prêtons à la relation causale $X_i = d^X(X_{Pa(i)})$ établie par le décideur la signification suivante : le décideur croit que l'aspect de son état mesuré par X_i dépend de la valeur des parents de X_i et, pour le reste, de la réalisation d'un aléa Y_i . En particulier, cet aspect est indépendant de la valeur des autres prédécesseurs de X_i une fois que la valeur de ses parents est connue. Comme il en est de même pour tout i , l'*appréciation* des conséquences en i par le décideur peut être estimée indépendante de trajectoire. Toute dépendance de la relation $\succeq_i$ vis-à-vis de la trajectoire proviendrait alors des croyances du décideur, et signifierait qu'il ne considère pas les variables Y_i comme mutuellement indépendantes. Nous excluons cette possibilité par un axiome.

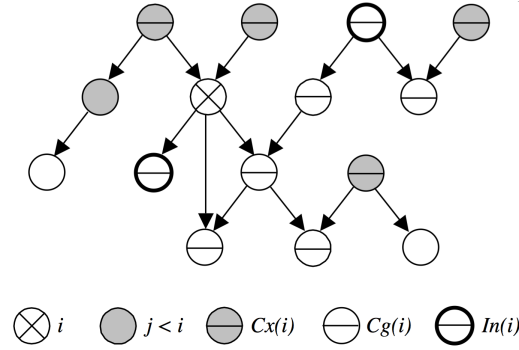
Axiome D2 : La relation $\succeq_i$ est indépendante de la trajectoire en i .

L'axiome *D2* implique que les croyances dérivées de la relation de préférences $\succeq_i$ sont elles-mêmes indépendantes de la trajectoire, et donc qu'elles vérifient la condition de Markov.

Sous les axiomes *D1* et *D2*, nous qualifierons le modèle décisionnel D^X de markovien, et nous dirons qu'associé à X , il constitue un réseau décisionnel bayésien. Il faut noter que la propriété de Markov est préservée lors du passage d'un réseau à un sous-réseau par c -séparation.

8.2.3 Synthèse

Le diagramme de la figure 8.2.2 fournit une illustration des notions développées dans cette section. Le choix d'un acte en i constitue un arbitrage du décideur entre des conséquences directes et indirectes. Pour chaque état du monde, les conséquences directes d'un acte sont une valeur de X_i , et ses conséquences indirectes sont un sous-ensemble, désigné par l'ensemble d'incides $In(i)$, des problèmes de décision conditionnés par cette valeur, soit $Cg(i)$. Ces problèmes sont tous modifiés par la réalisation d'une valeur de X_i . Cependant, leur ensemble admet une partition en sous-ensembles r -séparés par X_i (ici au nombre de deux), de sorte qu'avec la réalisation de X_i , ces sous-ensembles

Figure 8.2.2: Détermination des conséquences indirectes d'un acte en i 

de problèmes deviennent causalement indépendants : les conséquences prises en compte dans les problèmes d'un sous-ensemble ne portent pas sur ceux d'un autre. Les conséquences indirectes de l'acte en i sont alors constitués du premier problème de décision rencontré par le décideur à l'intérieur de chacun de ces sous-ensembles.

Lorsqu'il considère le problème de décision en i , le décideur a déjà traité les problèmes dont l'indice est inférieur à i . Dans certains cas, le problème de décision en i était contingent aux variables concernées. Ces choix sont ceux qui ont déterminé le problème de décision en i .

Pour évaluer les conséquences d'un acte, un décideur doit se donner une situation de référence pertinente, qui permette en particulier de les isoler de tout autre effet à l'intérieur du système de variables considéré. La démarche la plus communément admise, à cet égard, consiste à conditionner les composantes du problème de décision à la trajectoire $x_{P_r(i)}$. Good, par exemple, recommande le «conditionnement à l'état de l'univers juste avant l'occurrence de la cause» (Good, 1961), et Suppes au «passé tout entier du système» (Suppes, 1970). Nous avons indiqué une méthode pour sélectionner les aspects de l'«univers» qui sont pertinents sur le plan causal pour l'évaluation d'un problème par un décideur, et démontré que sous condition de respect de l'axiome de conséquentialisme, il lui suffisait de conditionner son problème aux seules composantes de la trajectoire qui mesurent ces aspects.

Définition 8.10. Dans le problème de décision en i , nous appelons *état du monde* chaque élément de l'ensemble Ω et *événement* chacun de ses sous-ensembles non vides, *conséquence* chaque élément de l'ensemble $\chi_i|_{x_{Cx(i)}} \times \Gamma_{In(i)|x_{Cx(i)}}$, où $\chi_i|_{x_{Cx(i)}}$ est l'ensemble des valeurs de X_i compatibles avec le contexte en i et $\Gamma_{In(i)|x_{Cx(i)}}$ l'ensemble des problèmes de décision déterminés par la valeur de X_i :

$$\Gamma_{In(i)|x_{Cx(i)}} = \bigotimes_{j \in In(i)} \Gamma_j|_{x_{Cx(i)}}$$

$$\begin{aligned}\Gamma_{i|x_{Cx(i)}} &= (\chi_{i|x_{Cx(i)}} \times \bigotimes_{j \in In(i)} \Gamma_{j|x_{Cx(i)}})^\Omega \\ \Gamma_{j|x_{Cx(i)}} &= \left\{ (\chi_{j|x_{Cx(i)}, x_i} \times \bigotimes_{k \in In(j)} \Gamma_{k|x_{Cx(i)}, x_i})^\Omega, x_i \in \chi_{i|x_{Cx(i)}} \right\}\end{aligned}$$

Nous appelons *acte* chaque élément de l'ensemble $\Gamma_{i|x_{Cx(i)}}$ des fonctions de l'ensemble des états du monde dans celui des conséquences, c'est-à-dire toute fonction qui attribue à chaque état du monde ω une conséquence $(x_i, \hat{\Gamma}_{In(i)}(x_i))$, où :

$$\begin{aligned}\hat{\Gamma}_{In(i)|x_{Cx(i)}}(x_i) &= \bigotimes_{j \in In(i)} \Gamma_{j|x_{Cx(i)}, x_i} \\ \hat{\Delta}_{In(i)|x_{Cx(i)}}(x_i) &= \bigotimes_{j \in In(i)} \Delta_{j|x_{Cx(i)}, x_i}\end{aligned}$$

8.3 Articulation avec la théorie de la décision

8.3.1 Décisions dans le risque

Lorsque les probabilités des événements sont pré-établies, les actes constituent des loteries (c'est-à-dire des distributions de probabilités) sur l'ensemble des conséquences :

$$\begin{aligned}\Delta_{i|x_{Cx(i)}} &= \Delta(\chi_{i|x_{Cx(i)}} \times \Delta_{In(i)|x_{Cx(i)}}) \\ \Delta_{j|x_{Cx(i)}} &= \left\{ \Delta(\chi_{j|x_{Cx(i)}, x_i} \times \bigotimes_{k \in In(j)} \Delta_{k|x_{Cx(i)}, x_i}), x_i \in \chi_{i|x_{Cx(i)}} \right\}\end{aligned}$$

où $\Delta_{In(i)|x_{Cx(i)}} = \bigotimes_{j \in In(i)} \Delta_{j|x_{Cx(i)}}$, et $\Delta_{j|x_{Cx(i)}}$ est l'ensemble des loteries $\Delta_{j|x_{Cx(i)}, x_i} = \Delta(\chi_{j|x_{Cx(i)}, x_i} \times \Delta_{In(j)|x_{Cx(i)}, x_i})$ pour chaque valeur possible de x_i / parties de $\Delta_{j|x_{Cx(j)}} = \Delta(\chi_{j|x_{Cx(j)}} \times \Delta_{In(j)|x_{Cx(j)}})$. En d'autres termes, les prix attachés à un billet de loterie en i sont constitués d'une valeur de X_i et de billets pour des loteries en $j \in In(i)$ qui dépendent de cette valeur.

La relation de préférences $\succeq_i$ porte ici sur des loteries, de sorte que nous devons imposer le respect de la condition de Markov séparément de son indépendance de la trajectoire (axiome $D2$). Nous supposons donc que pour tout i , toute trajectoire $x_{Pr(i)}$ en i et toute suite de loteries $\delta_1 \in \Delta_{1|x_{Cx(1)}}, \dots, \delta_i \in \Delta_{i|x_{Cx(i)}}$, $(\delta_1, \dots, \delta_i)$ sont mutuellement indépendants.

Notre modèle représente alors une version causale du modèle de Kreps et Porteus (1978) (KP dans la suite), qui ont proposé la première représentation d'un problème de choix dynamique par une espérance d'utilité récursive. À leur suite, nous supposons que chaque $\chi_{i|x_{Cx(i)}}$ est un espace compact polonais (c'est-à-dire métrique, séparable et complet). Nous munissons chaque ensemble $\Delta(\chi_{i|x_{Cx(i)}})$ de la métrique de Prohorov, et chaque ensemble $\Delta_{In(i)|x_{Cx(i)}}$ de la métrique

de Hausdorff⁹.

Afin d'alléger les notations, nous omettrons dans la suite l'indice $x_{C(x(i))}$ à chaque fois que cela ne portera pas à confusion.

8.3.1.1 Représentation en utilité espérée

Axiome R1 (préordre) : $\succeq_i$ est complète et transitive.

Axiome R2 (continuité) : Pour tout $\delta_i \in \Delta_i$, $\{\delta_i \in \Delta_i \mid \delta_i \succeq_i \delta'_i\}$ et $\{\delta_i \in \Delta_i \mid \delta'_i \succeq_i \delta_i\}$ sont fermés.

Axiome R3 (indépendance) : Pour tous $\delta_i, \delta'_i \in \Delta_i$, si $\delta_i \succ_i \delta'_i$ alors $\forall \alpha \in [0, 1], \delta''_i \in \Delta_i$ $\alpha \cdot \delta_i + (1 - \alpha) \cdot \delta''_i \succ_i \alpha \cdot \delta'_i + (1 - \alpha) \cdot \delta''_i$.

Proposition 8.11. *Pour tout i et pour tout contexte $x_{C(x(i))}$ en i , $\succeq_i$ satisfait les axiomes 1 à 3 si et seulement s'il existe une fonction continue et bornée $u_i : \chi_i \times \Delta_{In(i)} \rightarrow \mathbb{R}$ telle que pour tous $\delta_i, \delta'_i \in \Delta_i$:*

$$\delta_i \succeq_i \delta'_i \Leftrightarrow E_{\delta_i} u_i \geq E_{\delta'_i} u_i \quad (8.3.1)$$

$$\delta_i \succeq_i \delta'_i \Leftrightarrow \sum_{x_i \in \chi_i} \delta_i(x_i) \cdot u_i(x_i, \hat{\Delta}_{In(i)|x_{C(x(i))}}(x_i)) \geq \sum_{x_i \in \chi_i} \delta'_i(x_i) \cdot u_i(x_i, \hat{\Delta}_{In(i)|x_{C(x(i))}}(x_i))$$

Démonstration. La proposition est une simple application de la version continue du résultat de von Neumann et Morgenstern (voir, par exemple, Grandmont (1972)).

□

8.3.1.2 Représentation récursive

Il nous reste à lier les choix opérés par le décideur aux différents nœuds du réseau, de façon à en garantir la cohérence. La difficulté, pour cela, est de mettre en relation des problèmes en i et en j sans que se résolve, entre les points où le décideur considère ces deux problèmes, la moindre

9. Cet équipement topologique relativement lourd nous est nécessaire pour obtenir des fonctions d'utilité continues, qui atteignent leurs maxima sur des ensembles de distributions de probabilités compacts. Nous faisons appel à deux résultats mathématiques fondamentaux :

1. Si (X, d) est un espace métrique séparable et complet et $P(X)$ est l'ensemble des mesures de Borel sur X , alors la convergence faible est la convergence dans la topologie définie sur $P(X)$ par la métrique de Prohorov π , et $(P(X), \pi)$ est un espace métrique séparable et complet.
2. Soit (X, d) est un espace métrique et H l'ensemble de toutes les parties fermées, bornées et non vides de X , muni de la métrique de Hausdorff D . Si (X, d) est complet, alors (H, D) l'est aussi. Si X est compact pour la topologie de d , alors H est compact pour la topologie de D .

incertitude qui soit de nature à entraîner un changement de ses préférences. KP ont contourné cette difficulté à l'aide de loteries dégénérées (qui représentent des actes constants). Dans un cadre où les préférences sont dépendantes de la trajectoire, leur axiome de cohérence temporelle lie ainsi la préférence du décideur en i , conditionnée par $x_{PR(i)}$, entre deux *loteries dégénérées* (x_i, δ_{i+1}) et (x_i, δ'_{i+1}) à sa préférence en $i + 1$, conditionnée par $(x_{PR(i)}, x_i)$, entre les deux *problèmes de décision* δ_{i+1} et δ'_{i+1} . Notons que la condition de cohérence temporelle, telle que formulée, à titre d'exemple, par Raiffa (1968) dans son traitement du paradoxe d'Allais ou par Hammond (1976) dans un problème de choix sans incertitude, porte sur les actes. Pour leurs propres besoins, KP l'appliquent aux problèmes de décision. Pour cela, ils définissent l'utilité d'un problème de décision comme l'espérance d'utilité maximale des actes qui la composent, et associent à cette fonction d'utilité une relation de préférence étendue aux problèmes de décision.

Dans le modèle de KP, la dépendance à la trajectoire ne constitue donc pas simplement un enrichissement permettant de rendre compte de l'évolution de l'utilité du décideur à mesure qu'il observe certaines conséquences. Elle est un moyen de relier ses choix en l'absence de rapport interne entre les conséquences directes et indirectes d'un acte, apparente dans l'association de deux problèmes de décision différents en $i + 1$ à une même valeur de x_i dans la condition de cohérence temporelle. Ce moyen est coûteux, toutefois, puisqu'il impose de définir une relation de préférences à chaque point de chaque trajectoire possible (soit de l'ordre de m^{n-1} relations pour n loteries à m issues chacune), et qu'il impose en outre à KP de démontrer la continuité des fonctions d'utilité dérivées *en la trajectoire*.

Dans notre modèle, à l'inverse, il existe une relation de cause à effet entre la réalisation d'une conséquence directe x_i et celle des conséquences indirectes $\Delta_{In(i)}$, et il n'est pas nécessaire de faire porter un tel lien par la relation de préférences. Mais de ce fait même, il est impossible de comparer des actes constants dont les conséquences directes seraient identiques, et les conséquences indirectes différentes. Il nous faut alors étendre la relation de préférences non pas aux problèmes de décision, mais à leurs parties, c'est-à-dire aux ensembles d'actes.

Pour toute paire π_i, π'_i de sous-ensembles non-vides de Δ_i , nous définissons ainsi les extensions suivantes de la fonction d'utilité u_i et de la relation de préférences $\succeq_i$:

$$\begin{aligned} u_i(\pi_i) &= \max_{\delta_i \in \pi_i} E_{\delta_i} u_i \\ \pi_i \succeq_i \pi'_i &\Leftrightarrow \forall \delta'_i \in \pi'_i, \exists \delta_i \in \pi_i \quad \delta_i \succeq_i \delta'_i \end{aligned}$$

Axiome R4 (cohérence temporelle) : Pour tout $x_i \in \chi_i$, pour toute suite de paires π_j, π'_j de

sous-ensembles de $\Delta_{j|x_{Cx(i)}, x_i}$ pour chaque $j \in In(i)$, pour tout $k \in In(i)$:

$$(x_i, \bigotimes_{j \in In(i) \setminus \{k\}} \pi_j \times \pi_k) \succeq_i (x_i, \bigotimes_{j \in In(i) \setminus \{k\}} \pi_j \times \pi'_k) \Leftrightarrow \pi_k \succeq_k \pi'_k$$

Proposition 8.12. *Le respect des axiomes 1 à 4 par la relation $\succeq_i$ pour tout i et pour tout contexte $x_{Cx(i)}$ est nécessaire et suffisant à l'existence d'une fonction continue $u_i : \chi_i \rightarrow \mathbb{R}$ pour chaque feuille et d'une fonction continue $V_i : \chi_i \times \mathbb{R}^{|In(i)|} \rightarrow \mathbb{R}$, strictement croissante en son second argument, pour les autres nœuds, telles qu'en définissant récursivement :*

$$\begin{aligned} u_i(x_i, \Delta_{In(i)}) &= V_i(x_i, (u_j(\Delta_j))_{j \in In(i)}) \\ &= \max_{(\delta_j \in \Delta_j)_{j \in In(i)}} V_i(x_i, (E_{\delta_j} u_j)_{j \in In(i)}) \end{aligned} \quad (8.3.2)$$

chaque fonction u_i représente $\succeq_i$ au sens de la proposition 8.11.

Démonstration. Pour la suffisance, considérons les fonctions u_i de la proposition 8.11, et fixons $i, x_{Cx(i)}$ et $x_i \in \chi_i|_{x_{Cx(i)}}$. Alors, comme nous l'avons montré dans la section précédente, x_i définit un unique problème de décision $\Delta_{j|x_{Cx(i)}, x_i}$ pour chaque $j \in In(i)$. Il existe donc un unique vecteur $(a_1, \dots, a_{|In(i)|})$ de réels tel que :

$$\begin{aligned} (a_1, \dots, a_{|In(i)|}) &= (u_j(\Delta_j))_{j \in In(i)} \\ &= \max_{(\delta_j \in \Delta_j)_{j \in In(i)}} (E_{\delta_j} u_j)_{j \in In(i)} \end{aligned} \quad (8.3.3)$$

La fonction V_i peut donc être définie par $V_i(x_i, a_1, \dots, a_{|In(i)|}) = u_i(x_i, \Delta_{In(i)})$ lorsque $(a_1, \dots, a_{|In(i)|})$ vérifie (8.3.3), et complétée de façon arbitraire. La monotonie de V_i en son second argument découle facilement de l'axiome R4.

En sens inverse, si les fonctions u_i représentent les relations $\succeq_i$, alors les trois premiers axiomes sont vérifiés par application du théorème R1, et le quatrième est impliqué par la monotonie de V_i . $\square$

La convexité de V_i par rapport à son second argument décrit le comportement du décideur vis-à-vis du caractère spécifique de l'incertitude en i . Les interprétations traditionnelles de cet agrégateur en font une mesure des préférences vis-à-vis de la dimension temporelle de l'incertitude (Kreps et Porteus, 1978; Epstein et Zin, 1989) : lorsque l'indice t se rapporte au temps, la convexité de la fonction V_t signifie que le décideur a une aversion pour la résolution tardive de l'incertitude portant sur l'issue de la loterie en $t + 1$, c'est-à-dire qu'il est impatient.

Notre représentation apporte un éclairage nouveau au rôle de l'agrégateur V_i : sa convexité peut être interprétée ici comme le signe que le décideur a de l'aversion pour le caractère contingent des

problèmes de décision $In(i)$, c'est-à-dire pour le fait que leur expression dépend de x_i , grandeur elle-même incertaine.

Notons aussi que le rôle de l'axiome de cohérence temporelle est ici plus limité que dans les représentations récursives s'inscrivant dans la tradition de KP. Comme chaque valeur de x_i définit les conséquences indirectes $\Delta_{In(i)|x_{C_{x(i)}}, x_i}$ de façon univoque, les fonctions d'utilité u_i sont en fait des fonctions d'une seule variable, et l'équation (8.3.3) suffit à assurer que V_i est bien définie. L'axiome $R4$ n'est nécessaire qu'à la monotonie de V_i .

8.3.2 Décision dans l'incertain

Nous revenons maintenant au cas général où les probabilités ne sont pas pré-établies. Nous continuerons d'omettre l'indice $x_{C_{x(i)}}$ dans cette section.

8.3.2.1 Représentation en utilité espérée

Postulat 1 : $\succeq_i$ est complète et transitive.

Postulat 2 : Pour tout $\mathcal{E}_i \subset \Omega$ et $\gamma_i, \gamma'_i, \beta_i, \beta'_i \in \Gamma_i$ tels que $\gamma_i = \gamma'_i, \beta_i = \beta'_i$ sur $\mathcal{E}_i$ et $\gamma_i = \beta_i, \gamma'_i = \beta'_i$ sur $\mathcal{E}_i^c$:

$$\gamma_i \succeq_i \beta_i \Leftrightarrow \gamma'_i \succeq_i \beta'_i$$

Définition : Un acte constant est un acte γ_i tel que $\exists x_i \in \chi_i, \forall \omega \in \Omega \quad \gamma_i(\omega) = (x_i, \hat{\Gamma}_{In(i)}(x_i))$.

Nous noterons un tel acte $\tilde{x}_i$.

Postulat 3 : Pour tout événement non nul $\mathcal{E}_i \subset \Omega$, pour tous $\gamma_i \in \Gamma_i$ et $x_i, x'_i \in \chi_i$:

$$(\tilde{x}_i, \mathcal{E}_i; \gamma_i, \mathcal{E}_i^c) \succeq_i (\tilde{x}'_i, \mathcal{E}_i; \gamma_i, \mathcal{E}_i^c) \Leftrightarrow \tilde{x}_i \succeq_i \tilde{x}'_i$$

Postulat 4 : Pour tous $\mathcal{E}_i, \mathcal{E}'_i \subset \Omega$ et $x_i, x'_i, w_i, w'_i \in \chi_i$ tels que $\tilde{x}_i \succ_i \tilde{x}'_i$ et $\tilde{w}_i \succ_i \tilde{w}'_i$:

$$(\tilde{x}_i, \mathcal{E}_i; \tilde{x}'_i, \mathcal{E}_i^c) \succeq_i (\tilde{x}_i, \mathcal{E}'_i; \tilde{x}'_i, \mathcal{E}'_i^c) \Leftrightarrow (\tilde{w}_i, \mathcal{E}_i; \tilde{w}'_i, \mathcal{E}_i^c) \succeq_i (\tilde{w}_i, \mathcal{E}'_i; \tilde{w}'_i, \mathcal{E}'_i^c)$$

Postulat 5 : Il existe $x_i, x'_i \in \chi_i$ tels que $\tilde{x}_i \succ_i \tilde{x}'_i$.

Postulat 6 : Pour tous $\gamma_i, \gamma'_i \in \Gamma_i$ et $x_i \in \chi_i$, si $\gamma_i \succeq_i \gamma'_i$ alors il existe une partition $(\mathcal{E}_i^1, \dots, \mathcal{E}_i^k)$ de Ω telle que pour tout $j = 1, \dots, k$:

$$(\tilde{x}_i, \mathcal{E}_i^j; \gamma_i, \mathcal{E}_i^{jc}) \succeq_i \gamma'_i \quad \text{et} \quad \gamma_i \succeq_i (\tilde{x}_i, \mathcal{E}_i^j; \gamma'_i, \mathcal{E}_i^{jc})$$

Proposition 8.13. *Pour tout i et pour tout contexte $x_{C^x(i)}$, si $\succeq_i$ satisfait les postulats 1 à 6, alors il existe une distribution de probabilités non atomique et additive π_i sur Ω et une fonction non constante $u_i : \chi_i \times \Gamma_{In(i)} \rightarrow \mathbb{R}$ telles que pour tous $\gamma_i, \gamma'_i \in \Gamma_i$:*

$$\begin{aligned} \gamma_i \succeq_i \gamma'_i &\Leftrightarrow \int_{\Omega} u_i(\gamma_i(\omega)) d\pi_i(\omega) \geq \int_{\Omega} u_i(\gamma'_i(\omega)) d\pi_i(\omega) \\ &\Leftrightarrow E_{\pi_i} u_i(\gamma_i) \geq E_{\pi_i} u_i(\gamma'_i) \end{aligned} \quad (8.3.4)$$

De plus, π_i est unique et u_i est unique à une transformation positive affine près.

Proof. La proposition est une simple application du principal théorème de Savage (1954). $\square$

π_i est la distribution de probabilités subjectives du décideur sur l'ensemble Ω . L'hypothèse d'indépendance deux à deux des sources d'incertitude, que nous avons avancée dans le cadre des réseaux bayésiens causaux, se traduit ici par l'indépendance des distributions de probabilités subjectives vis-à-vis de la trajectoire, qui découle de l'axiome $D2$.

8.3.2.2 Représentation récursive

L'obtention d'une représentation récursive pose des problèmes particuliers dans le cadre de Savage, où le passage par une fonction d'utilité continue n'est pas possible. À notre connaissance, la seule représentation récursive dans un cadre entièrement subjectif est celle de Skiadas (1998), qui repose sur une démarche originale où la dynamique de l'information et de l'incertitude est représentée par une succession de filtrages plutôt qu'une séquence d'actes. L'application du modèle de Skiadas à notre cadre, que nous n'avons pas réalisée, pourrait ouvrir des pistes de réflexion intéressantes.

8.4 Décision dans l'ambiguïté

Peut-on traduire toute forme d'incertitude par une distribution de probabilités ? Cette question fait l'objet d'un débat très ancien entre les partisans du bayésianisme le plus pur et les autres praticiens des probabilités. Pour les économistes, la question est généralement associée au nom de Knight et à sa distinction entre risque et incertitude (Knight, 1921). Mais elle est plus vieille, puisqu'à la fin du XIXe siècle, von Kries estimait déjà que le « principe de raison insuffisante »¹⁰ ne pouvait avoir la même valeur lorsque son application reposait sur une connaissance précise des conditions de survenue des événements que lorsqu'elle était, au contraire, un choix par défaut imposé par l'ignorance (von Kries, 1886 cité par Bouveresse, 1993, p. 110).

10. Le principe de raison insuffisante, d'équipossibilité ou encore d'indifférence, selon la dénomination plus satisfaisante due à Keynes (1921), établit que deux événements doivent se voir attribuer la même probabilité lorsque rien ne permet de privilégier les chances de survenue de l'un par rapport à l'autre.

Si le premier groupe, celui des partisans du oui, semblait disposer d'arguments définitifs au milieu du siècle dernier, les dernières décennies ont été incontestablement favorables au second.

En premier lieu, la représentation même de l'incertitude a été un domaine actif de recherche et une multitude de propositions se sont posées en alternatives aux probabilités : fonctions de croyance et théorie de l'évidence (Dempster, 1967 ; Shafer, 1976), capacités (Choquet, 1953), ensembles de probabilités (Smith, 1961), mesures de possibilité (Zadeh, 1978), fonctions de rang (Shackle, 1961) ou encore probabilités imprécises (Walley, 1991).¹¹

Par ailleurs, une vague sans cesse croissante de la littérature en théorie de la décision a commencé, dès les années 1950, à critiquer l'approche bayésienne, puis tenté, à partir des années 1980, de la dépasser. Cette littérature, qui est d'une grande diversité, peut être schématiquement répartie en deux courants, l'un empirique ou à visée descriptive, l'autre centré sur les ambitions normatives de la théorie de la décision. C'est au second que nous nous intéressons ici.

Le principal point d'ancrage de ce courant avec les critiques empiriques de la théorie bayésienne traditionnelle, et son principal axe de rénovation de celle-ci, provient du paradoxe d'Ellsberg (1961). Dans une expérience (mentale) de tirage de boules dans des urnes, Ellsberg a observé qu'il était naturel de préférer strictement un pari dont on connaît précisément les probabilités de gain et de perte qu'un pari similaire mais dont les probabilités ne sont connues que de façon vague. Ce trait de comportement a été caractérisé comme une aversion pour l'ambiguïté, c'est-à-dire pour la part d'incertitude qui échappe à une représentation en termes de probabilités.

Le sens dans lequel la théorie de la décision s'est développée au cours des trente dernières années accorde donc une valeur normative à la distinction entre risque et ambiguïté, par opposition à la plupart des autres formes de violation des axiomes de la théorie bayésienne.

Dans cette partie, nous nous intéressons à ces développements, en essayant de montrer à la fois l'étendue de leur apport à la théorie de la décision et les risques de confusion que leur généalogie particulière fait peser sur leur mise en œuvre concrète. En nous appuyant sur l'analyse de la sûreté nucléaire menée dans les chapitres précédents, nous défendons une interprétation particulière de l'ambiguïté et esquissons une extension du modèle théorique de ce chapitre dans cette direction.

La première section passe en revue la littérature sur la décision en présence d'ambiguïté à partir de la contribution séminale d'Ellsberg, et situe le modèle de décision causale présenté plus haut dans ce contexte. Dans la deuxième section, nous nous intéressons aux applications des modèles de décision en présence d'ambiguïté dans des domaines proches des risques sociaux, et nous proposons une critique de ces applications. Dans la dernière section, enfin, nous plaçons pour la distinction d'une forme particulière d'ambiguïté, qui nous semble le mieux rendre compte des cas relevant du principe de précaution : celle qui résulte d'une incertitude relative à la structure causale d'un

11. Pour une revue de ces représentations et des résultats d'équivalence, on se reportera à Halpern (2003, chapitre 2).

problème.

8.4.1 Le paradoxe d'Ellsberg et l'incertitude de second ordre

8.4.1.1 Survol de la littérature

L'une des principales limites de la représentation en utilité espérée est mise en évidence par le paradoxe d'Ellsberg (1961). L'expérience d'Ellsberg, dans sa forme la plus simple, fait appel à deux urnes contenant chacune 100 boules rouges ou noires. Les proportions sont de 50 % pour chaque couleur dans l'urne A, alors qu'elles sont inconnues dans l'urne B. Les probabilités de gain de loteries portant sur l'issue d'un tirage sont les mêmes pour les deux urnes : l'urne B représentant une situation symétrique entre le rouge et le noir, les probabilités de tirage des deux couleurs sont égales en vertu du principe d'indifférence, et valent donc 0,5, comme pour l'urne A. Dès lors, les représentations en utilité espérée, et plus généralement toute formalisation de l'incertitude fondée sur des distributions simples de probabilités, ne permettent pas de distinguer ces deux loteries.

Or les études expérimentales ont montré qu'une telle distinction était implicitement faite par une majorité d'individus, qui ne peuvent de ce fait se conformer aux postulats de Savage, ni, plus largement, aux exigences du bayésianisme (Camerer, 1995). À partir de la fin des années 1970, des modèles alternatifs de décision ont commencé à être proposés afin de rendre compte de cette différence : théorie des perspectives (Kahneman et Tversky, 1979), utilité dépendante du rang (Quiggin, 1982), espérance d'utilité fondée sur les capacités de Choquet (Schmeidler, 1989), *a priori* multiples (Bewley, 2002 ; Gilboa et Schmeidler, 1989), préférences « lisses » (*smooth*) (Klibanoff, Marinacci et Mukerji, 2005 ; Seo, 2009), préférences variationnelles (Maccheroni, Marinacci et Rustichini, 2006), sensibilité à la source de l'incertitude (Ergin et Gul, 2009) ou encore aversion pour l'imprécision (Gajdos et al., 2008). Il est impossible de proposer ici une synthèse de cette littérature ; nous nous contenterons de présenter deux catégories de modèles qui sont d'un intérêt particulier du point de vue des applications qui nous intéressent, et d'énoncer quelques résultats illustratifs pour chacune d'elles.

Les modèles à *a priori* multiples représentent les croyances du décideur non par une distribution de probabilités, mais par un ensemble de distributions. Le degré de pessimisme appliqué par le décideur dans le choix de sa distribution de référence au sein de cet ensemble caractérise son attitude vis-à-vis de l'ambiguïté. L'axiome d'indépendance requiert, comme nous l'avons vu, l'indifférence du décideur à la transformation des actes (ou des loteries) par combinaison linéaire (appelée *mélange*). En se plaçant dans le cadre d'Anscombe et Aumann, Gilboa et Schmeidler (1989) restreignent la condition habituelle d'indépendance à un sous-ensemble des actes, et lui adjoignent un axiome d'aversion pour l'ambiguïté, caractérisée comme préférence pour le mélange des actes : si γ et γ' sont deux actes tels que $\gamma \sim \gamma'$ et α est un entier strictement compris entre 0 et 1, alors

$\alpha\gamma + (1 - \alpha)\gamma' \succeq \gamma$. Ils démontrent alors l'existence d'un ensemble non vide, fermé et convexe $\mathcal{C}$ de distributions de probabilités tel que les préférences du décideur sont représentées par l'espérance d'utilité minimale atteinte sur $\mathcal{C}$:

$$\gamma \succeq \gamma' \Leftrightarrow \min_{\pi \in \mathcal{C}} \int u(\gamma) d\pi \geq \min_{\pi \in \mathcal{C}} \int u(\gamma') d\pi$$

Ce critère de décision, dit *maxmin* (maximisation de l'utilité sous la distribution de probabilités la moins favorable), a été fréquemment critiqué pour son caractère pessimiste. Il convient de noter, cependant, que l'ensemble de probabilités $\mathcal{C}$ n'est pas déterminé dans le modèle de Gilboa et Schmeidler, alors qu'il doit aussi entrer en compte dans la détermination de l'attitude face à l'ambiguïté du décideur. Si, de façon extrême, $\mathcal{C}$ est limité à un singleton, alors cette attitude correspond à l'indifférence et le modèle maxmin coïncide avec l'espérance d'utilité.

La seconde catégorie de modèles représente l'ambiguïté par une probabilité de second ordre. Lorsque les distributions ont des supports finis, l'incertitude se trouve ainsi traduite par une loterie d'ordre deux, et l'axiome d'indépendance doit être reconsidéré dans ce cadre. En effet, tant que le cadre est limité à un ensemble de loteries simples, le mélange auquel l'axiome fait appel est aussi une loterie simple, qui associe à chaque conséquence la combinaison linéaire des probabilités qu'y attachent les loteries élémentaires. Si toutefois nous autorisons l'existence de loteries composées, alors le mélange peut être interprété comme une probabilisation des loteries mélangées, c'est-à-dire comme une loterie à deux niveaux. Si le décideur n'a pas de préférence pour l'une ou l'autre de ces formes, il souscrit au principe de réduction des loteries composées. C'est dans ce cas seulement que l'axiome d'indépendance traditionnel implique l'indifférence entre une loterie simple et une loterie composée de même moyenne, ce qui conduit au paradoxe d'Ellsberg (Segal, 1990).

Il est possible de modifier ce résultat en renonçant d'une part au principe de réduction des loteries composées, pour séparer l'ambiguïté du risque (Segal, 1987), et en distinguant d'autre part la relation de préférence du décideur sur les loteries d'ordre deux de celle sur les loteries simples, afin de caractériser son attitude face à l'ambiguïté (Seo, 2009). À partir d'une démarche voisine, mais adaptée à des distributions de probabilités continues, Klibanoff, Marinacci et Mukerji (2005) obtiennent une représentation de la forme :

$$\gamma \succeq \gamma' \Leftrightarrow \int \varphi \left(\int u(\gamma) d\pi \right) d\mu \geq \int \varphi \left(\int u(\gamma') d\pi \right) d\mu \quad (8.4.1)$$

où π est une distribution de probabilités objectives sur l'ensemble des états, μ une distribution de probabilités subjectives sur l'ensemble des distributions objectives possibles, et φ est une fonction réelle dont la concavité caractérise l'aversion du décideur pour l'ambiguïté. Lorsque φ est linéaire, les deux niveaux de probabilités dans (8.4.1) peuvent être agrégés, et l'on retrouve une représentation

en espérance d'utilité. Lorsque l'aversion pour l'ambigüité est infinie, le modèle est équivalent au maxmin de Gilboa et Schmeidler.

Le lien entre différenciation des loteries d'ordre un et deux et aversion pour l'ambigüité a par ailleurs été confirmée sur le plan empirique par Halevy (2007).

8.4.1.2 Représentation causale

La figure 8.4.1 présente une traduction possible de l'expérience d'Ellsberg par un RBC. X_R et X_I désignent le résultat du tirage respectivement dans l'urne risquée et dans l'urne incertaine, et G_R et G_I les paris du décideur sur leurs valeurs¹². X_R est déterminée, à partir d'une composition connue de l'urne, par la géométrie des boules, leurs positions initiales, la procédure de tirage et d'autres facteurs que l'on peut raisonnablement supposer inobservables par le décideur dans ce contexte particulier. Y_R agrège leurs effets sous une forme probabiliste. Dans le tirage de droite, cependant, la composition X_C de l'urne est également indéterminée, et deux arguments militent pour qu'elle soit singularisée par rapport aux autres facteurs : la prépondérance de son rôle dans le mécanisme de détermination du résultat X_I et le fait que le décideur dispose, en principe, des moyens de se représenter ce rôle, au moins de façon approximative. X_C est déterminée par le seul aléa Y_C et le décideur est appelé à prévoir sa valeur (à travers un premier pari fictif G_C) afin de pouvoir engager son pari final.

Notons par $\Delta_{I|x_C}$ l'ensemble des paris sur le résultat du tirage proposés au décideur conditionnellement à une composition x_C de l'urne, avec $\delta_{I|c}$ comme élément typique, et par δ_C et δ'_C deux distributions de probabilités sur la valeur de x_C , que le décideur compare mentalement. Alors, s'il souscrit aux axiomes $D1$, $D2$ et $R1$ à $R4$, nous savons, d'après les propositions 8.11 et 8.12, qu'il existe des fonctions u_C , V_C et u_I telles que :

$$\begin{aligned} \delta_C \succeq_C \delta'_C &\Leftrightarrow E_{\delta_C} u_C(x_C, \Delta_{I|x_C}) \geq E_{\delta'_C} u_C(x_C, \Delta_{I|x_C}) \\ &\Leftrightarrow E_{\delta_C} \max_{\delta_{I|x_C} \in \Delta_{I|x_C}} V_C(E_{\delta_{I|x_C}} u_I(x_I)) \geq E_{\delta'_C} \max_{\delta_{I|x_C} \in \Delta_{I|x_C}} V_C(E_{\delta_{I|x_C}} u_I(x_I)) \end{aligned}$$

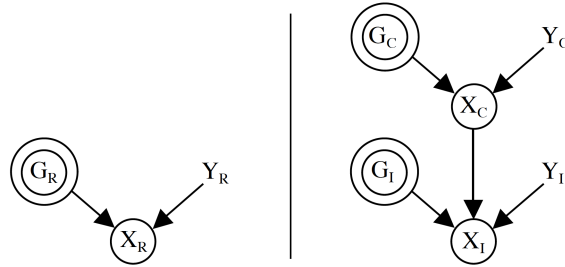
où x_I est le gain final et où l'on a supposé que le décideur n'attachait aucune utilité à la valeur de x_C en soi. On voit ici que la non linéarité de l'agrégateur V_C empêche la réduction des deux niveaux de probabilités, à l'image de la fonction φ dans (8.4.1)¹³.

La représentation avec une distribution de probabilités du second ordre est à rapprocher de la

12. Pour être tout à fait en accord avec nos conventions précédentes, il faudrait ajouter à chaque graphe une variable mesurant le gain du décideur, déterminée uniquement par le résultat du tirage. C'est vers ces variables que devraient pointer les variables de décision G_R et G'_R .

13. Les liens entre les modèles de décision avec incertitude du second ordre et les représentations sous forme d'espérance d'utilité récursive ont été récemment explorés par Ergin et Gul (2009), Klibanoff, Marinacci et Mukerji (2009) et Strzalecki (2010).

FIGURE 8.4.1 – Une représentation causale de l'expérience d'Ellsberg



démarche probabiliste décrite dans le cas de la sûreté nucléaire. Il est intéressant, à ce titre, de noter la proximité entre le modèle de Klibanoff, Marinacci et Mukerji (2005) et le « cadre de la probabilité de fréquence » exposé au chapitre 3, et notamment entre l'équation (8.4.1) et les équations (3.4.1) et (3.4.2) en page 99.

Par contraste, l'application du critère de décision maxmin est à l'image des méthodes déterministes. Dans le cas présent, elle conduirait le décideur à choisir une valeur conservatrice de la composition X_C , et à l'injecter dans le mécanisme de détermination du résultat X_I . À titre d'exemple, si l'ensemble $\mathcal{C}$ des distributions de probabilités retenu par le décideur est l'ensemble des lois de Bernoulli dont le paramètre est de la forme $\frac{n}{100}$ où n est compris entre 1 et 99, alors le critère maxmin est équivalent au choix déterministe d'une valeur $\bar{x}_C = 0,01$.

8.4.2 Quelle interprétation ?

8.4.2.1 Le passage incertain de l'ambiguïté à l'incertitude paramétrique

Après ce survol des grandes lignes de ces deux catégories de modèles de décision dans l'ambiguïté, il faut s'interroger sur leurs usages pratiques. Pour indiquer les applications possibles de leur modèle, Klibanoff, Marinacci et Mukerji citent l'exemple suivant :

"Think of a monetary policy maker setting policy on the basis of a parametric model that solves to yield a probability distribution on a set of macroeconomic variables of interest. However, the probability distribution on variables is conditional on the value of the parameters, which, in turn, is uncertain. That might cause the DM to be concerned enough to seek a policy whose performance is more robust to the uncertainty as to which probability applies. Indeed, such a concern is central to the recent literature investigating decision rules robust to model misspecification or 'model uncertainty' (e.g., Hansen et Sargent (2000))" (Klibanoff, Marinacci et Mukerji, 2005, p.1850).

Or cet exemple, loin de confirmer la pertinence de leur modèle, nous semble en révéler une faiblesse fondamentale.

Les modèles macroéconomiques auxquels les auteurs font référence sont les modèles à équations structurelles qui, comme nous l'avons dit précédemment, sont à l'origine de la version fonctionnelle des réseaux bayésiens causaux et en demeurent un exemple typique. Chaque équation y figure un mécanisme, caractérisé par une forme fonctionnelle et un ensemble de paramètres.

La situation décrite est donc celle d'un décideur qui s'appuierait sur un RBC, mais serait incertain de la valeur de certains paramètres. La figure 8.4.2 en offre une illustration à partir de l'exemple de santé publique discuté au chapitre précédent. On considère ici une incertitude relative à l'élasticité μ de la relation dose-réponse, qui détermine l'ampleur de l'épidémie x_M à partir des valeurs de rejet x_R et de consommation x_C . Le paramètre μ est une grandeur incertaine qui entre en jeu dans le mécanisme de détermination d'une variable ; c'est donc l'un de ses parents markoviens dont le traitement, du point de vue de notre modèle, est similaire à celui de toutes les autres variables.

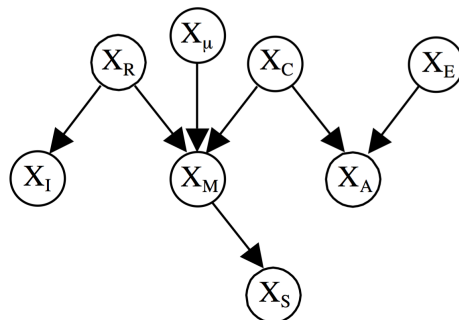
La démarche proposée par Klibanoff, Marinacci et Mukerji apparaît alors reposer sur trois hypothèses :

1. Le décideur distingue certaines variables du réseau, qu'il désigne comme paramètres. En effet, l'une des particularités du cadre bayésien, que les auteurs adoptent, par rapport à l'approche dite classique des statistiques, est qu'il ne fait aucune distinction entre un paramètre incertain et une variable aléatoire.
2. Son attitude face à l'incertitude est du type utilité espérée pour les autres variables du réseau, mais qu'il a une aversion particulière pour l'incertitude relative aux paramètres. Autrement dit, il refuse l'agrégation des distributions de probabilités pour les paramètres mais l'accepte toutes les autres variables du réseau.
3. Il place les paramètres en tête dans son ordre de prise en compte des variables. La non linéarité de l'agrégateur V_μ fige en effet l'ordre de traitement des problèmes de décision dans (8.3.2). Dans un cadre de loteries composées, Sarin and Wakker (1994) ont montré que le relâchement de l'axiome d'indépendance (ou du principe de la chose sûre) imposait de renoncer à la possibilité de combiner et d'invertir les nœuds de l'arbre de probabilités.

Or aucun de ces trois particularités comportementales ne découle de l'axiomatique proposée. Les auteurs traitent le modèle (ou le RBC) à la manière d'une *boîte noire*, dont on évaluerait les paramètres incertains depuis l'extérieur. Ce faisant, ils semblent oublier la diversité des sources d'incertitude qu'ils ont enfermées dans la boîte.

Au-delà du caractère inapproprié de l'exemple choisi par les auteurs, cette discussion montre que l'urne incertaine ne peut servir de modèle qu'à des problèmes de décision soumis à deux sources d'incertitude. L'approche consistant à séparer les niveaux de probabilités impose, en toute rigueur, d'en affecter un à chaque source d'incertitude et de chercher à distinguer, dans le comportement

FIGURE 8.4.2 – Une représentation de l'incertitude paramétrique dans un RBC



du décideur, ce qui relève de l'attitude vis-à-vis de cette source particulière. Dans le problème de la figure (8.4.2), il faut donc considérer huit niveaux de loterie et huit fonctions mesurant le degré d'aversion du décideur pour l'incertitude.

Tant que les trois points relevés ci-dessus n'ont pas trouvé de réponse, la seule alternative à cette inflation semble être l'acceptation des hypothèses bayésiennes traditionnelles, et donc de l'utilité espérée. Il semble donc bien que la littérature sur l'incertitude de second ordre n'ait pas encore de réponse à cette objection soulevée par Savage plusieurs années *avant* la formulation du paradoxe d'Ellsberg :

"Once second order probabilities are introduced, the introduction of an endless hierarchy seems inescapable. Such a hierarchy seems very difficult to interpret, and it seems at best to make the theory less realistic, not more"(Savage, 1954, p.58).

Il est intéressant de noter que Hansen et Sargent, auxquels Klibanoff, Marinacci et Mukerji font référence, évitent précisément de définir une probabilité de second ordre dans leurs travaux sur la robustesse en macroéconomie et en finances¹⁴, se réclamant en cela de Friedman :

"his treatise on monetary policy [...] did not advocate forming prior beliefs over alternative specifications of the dynamic models in response to this concern about model misspecification" (Hansen et Sargent, 2000, p.6).

Le cadre de référence retenu par ces auteurs pour rendre compte de cette incertitude non probabilisable est le modèle à *a priori* multiples de Gilboa et Schmeidler.

8.4.2.2 Incertitude causale et principe de précaution

Nous avons rencontré, au fil des chapitres précédents, deux cas de figure dans lesquels il peut être justifié, sur un plan normatif, de recourir à un raisonnement déterministe. Il s'agit de situations où :

14. Voir par exemple leurs travaux cités dans l'introduction à ce chapitre.

- le décideur ne parvient pas, sur la base de l’information existante, à valider ni à exclure une relation causale entre deux variables d’intérêt ;
- le décideur pense qu’un mécanisme causal est mal spécifié, ce qui le conduit à douter de la distribution de probabilités *a priori* d’une variable exogène.

Ces deux cas relèvent d’une incertitude causale. Dans le premier, le décideur hésite entre deux représentations du monde également satisfaisantes. Dans le second, il n’est satisfait d’aucune des représentations dont il dispose. Naturellement, s’il obéit aux stricts préceptes du bayésianisme, le décideur peut (et doit) exprimer cette incertitude par une distribution de probabilités. Cependant, *s’il suit la démarche prioritairement causale que nous avons décrite*, il est dans l’incapacité de traduire de cette façon son état d’incertitude. Dans les cas que nous avons décrits, il semble effectivement *souhaitable* que le décideur consente à suspendre son jugement probabiliste pour se placer sous l’hypothèse de validité de l’hypothèse causale la moins favorable.

L’incertitude relative à la structure causale d’un problème de décision semble bien caractériser les situations pratiques relevant du principe de précaution. Certaines formulations de ce principe, comme celle de la Conférence de Wingspread sur l’implémentation du Principe de précaution, font même une référence explicite à la notion de causalité :

‘When an activity raises threats of harm to human health or the environment, precautionary measures should be taken even if some cause-and-effect relationships are not fully established scientifically.’ (Raffensperger et Tickner, 1999)

Comme nous l’avons vu au chapitre 2, les tribunaux internationaux et nationaux et certains experts ont également privilégié, au cours des années récentes, une interprétation du principe de précaution comme possibilité d’admettre une hypothèse causale pessimiste mais plausible afin de conduire, sous cette hypothèse, une évaluation normale du risque.

8.5 Conclusion

L’approche fonctionnelle des RBC semble ouvrir de nouvelles perspectives, que nous n’avons souvent pu qu’esquisser ici, dans plusieurs directions : la théorie de la décision causale, les représentations graphiques de la théorie de la décision, l’usage de modèles comme appuis à la décision ou encore la décision en situation d’ambiguïté.

La définition des réseaux décisionnels comporte incontestablement davantage de structure *a priori* que celle des problèmes de décision traditionnels, mais ce prix nous semble raisonnable au regard des avantages obtenus : *décrire en termes naturels* les grandeurs caractéristiques d’un problème de décision et *caractériser leurs rapports causaux*. L’une de ses particularités est de se prêter à une formulation des actions d’un décideur comme choix d’une relation entre une variable et

ses causes directes, en ligne avec une pratique bien ancrée en modélisation économique, mais aussi en sciences, en ingénierie et, éventuellement, dans la vie quotidienne. Par ce biais, il devient notamment possible de rendre compte de façon endogène des effets d'une décision pour des problèmes de décision futurs et d'examiner ainsi les questions de choix dynamique dans un cadre causal. L'approche apporte également un éclairage original à la question de l'ambiguïté, qui a été l'axe principal de renouveau de la théorie de la décision au cours des trois dernières décennies. Elle permet d'améliorer l'expression théorique de notions importantes pour l'étude des risques sociaux, telles que la gestion déterministe du risque ou le principe de précaution.

Conclusion générale

La probabilité et la causalité sont deux notions indispensables à la prise en compte des situations de risque social. Le simple survol historique proposé en introduction nous orientait déjà dans cette direction. Le tour d’horizon de la terminologie du risque dans différentes catégories de risque social, puis l’analyse détaillée des méthodes de la sûreté nucléaire ont confirmé cette impression. Les relations causales sont le fondement des représentations à partir desquelles on peut évaluer le risque et concevoir des actions de prévention, de mitigation ou d’indemnisation. La probabilité permet de réaliser cette évaluation et de calibrer ces actions.

Dès lors, il semble non seulement naturel, mais nécessaire d’intégrer les deux notions à la définition du problème de décision en situation de risque.

En généralisant la version fonctionnelle des réseaux bayésiens causaux, nous avons développé un outil théorique qui répond à ce besoin. Nous avons proposé, par son entremise, une définition du risque qui semble avoir un large champ d’application. Nous avons ensuite proposé des applications simples de ce modèle à des problèmes de sûreté nucléaire. Outre les enseignements spécifiques que nous en avons tirés, ceci nous a conduit à préciser un constat établi lors de l’analyse descriptive de ce domaine, à savoir la nécessité d’une démarche systématique d’identification des incertitudes portant sur la représentation du risque.

Étendu en direction de la théorie de la décision, notre outil débouche naturellement sur un modèle de décision dynamique dans lequel les actes *causent* les conséquences et sont causalement liés entre eux. Il apporte en outre une interprétation causale au cadre conceptuel de Savage et permet d’en résoudre certains paradoxes et lever certaines ambiguïtés.

Ayant ainsi rassemblé toutes les informations pertinentes pour juger des issues d’une situation donnée, le décideur doit-il se désintéresser *de ce qu’il ne sait pas* ? Plusieurs cas pratiques nous ont incliné à penser que non. Dans certaines conditions, l’attitude raisonnable d’un décideur paraît être de suspendre ses croyances probabilistes pour recourir à un raisonnement déterministe. Celui-ci consiste à choisir, entre plusieurs représentations causales alternatives, celle qui décrit la situation la moins favorable. Se fondant sur cette représentation, le décideur peut alors considérer son problème de décision en situation de risque. Une telle attitude correspond précisément au principe

de précaution qui reste, pour les théoriciens de la décision, un sujet difficile, peu adapté au cadre bayésien, même élargi.

Bibliographie

- ABRAMSON, L. R. (1981), 'Some Misconceptions About the Foundations of Risk Analysis', *Risk Analysis*, 1(4), pp. 229-30.
- ADLER, M. D. et E. A. Posner (1999), Rethinking Cost-Benefit Analysis, *Yale Law Journal*, 109(2), pp. 165-230.
- AEC (1957), *Theoretical Possibilities and Consequences of Major Accidents in Large Nuclear Power Plants*. WASH-740. United States Atomic Energy Commission, Washington, D. C.
- AEN (1991), *Disposal Of Radioactive Waste : Can Long-Term Safety Be Evaluated ? An International Collective Opinion*. Agence pour l'Énergie Nucléaire, Organisation pour la Coopération et le Développement Économiques, Paris.
- AEN (1997), *Disposal Of Radioactive Waste : The Probability System Assessment Group - History and Achievements 1985-1994*. Agence pour l'Énergie Nucléaire, Organisation pour la Coopération et le Développement Économiques, Paris.
- AEN (1999), *Confidence in the Long-term Safety of Deep Geological Repositories - Its Development and Communication*. Agence pour l'Énergie Nucléaire, Organisation pour la Coopération et le Développement Économiques, Paris.
- AEN (2000), *Methodologies for Assessing the Economic Consequences of Nuclear Reactor Accidents*. Agence pour l'Énergie Nucléaire, Organisation pour la Coopération et le Développement Économiques, Paris.
- AEN (2007), *Safety Margins Action Plan - Final Report*. Task Group on Safety Margins Action Plan, Committee on the Safety of Nuclear Installations, Agence pour l'Énergie Nucléaire.
- AEN et AIEA (2002), *An International Peer Review of the Yucca Mountain Project TSPA-SR*. Agence pour l'Énergie Nucléaire, Organisation pour la Coopération et le Développement Économiques, Paris.

- AGUILLON, L. (1889), 'Notice historique sur l'Ecole des Mines de Paris', *Annales des mines*, mai-juin. Ecole des mines, Paris.
- AIEA (1979), *Earthquakes and Associated Topics in Relation to Nuclear Power Plant Siting*. Safety guide 50-SG-S1. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (1992), *Procedures for Conducting Probabilistic Safety Assessments of Nuclear Power Plants (Level 1)*. Safety series, n° 50-P-4. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (1995), *Procedures for Conducting Probabilistic Safety Assessments of Nuclear Power Plants (Level 2)*. Safety series, n° 50-P-8. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (1996), *Procedures for Conducting Probabilistic Safety Assessments of Nuclear Power Plants (Level 3)*. Safety series, n° 50-P-12. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (1996b), *Defence in Depth in Nuclear Safety*. Report by the International Nuclear Safety Advisory Group (INSAG-10). Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (2001), *Safety assessment and verification for nuclear power plants*. Safety guide NS-G-1.2. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (2005), *Sûreté des centrales nucléaires : Conception*. Collection Normes de sûreté NS-R-1. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (2007), *Glossaire de sûreté de l'AIEA*. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (2008), *Best estimate safety analysis for nuclear power plants : Uncertainty evaluation*. Safety report series, n° 52. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (2010), *Meteorological and Hydrological Hazards in Site Evaluation for Nuclear Installations*. Draft Safety Guide DS417, Commission on Safety Standards. Agence Internationale de l'Énergie Atomique, Vienne.
- AIEA (2011), *IAEA International Fact-Finding Expert Mission of the Fukushima Dai-Ichi NPP Accident Following the Great East Japan Earthquake and Tsunami*. Agence Internationale de l'Énergie Atomique, Vienne.
- ALDRICH, J. (1989), 'Autonomy', *Oxford Economic Papers*, 41, pp. 15-34.
- ALEMANNO, A. (2007), 'The Shaping of the Precautionary Principle by European Courts : From Scientific Uncertainty to Legal Certainty', in L. Cuocolo et L. Luparia (sous la direction de), *Valori Costituzionali e Nuove Politiche del Diritto*, Cahiers Européens, Halley.

- ALEMANNO, A. (2008), 'The Shaping of European Risk Regulation by Community Courts', *The Jean Monnet Working Papers*, n°18.
- ALLAIS, M. (1953), 'Le comportement de l'homme rationel devant le risque : critique de l'école américaine', *Econometrica*, 21, pp. 503-46.
- ALTEZ, R. (2007), 'Muertes Bajo Sospecha : Investigacion Sobre el Numero de Fallecidos en el Desastre del Estado de Vargas, Venezuela, en 1999', *Cuadernos de Medicina Forense*, 13(50).
- ANDO, M., Ishida, M., Hayashi, Y. et C. Mizuki (2011), 'Interviews With Survivors of Tohoku Earthquake Provide Insights Into Fatality Rate', *Eos*, 92(46), pp. 411-2.
- ANSI (1973), *American national standard : Nuclear safety criteria for the design of stationary pressurized water reactor plants - ANSI 18.2*. Elaboré par l'American Nuclear Society. American National Standards Institute, Washington, D. C.
- ANG, A. H.-S. et W. H. Tang (1984), *Probability Concepts in Engineering Planning and Design*. Volumes 1 et 2, Wiley, New York.
- ANSCOMBE, F. J. et R. J. Aumann (1963), 'A definition of subjective probability', *Annals of Mathematical Statistics*, 34, pp. 199-205.
- APOSTOLAKIS, G. (1990), 'The Concept of Probability in Safety Assessments of Technological Systems', *Science*, 250, pp.1359-64.
- APOSTOLAKIS, G. (2004), 'How useful is quantitative risk assessment?', *Risk Analysis*, 24(3), pp. 515-20.
- APOSTOLAKIS, G. et S. Kaplan (1981), 'Pitfalls in risk calculations', *Reliability Engineering*, 2(2), pp.135-45.
- ARROW, K. J. (1963), 'Uncertainty and the Welfare Economics of Medical Care', *American Economic Review*, 53(5), pp. 941-73.
- ARROW, K. J. (1971), *Essays in the Theory of Risk Bearing*. Markham Publishing, Chicago.
- ARROW, K. J. et R. C. Lind (1970), 'Uncertainty and the Evaluation of Public Investment Decisions', *American Economic Review*, 60, pp. 364-78.
- ASME/ANS (2009), *Standard for Level 1/Large Early Release Frequency Probabilistic Risk Assessment for Nuclear Power Plant Applications*. American Society of Mechanical Engineers et American Nuclear Society, New York.

- ASN (2002), *Règle fondamentale de sûreté n° 2002-01 relative au développement et à l'utilisation des études probabilistes pour la sûreté des installations nucléaires de base*. Autorité de Sûreté Nucléaire, Paris.
- ASN (2002b), *Note d'information du 26 décembre 2002*. Consultée le 8 mai 2011 sur le site de l'Autorité de Sûreté Nucléaire (<http://www.asn.fr/index.php/S-informer/Actualites/2002/Regle-fondamentale-de-surete-RFS-n-2002-01>).
- ASN (2010), *Protection des installations nucléaires de base contre les inondations externes*. Projet de guide de l'ASN soumis à consultation du public, version du 17/05/2010. Autorité de Sûreté Nucléaire, Paris.
- AUMANN, R. J. et L. Savage (1971), 'Letter from Robert Aumann to Leonard Savage and Letter from Leonard Savage to Robert Aumann', in Drèze (1987), pp. 76-81.
- AUTEURS DIVERS (1789), *Encyclopédie méthodique. Jurisprudence. Tome neuvième contenant la police et les municipalités*. Panckouke, Paris et Plomteux, Liège.
- AVEN, T. (2003), *Foundations of Risk Analysis—A Knowledge and Decision Oriented Perspective*. John Wiley and Sons, Chichester.
- AVEN, T. et O. Renn (2009), 'The Role of Quantitative Risk Assessments for Characterizing Risk and Uncertainty and Delineating Appropriate Risk Management Options, with Special Emphasis on Terrorism Risk', *Risk Analysis*, 29(4), pp. 587-600.
- BACZKO, B. (1997), *Job, mon ami : promesses du bonheur et fatalité du mal*. Éditions Gallimard, Paris.
- BAGEHOT, W. (1873), *Lombard Street : A Description of the Money Market*. H. S. King, Londres.
- BALOGH, B. (1991), *Chain Reaction : Expert Debate and Public Participation in American Commercial Nuclear Power, 1945-1975*. Cambridge University Press, Cambridge.
- BARILLAS, F., Hansen, L. P. et T. J. Sargent (2009), 'Doubts or variability?', *Journal of Economic Theory*, 144, pp. 2388-418.
- BECK, C. K., Mann, M. M. et P. A. Morris (1958), 'Reactor Safety, Hazards Evaluation and Inspection', *Proceedings of the Second International Conference on Peaceful Uses of Atomic Energy*, vol. XI, pp. 17-20. Organisation des Nations Unies, New York, 1959.
- BECK, U. (1986), *Risikogesellschaft : Auf dem Weg in eine Andere Moderne*. Suhrkamp Verlag, Francfort sur le Main. Traduction anglaise de Mark Ritter, *Risk Society : Towards a New Modernity*. 3ème édition (2002), Sage Publications, Londres.

- BECKJORD, E. S., M. A. Cunningham et J. A. Murphy (1993), 'Probabilistic safety assessment development in the United States 1972–1990', *Reliability Engineering and System Safety*, 39(2), pp. 159-70.
- BEDFORD, T. et R. Cooke (2001), *Probabilistic Risk Analysis : Foundations and Methods*. Cambridge University Press, Cambridge.
- BERGER, J. O. (1980), *Statistical Decision Theory and Bayesian Analysis*. Springer-Verlag, New York.
- BEWLEY, T. (2002), 'Knightian Decision Theory : Part I', *Decisions in Economics and Finance*, 25, pp. 79-110. (publication d'un document de travail de 1986).
- BIRKMANN, J. (2006), 'Measuring vulnerability to promote disaster-resilient societies : conceptual frameworks and definitions', in J. Birkmann (sous la direction de), *Measuring Vulnerability to Natural Hazards : Towards Disaster Resilient Societies*. United Nations University Press, Tokyo, pp. 9-54.
- BOBBIO A, L. Portinale, M. Minichino et E. Ciancamerla (2001) 'Improving the analysis of dependable systems by mapping fault trees into Bayesian networks', *Reliability Engineering and System Safety*, 71(3), pp. 249–60.
- BOITEUX, L. A. (1968), *La fortune de mer - Le besoin de sécurité et les débuts de l'assurance maritime*. SEVPEN, Paris.
- BOITEUX, M. (1993), *Haute tension*. Éditions Odile Jacob, Paris.
- BOURGEOIS, J., P. Tanguy, F. Cogné et F. Petit (1996), *La sûreté nucléaire en France et dans le monde*. Polytechnia, Paris.
- BOURRELIER, P.-H. (1997), *La prévention des risques naturels – Rapport d'évaluation*, Comité interministériel de l'évaluation des politiques publiques, La Documentation Française, Paris.
- BOUVERESSE, J. (1993), *L'homme probable. Robert Musil, le hasard, la moyenne et l'escargot de l'histoire*. Éditions de l'éclat, Paris.
- BOUYSSOU, D., Marchant, T., Pirlot, M., Tsoukiàs, A. et P. Vincke (2006), *Evaluation and Decision Models with Multiple Criteria - Stepping Stones for the Analyst*. Springer, New York.
- BRAINARD, W. (1967), 'Uncertainty and the effectiveness of policy', *American Economic Review*, 57, pp. 411-25.
- BRIGUGLIO, L., Cordina, G., Farrugia, N. et S. Vella (2009), 'Economic Vulnerability and Resilience : Concepts and Measurements', *Oxford Development Studies*, 37(3), pp. 229-47.

- BROOME, J. (1991), *Weighing Goods*. Blackwell, Oxford.
- BUHL, A. R., Carter, J. C., Fontana, M. H., Henry, R. E. et H. A. Mitchell (1987), 'The IDCOR program - Severe accident issues, individual plant examinations and source term developments', in Lave, L. B. (sous la direction de), *Risk assessment and management*. Plenum Press, New York, pp. 205-17.
- BUSTANY, C. (1992), *Maîtrise du sol et urbanisme à Rome à l'époque républicaine*. Thèse de doctorat, Université de Paris IV.
- CAMERER, C. F. (1995), 'Individual decision making', in Kagel, J.H. et A. E. Roth (sous la direction de), *Handbook of Experimental Economics*. Princeton University Press, Princeton (New Jersey), pp. 587-703.
- CAMERER, C. F. (2000), 'Prospect Theory in the Wild : Evidence from the Field', in Kahneman, D. et A. Tversky (sous la direction de), *Choices, Values, and Frames*. Cambridge University Press, New York, pp. 288-300.
- CANGUILHEM, Georges (1966), *Le normal et le pathologique*. Presses universitaires de France, Paris.
- CARPENTER, D. P. (2004), 'The Political Economy Of FDA Drug Review : Processing, Politics, And Lessons For Policy', *Health Affairs*, 23(1), pp. 52-63.
- CBO (2008), *Nuclear Power's Role in Generating Electricity*. Congressional Budget Office, Washington, D.C.
- CCE (2000), *Communication sur le Principe de précaution*. Document COM(2000)1. Commission des Communautés européennes, Bruxelles.
- CEA (1960), *Note de Service n° C 278 du 27 janvier 1960*. Commissariat à l'Énergie Atomique, archives, fonds du Haut-Commissaire.
- CEC (1981), *Zion Probabilistic Safety Study*. Commonwealth Edison Company, Chicago.
- CHAMBERLAIN, G. (2000), 'Econometrics and decision theory', *Journal of Econometrics*, 95, pp. 255-83.
- CHAMBERLAIN, G. (2007), 'Decision Theory Applied to an Instrumental Variables Model', *Econometrica*, 75(3), pp. 609-52.
- CGTN (2006), *Vocabulaire de l'ingénierie nucléaire*. Commission générale de terminologie et de néologie, Ministère de la culture et de la communication, Paris.

- CHEW, S. H. et L. Epstein (1991), 'Recursive utility under uncertainty', in A. Kahn et N. Yannelis (sous la direction de), *Equilibrium theory with an infinite number of commodities*. Springer, New York.
- CHOQUET, G. (1953) 'Theory of capacities', *Annales de l'Institut Fourier*, 5, pp. 131– 295.
- CLEMEN, R. T. (1996), *Making hard decisions : An introduction to decision analysis*. Wadsworth, Belmont.
- COGNÉ, François, 'Evolution de la sûreté nucléaire', *Revue Générale Nucléaire*, 1, pp. 18-32.
- COMMISSION DES FINANCES, DE L'ÉCONOMIE GÉNÉRALE ET DU PLAN (1977), *Rapport sur la Loi de finances pour 1978*. Rapporteur spécial E. Schloesing, Annexe n° 23 au procès verbal de la séance du 5 octobre 1977, Assemblée Nationale.
- COOKE, R. M. et L. H. J. Goossens (1999), *Procedures Guide for Structured Expert Judgement*. Nuclear Science and Technology, Rapport de projet. Commission des Communautés Européennes, Bruxelles.
- CORNELL, C. A. (1968), 'Engineering seismic risk analysis', *Bulletin of the Seismological Society of America*, 58, pp. 1583-1606.
- CORNU, G. (sous la direction de) (2007), *Vocabulaire juridique*. Association Henri Capitant, 8ème édition, Presses Universitaires de France, Paris.
- COUR DES COMPTES (1999), *Rapport public 1999*. Cours des comptes, Paris.
- CPP (2010), *La décision publique face à l'incertitude*. Avis du Comité de la Prévention et de la Précaution. Accessible sur le site du Ministère du Développement Durable (www.developpement-durable.gouv.fr/IMG/pdf/avis-CPP_Decision_et_incertitude.pdf, consulté le 28 décembre 2011).
- CRS (1994), *Risk Analysis and Cost-Benefit Analysis of Environmental Regulations*. Report 94-961. Congressional Research Service, Washington, D.C.
- CROIX ROUGE JAPONAISE (2011), *Japan : Earthquake and Tsunami, 6-month Report*. 31 août 2011.
- DARWICHE, A. (2009), *Modeling and Reasoning with Bayesian Networks*. Cambridge University Press, Cambridge, Royaume-Uni.
- DASTON, L. (1988), *Classical Probability in the Enlightenment*. Princeton University Press, Princeton, New Jersey.

- DAWID, A. P. (1979), 'Conditional Independence in Statistical Theory', *Journal of the Royal Statistical Society, Series B*, 41(1), pp. 1-31.
- DAWID, A. P. (2004), 'Probability, Causality and the Empirical World : A Bayes-de Finetti-Popper-Borel Synthesis', *Statistical Science*, 19(1), pp.44-57.
- DE FINETTI, B. (1937), 'La prévision : ses lois logiques, ses sources subjectives', *Annales de l'Institut Henri Poincaré*, 7, pp. 1-68.
- DE VALTHAIRE, F. (1960), 'Sûreté des installations atomiques', *Génie atomique*, volume I, section C. Bibliothèque des sciences et techniques nucléaires, Presses Universitaires de France, Paris.
- DEMPSTER, A. (1967), 'Upper and lower probabilities induced by a multivalued mapping', *Annals of Mathematical Statistics*, 38, pp. 325-339.
- DEFRA (2004), *Best Practice in Coastal Flood Forecasting*. DEFRA / Environment Agency Flood and Coastal Defence R&D Programme, R&D Technical Report FD2206/TR1. Department for Environment, Food and Rural Affairs, Londres.
- DER KIUREGHIAN, A. et O. Ditlevsen (2009), 'Aleatory or epistemic? Does it matter?', *Structural Safety*, 31, pp. 105-112.
- DESROSIÈRES, A. (1993), *La politique des grands nombres - Histoire de la raison statistique*. La Découverte, Paris.
- DESTERCKE, S. (2008), *Représentation et combinaison d'informations incertaines : nouveaux résultats avec applications aux études de sûreté nucléaires*. Thèse de doctorat, Université de Toulouse.
- DEVICTOR, N. et R. Bolado Lavín (2005), 'Uncertainty and sensitivity methods in support of PSA level 2', *CSNI Workshop on Evaluation of Uncertainties in Relation to Severe Accidents and Level 2 PSA*. Agence pour l'Énergie Nucléaire.
- DILLEY, M., R. S. Chen, U. Deichmann, A. L. Lerner-Lam et M. Arnold (2005), *Natural Disaster Hotspots : A Global Risk Analysis*. Banque Mondiale, Washington, D. C.
- DITLEVSEN, O. et H. O. Madsen (1996), *Structural Reliability Methods*. Seconde édition (2007), Université technique du Danemark, <http://www.web.mek.dtu.dk/staff/od/books.htm>.
- DOLL, R. et A. B. Hill (1950), 'Smoking and carcinoma of the lung : preliminary report', *British Medical Journal*, 4682, pp. 739-48.
- DOLL, R. et A. B. Hill (1954), 'The mortality of doctors in relation to their smoking habits : a preliminary report', *British Medical Journal*, 4877, pp. 1451-5.

- DOUGLAS, M. (1992), *Risk and blame - Essays in cultural theory*, Routledge, Londres.
- DOUGLAS, M. et A. Wildavsky (1982), *Risk and Culture*. California University Press, Berkeley.
- DRÈZE, J. (1961), 'Les fondements logiques de l'utilité cardinale et de la probabilité subjective', in CNRS (éditeur), *La décision*, pp. 73-97. Colloques internationaux du CNRS, Paris.
- DRÈZE, J. H. (1987) (sous la direction de), *Essays on Economic Decisions under Uncertainty*. Cambridge University Press, Cambridge.
- DRUZDZEL, M. J. et H. A. Simon (1993), 'Causality in Bayesian Belief Networks', in D. Heckerman et A. Mamdani (eds), *Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93)*. Morgan Kaufman, San Mateo, Californie.
- DUCOIN, J. (1993), *Naufrages, conditions de navigation et assurances dans la marine de commerce au XVIIIe siècle*. Librairie de l'Inde, Paris.
- DURKHEIM, E (1937), *Les règles de la méthode sociologique*. Presses universitaires de France, Paris.
- EASTMAN J. W. (1984), *Styling vs. Safety. The American Automobile Industry and the Development of Automotive Safety, 1900-1966*. University Press of America, Lanham.
- ECLAC (2003), *Handbook for Estimating the Socio-Economic and Environmental Effects of Disasters*. Commission économique des Nations Unies pour l'Amérique latine et les Caraïbes, Mexico et Banque Mondiale, Washington, D. C.
- ÉLECTRICITÉ DE FRANCE (1990), *Probabilistic Safety Assessment of reactor unit 3 in the Paluel Nuclear Power Centre (1300 MWe), Overall Report*. Électricité de France, Paris.
- ELLSBERG, D. (1961), 'Risk, Ambiguity, and the Savage Axioms', *Quarterly Journal of Economics*, 75(4), pp. 643-69.
- EPSTEIN, L. G. et M. Schneider (2003), 'Recursive multi-priors', *Journal of Economic Theory*, 113, pp. 1-31.
- EPSTEIN, L. G. et S. E. Zin (1989), 'Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns : A Theoretical Framework', *Econometrica*, 57(4), pp. 937-69.
- ERGIN, H. et F. Gul (2009), 'A Theory of Subjective Compound Lotteries', *Journal of Economic Theory*, 144, pp. 899-929.
- ESTÈVE, J. (2009), 'Is the "largest study ever conducted" on nuclear industry workers really the largest?', *Radioprotection*, 44(1), pp. 13-20.

- EWALD, F. (2002), 'Philosophie politique du principe de précaution', in F. Ewald, C. Gollier et N. de Sadeleer, *Le principe de précaution*. Presses universitaires de France, Paris.
- FABER, M. H. (2005), 'On the Treatment of Uncertainties and Probabilities in Engineering Decision Analysis', *Journal of Offshore Mechanics and Arctic Engineering*, 127, pp. 243-8.
- FARMER, F. R. (1967), 'Siting Criteria - A New Approach', *Proceedings of the IAEA symposium on nuclear siting*, pp. 303-29. Agence Internationale de l'Énergie Atomique, Vienne.
- FARMER, F. R., Fletcher, P. T. et T. M. Fry (1958), 'Safety Considerations for Gas-Cooled Thermal Reactors of the Calder Hall Type', *Proceedings of the Second International Conference on Peaceful Uses of Atomic Energy*, vol. XI, pp. 202-15. Organisation des Nations Unies, New York, 1959.
- FERSON, S. et L. R. Ginzburg (1996), 'Different Methods are Needed to Propagate Ignorance and Variability', *Reliability Engineering and System Safety*, 54, pp. 133-44.
- FERSON, S., Joslyn, C. A., Helton, Oberkampf, W. L. et K. Sentz (2004), 'Summary from the epistemic uncertainty workshop : consensus amid diversity', *Reliability Engineering and System Safety*, 85, pp. 355-69.
- FIELD, R. V. et M. Grigoriu (2007), 'Model Selection in Applied Science and Engineering : A Decision-Theoretic Approach', *Journal of Engineering Mechanics*, 133, pp. 780-91.
- FISHER, F. M. (1970), 'A correspondence principle for simultaneous equations models', *Econometrica*, 38, pp. 73-92.
- FISHER, E. (2010), 'Risk Regulatory Concepts and the Law', in OCDE, *Risk and Regulatory Policy: Improving the Governance of Risk*. Organisation pour la Coopération et le Développement Économiques, Paris.
- FISCHHOFF, B., Combs, B., Read, S., Lichtenstein, S. et P. Slovic (1978), 'How safe is safe enough? A psychometric study of attitudes towards technological risks and benefits', *Policy Sciences*, 9, pp. 127-52.
- FLEMING, K. N. (2003), *Issues and Recommendations for Advancement of PRA Technology in Risk-Informed Decision Making*, NUREG/CR-6813. Nuclear Regulatory Commission, Washington, D.C.
- FLEMING, K. N., A. Mosleh et A. P. Kelley (1983), 'Analysis of dependent failures in risk assessment and reliability evaluation', *Nuclear Safety*, 24(5), pp. 637-57.
- FLEURBAEY, M. (2010), 'Assessing Risky Social Situations', *Journal of Political Economy*, 118(4), pp. 649-80.

- FOASSO, C. (2003), *Histoire de la sûreté de l'énergie nucléaire civile en France (1945-2000)*. Thèse de doctorat en histoire moderne et contemporaine, Université Lumière-Lyon II.
- FRANCA, J. A. (1965), *Une ville des Lumières : la Lisbonne de Pombal*. SEVPEN, Paris.
- FREEMAN, P. K., L. A. Martin, R. Mechler, K. Warner et P. Hausman (2002), 'Catastrophes and Development : Integrating Natural Catastrophes into Development Planning', *Disaster Risk Management Working Paper Series*, 4. Banque Mondiale, Washington, D. C.
- FREIXAS, X., Giannini, C., Hoggarth, G. et F. Soussa (2000), 'Lender of Last Resort : What Have We Learned Since Bagehot', *Journal of Financial Services Research*, 18(1), pp. 63-84.
- FRIEDMAN, M. (1953), 'The effects of a full-employment policy on economic stability : A formal analysis', in M. Friedman (sous la direction de), *Essays in Positive Economics*. University of Chicago Press, Chicago.
- GAJDOS, T., Tallon, J.-M. et J.-C. Vergnaud (2008), 'Representation and aggregation of preferences under uncertainty', *Journal of Economic Theory*, 141(1), pp. 68-99.
- GAJDOS, T., Hayashi, T., Tallon, J.-M. et J.-C. Vergnaud (2008), 'Attitude toward imprecise information', *Journal of Economic Theory*, 140(1), pp. 23-56.
- GEIGER, D. et J. Pearl (1988), 'On the Logic of Causal Models', *Proceedings, 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, pp. 136-47. Réédité dans Shachter, R., T. S. Levitt et L. N. Kanal (sous la direction de), *Uncertainty in Artificial Intelligence*, vol. 4. Elsevier Science Publishers, 1990.
- GELLER, R. J. (2011), 'Shake-up time for Japanese seismology', *Nature*, 472, pp. 407-9.
- GHIRARDATO, P. (2001), 'Coping with Ignorance : Unforeseen Contingencies and Non-Additive Uncertainty', *Economic Theory*, 17, pp. 247-76.
- GIBBARD, A. et L. Harper (1978), 'Counterfactuals and two kinds of expected utility', in W. L. Harper, R. Stalnaker et G. Pearce (sous la direction de), *Ifs*. Reidel, Dordrecht.
- GILBOA, I. (2003), *Hempel, Good, and Bayes*. Manuscrit.
- GILBOA, I. (2009), *Theory of Decision under Uncertainty*. Cambridge University Press, Cambridge, Massachussets.
- GILBOA, I., D. Samet et D. Schmeidler (2004), 'Utilitarian Aggregation of Beliefs and Tastes', *Journal of Political Economy*, 112(4), pp. 932-8.

- GILBOA, I. et D. Schmeidler (1989), 'Maxmin Expected Utility with a Non- Unique Prior', *Journal of Mathematical Economics*, 18, pp. 141-53.
- GILBOA, I. et D. Schmeidler (2001), *A Theory of Case-Based Decisions*. Cambridge University Press, Cambridge, Massachussets.
- GILBOA, I. et D. Schmeidler (2010), 'Simplicity and likelihood : An axiomatic approach', *Journal of Economic Theory*, 145, pp. 1757-75.
- GODARD, O., C. Henry, P. Lagadec et E. Michel-Kerjan (2002), *Traité des Nouveaux Risques*. Gallimard, Paris.
- GOLDSCHMIDT, B. (1980), *Le complexe atomique*. Paris, Fayard.
- GOLDZMIDT, M. et J. Pearl (1992), 'Rank-based systems : a simple approach to belief revision, belief update and reasoning about evidence and actions', in B. Nebel, C. Rich et W. Swartout (sous la direction de), *Proceedings of the 3rd International Conference on Knowledge Representation and Reasoning*, pp. 661-72. Morgan Kauffman, San Mateo (Californie).
- GOLLIER, C. (2001), *The Economics of Risk and Time*, The MIT Press, Cambridge, Massachussets.
- GOLLIER, C., B. Jullien et N. Treich (2000), 'Scientific Progress and Irreversibility : An Economic Interpretation of the 'Precautionary Principle'', *Journal of Public Economics*, 75, pp. 229-53.
- GOOD, I. J. (1961), 'A causal calculus, I', *British Journal for the Philosophy of Science*, 11, pp. 305-18.
- GOOD, I. J. (1962), 'A causal calculus, II', *British Journal for the Philosophy of Science*, 12, pp. 43-51, addendum 13, p. 88.
- GOUVERNEMENT DU JAPON (2011), *Report of the Japanese Government to the IAEA Ministerial Conference on Nuclear Safety - The Accident at TEPCO's Fukushima Nuclear Power Stations*. Secrétariat du Cabinet du Premier Ministre du Japon, Tokyo.
- GOUVERNEMENT DU JAPON (2011b), *Additional Report of the Japanese Government to the IAEA - The Accident at TEPCO's Fukushima Nuclear Power Stations*. Nuclear Emergency Response Headquarters, Government of Japan, Tokyo.
- GOUVERNEMENT DU JAPON (2011c), *Road to Recovery*. Brochure d'information. Secrétariat du Cabinet du Premier Ministre du Japon, Tokyo.
- GRANDMONT, J.-M. (1972), 'Continuity Properties of a von Neumann-Morgenstern Utility', *Journal of Economic Theory*, 4, pp. 45-57.

- GRELOT, F., Poinard, D., Gendreau, N., Munier B. et E. Parent (2001), *Le risque d'inondation et la prise de décision publique*. Rapport final CEMAGREF, Paris.
- GROUPE PERMANENT RÉACTEURS (2008), *Avis relatif au bilan du réexamen de sûreté des réacteurs de 900 MWe dans le cadre de leurs troisièmes visites décennales*. Réunion du 20 novembre 2008, Groupe permanent d'experts pour les réacteurs nucléaires, Autorité de Sûreté Nucléaire.
- GROUPES PERMANENTS RÉACTEURS ET USINES (2011), *Avis relatif aux évaluations complémentaires de sûreté post-Fukushima réalisées en 2011 par les exploitants ÉdF, ILL, AREVA et CEA*. Réunions des 8, 9 et 10 novembre 2011, Groupes permanents d'experts pour les réacteurs nucléaires et pour les installations nucléaires de base autres que les réacteurs nucléaires à l'exception des installations destinées au stockage à long terme des déchets radioactifs, Autorité de Sûreté Nucléaire.
- GRNC (2002), *Analyse de sensibilité et d'incertitude sur le risque de leucémie attribuable aux installations nucléaires du Nord-Cotentin*. Rapport de synthèse du Groupe radioécologie Nord-Cotentin, Institut de radioprotection et de sûreté nucléaire.
- GUESNERIE, R. (2003), *Kyoto et l'Economie de l'Effet de Serre*. Rapport du Conseil d'Analyse Économique n° 39. La Documentation Française, Paris.
- HACKING, I. (1975), *The Emergence of Probability*. Cambridge University Press, New York.
- HACKING, I. (1990), *The Taming of Chance*. Cambridge University Press, New York.
- HALD, A. (1990), *A History of Probability and Statistics and Their Applications before 1750*. Wiley, New York.
- HALEVY, Y. (2007), 'Ellsberg revisited : an experimental study', *Econometrica*, 75, pp. 503-36.
- HALLEGATTE, S. (2008), 'An Adaptive Regional Input-Output Model and its Application to the Assessment of the Economic Cost of Katrina', *Risk Analysis*, 28(3), pp. 380-99.
- HALPERN, J. (2003), *Reasoning about uncertainty*. The MIT Press, Cambridge, Massachussetts.
- HAMMOND, P. J. (1976), 'Changing Tastes and Coherent Dynamic Choice', *Review of Economic Studies*, 43, pp. 159-73.
- HAMMOND, P. J. (1981), 'Ex-Ante and Ex-Post Welfare Optimality Under Uncertainty', *Economica*, 48, pp. 235-50.
- HAMMOND, P. J. (1983), 'Ex-Post Optimality As A Dynamically Consistent Objective for Collective Choice Under Uncertainty', in Pattanaik P. K. et M. Salles (sous la direction de), *Social Choice and Welfare*. North-Holland, Amsterdam.

- HAMMOND, P. J. (1988a), 'Consequentialism and the independence axiom', in B. Munier (sous la direction de), *Risk, Decision, and Rationality*. Reidel, Dordrecht.
- HAMMOND, P. J. (1988b), 'Consequentialist foundations for expected utility', *Theory and Decision*, 25, pp. 25-78.
- HAMMOND, P. J. (1989), 'Consistent plans, consequentialism, and expected utility', *Econometrica*, 57(6), pp. 1445-9.
- HANSEN, L. P. et T. J. Sargent (2000), *Wanting Robustness in Macroeconomics*. Manuscrit.
- HANSEN, L. P. et T. J. Sargent (2001), 'Robust Control and Model Uncertainty'. *American Economic Review*, 91, pp. 60-6.
- HELTON, J. C. (1994), 'Treatment of Uncertainty in Performance Assessments for Complex Systems', *Risk Analysis*, 14(4), pp. 483-511.
- HELTON, J. C., Johnson, J. D. et W. L. Oberkampf (2004), 'An Exploration of Alternative Approaches to the Representation of Uncertainty in Model Predictions', *Reliability Engineering and System Safety*, 85, pp. 39-71.
- HENRY, P. (2011), 'Le mégaséisme de subduction du Japon', *Pour la Science*, 403, pp. 34-9.
- HERMITTE, M.-A. (2007), 'La fondation juridique d'une société des sciences et des techniques par les crises et les risques', in C. Burton-Jeangros, C. Grosse et V. November, *Face au risque*. Éditions Georg – L'Equinoxe, Paris.
- HERP (1999), *The Promotion of Earthquake Research - Basic comprehensive policy for the promotion of earthquake observation, measurement, surveys and research*. Headquarters for Earthquake Research Promotion, Tokyo.
- HERP (2005), *National Seismic Hazard Maps for Japan*. Earthquake Research Committee, Headquarters for Earthquake Research Promotion, Tokyo.
- HICKS, J. R. (1931), 'The Theory of Uncertainty and Profit', *Economica*, 32, pp. 170-89.
- HILL, A.B. (1965), 'The environment and disease : association or causation?', *Proceedings of the Royal Society of Medicine*, 58, pp. 295-300.
- HILL, B. (2009), 'Confidence and ambiguity', *Cahiers de Recherche GREGHEC*, n° 914/2009. Groupement de Recherches en Gestion à HEC, Jouy-en-Josas.

- HOEVEN, Eugene (2005), 'Current state of the coverage for war and terrorism risks in the aviation sector', in OCDE (2005), *Catastrophic risks and insurance*. Policy issues in insurance, 8. Organisation pour la Coopération et le Développement Economiques, Paris.
- HOFER, E. (1996), 'When to Separate Uncertainties and When Not to Separate', *Reliability Engineering and System Safety*, 54(2-3), pp. 113-8.
- HOFFMAN, F. O. et J. S. Hammonds (1994), 'Propagation of Uncertainty in Risk Assessments : The Need to Distinguish Between Uncertainty Due to Lack of Knowledge and Uncertainty Due to Variability', *Risk Analysis*, 14(5), pp. 707-12.
- HOWARD, R. A. (1988), 'Uncertainty about Probability : A Decision Analysis Perspective', *Risk Analysis*, 8(1), pp. 91-8.
- HOWARD, R. A et J. E. Matheson (1981), 'Influence diagrams', in Howard, R. A. et J. E. Matheson (sous la direction de), *The Principles and Applications of Decision Analysis*, vol. II, 1984. Strategic Decisions Group, Menlo Park, Californie.
- HOWARD-SNYDER, F. (2007), 'Doing vs. Allowing Harm', *Stanford Encyclopedia of Philosophy*.
- HSE (1992), *The tolerability of risk from nuclear power stations*, revised version. Health and Safety Executive. HMSO, Londres.
- HSE (1999), *Reducing risks, protecting people - HSE's decision-making process*. Health and Safety Executive. HSE Books, Sudbury, Royaume-Uni.
- HUME, D. (1739), *A treatise of human nature*. Clarendon Press, Oxford, deuxième édition (1987).
- INPO (2011), *Special Report on the Nuclear Accident at the Fukushima Daiichi Nuclear Power Station*. Institute of Nuclear Power Operations, Atlanta.
- INSPECTION GÉNÉRALE DE L'ADMINISTRATION, Conseil général des ponts et chaussées, Conseil général du génie rural, des eaux et des forêts, et Inspection générale de l'environnement (2002), *Avis délibéré sur les retours d'expérience des inondations catastrophiques et les inspections des services déconcentrés en charge des risques naturels réalisés depuis l'année 1999*. Ministère de l'écologie, du développement durable, des transports et du logement.
- IPSN (1990), *EPS 900 - Etude probabiliste de sûreté des réacteurs à eau sous pression de 900 MWe - Introduction*. Rapport EPS 900/G/1, Institut de protection et de sûreté nucléaire, Commissariat à l'énergie atomique.

- IPSN (1990), *EPS 900 - Etude Probabiliste De Sûreté Des Réacteurs À Eau Sous Pression De 900 MWe - Événements Initiateurs*. Rapport EPS 900/G/3, Institut de protection et de sûreté nucléaire, Commissariat à l'énergie atomique.
- IPSN (1990), *EPS 900 - Etude Probabiliste De Sûreté Des Réacteurs À Eau Sous Pression De 900 MWe - Prise en compte du facteur humain*. Rapport EPS 900/G/5, Institut de protection et de sûreté nucléaire, Commissariat à l'énergie atomique.
- IPSN (2000), *Rapport sur l'inondation du site du Blayais survenue le 27 décembre 1999*. Institut de protection et de sûreté nucléaire.
- IRSN (2006), *EPS1/REP900 - Rapport de synthèse (conduite APE)*. Rapport DSR/SESPRI n°60, Institut de radioprotection et de sûreté nucléaire.
- IRSN (2007), *Protection des installations nucléaires contre les risques d'inondation externe*. Annexes au rapport DSR n°149, tome 2/5, Institut de radioprotection et de sûreté nucléaire.
- IRSN (2009), *EPS2/REP900 Version 4.0 - Rapport de synthèse. Tome 2 - Résultats en conséquences radiologiques*. Rapport DSR/SAGR n°41, Institut de radioprotection et de sûreté nucléaire.
- IRSN (2009b), *Etude des surcotes extrêmes le long du littoral français*. Rapport DEI/SARG/2009-018, Institut de radioprotection et de sûreté nucléaire.
- IRSN (2011a), 'L'IRSN a expertisé les démarches mises en œuvre par les exploitants pour réaliser les évaluations complémentaires de sûreté « post-Fukushima » des installations nucléaires françaises'. Communiqué du 6 juillet 2011, Institut de radioprotection et de sûreté nucléaire.
- IRSN (2011b), *Evaluations complémentaires de sûreté post-Fukushima : comportement des installations nucléaires françaises en cas de situations extrêmes et pertinence des propositions d'améliorations*. Rapport IRSN n° 708, Institut de radioprotection et de sûreté nucléaire.
- IRSN (2012), *L'économie de la sûreté*. Document de travail interne, Institut de radioprotection et de sûreté nucléaire.
- ISHIBASHI, K. (2007), 'Why Worry ? Japan's Nuclear Plants at Grave Risk From Quake Damage', *Asahi Shinbun* et *International Herald Tribune*, 11 août.
- ISO (2009), *Management du risque - Vocabulaire*. Guide ISO 73 :2009. Organisation internationale de normalisation, Genève.
- JSCE (2002), *Tsunami Assessment Method for Nuclear Power Plants in Japan*. Tsunami Evaluation Subcommittee, Nuclear Civil Engineering Committee, Japan Society of Civil Engineers, Tokyo.

- JEANNENEY, J.-M. (1981), 'La politique énergétique du gouvernement Debré', *Bulletin de l'Institut d'Histoire du Temps présent*, supplément n° 1.
- JEFFREY, R. C. (1965), *The Logic of Decision*. McGraw-Hill, New-York.
- JOYCE, J. (2007), 'Are Newcomb problems really decisions?', *Synthese*, 156, pp. 537–62.
- JOYCE, J. (2010), 'Causal reasoning and backtracking', *Philosophical Studies*, 147, pp. 139–54.
- KAHNEMAN, D., J. L. Knetsch et R. H. Thaler (1990), 'Experimental Test of the Endowment Effect and the Coase Theorem', *Journal of Political Economy*, 98(6), pp. 1325–48.
- KAHNEMAN, D. et A. Tversky (1979), 'Prospect theory : an analysis of decisions under risk', *Econometrica*, 2, pp. 263–91.
- KAPLAN, S. et B. J. Garrick (1981), 'On the Quantitative Definition of Risk', *Risk Analysis*, 1(1), pp. 11–27.
- KARNI, E. (1993), 'Subjective expected utility theory with state dependent preferences', *Journal of Economic Theory*, 60, pp. 428–38.
- KARNI, E. (2006), 'Subjective expected utility theory without states of the world', *Journal of Mathematical Economics*, 42(3), pp. 325–42.
- KARNI, E. et P. Mongin (2000), 'On the determination of subjective probability by choices', *Management Science*, 46, pp. 233–48.
- KARNI, E. et D. Schmeidler (1991), 'Atemporal dynamic consistency and expected utility theory', *Journal of Economic Theory*, 54, pp. 401–8.
- KARNI, E., D. Schmeidler et K. Vind (1983), 'On state dependent preferences and subjective probabilities', *Econometrica*, 51, pp. 1021–32.
- KEENEY, R. L. et H. Raiffa (1976), *Decisions with Multiple Objectives : Preferences and Value Trade-offs*. Wiley, New York.
- KELLER, W. et M. Modarres (2005), 'A historical overview of probabilistic risk assessment development and its use in the nuclear power industry : a tribute to the late Professor Norman Carl Rasmussen', *Reliability Engineering and System Safety*, 89, pp. 271–85.
- KEMENY, J. G. et al. (1979), *Report of the President's Commission on the Accident at Three Mile Island*. U.S. Government printing office, Washington, D.C.
- KEYNES, J. M. (1921), *A treatise on probability*. MacMillan, Londres.

- KIIVERI, H., T. P. Speed and J. B. Carlin (1984), 'Recursive causal models', *Journal of the Australian Mathematical Society*, 36, pp. 30-52.
- KLIBANOFF, P. (2001), 'Characterizing uncertainty aversion through preference for mixtures', *Social Choice and Welfare*, 18, pp. 289-301.
- KLIBANOFF, P., M. Marinacci et S. Mukerji (2005), 'A smooth model of decision making under ambiguity', *Econometrica*, 73, pp. 1849-92.
- KLIBANOFF, P., M. Marinacci et S. Mukerji (2009), 'Recursive smooth ambiguity preferences', *Journal of Economic Theory*, 144, pp. 930-76.
- KNIGHT, F. (1921), *Risk, uncertainty, and profit*. Houghton Mifflin, Boston.
- KOELZER, Winfried (2009), *Nuclear Glossary*. Forschungszentrum Karlsruhe GmbH, Karlsruhe.
- KOLM, S. C. (1998), 'Chance and Justice : Social Policies and the Harsanyi-Vickrey-Rawls Problem', *European Economic Review*, 42, pp. 1393-416.
- KREPS, D. M. et E. L. Porteus (1978), 'Temporal resolution of uncertainty and dynamic choice theory', *Econometrica*, 46, pp. 185-200.
- KOURILSKY, P. et G. Viney (2000), *Le Principe de précaution. Rapport au Premier Ministre*. Odile Jacob, Paris.
- KUNREUTHER, H. et E. Michel-Kerjan (2004), 'Policy Watch: Challenges for Terrorism Risk Insurance in the United States', *Journal of Economic Perspectives*, 18(4), pp. 201-14.
- KUROWICKA, D. et R. M. Cooke (2006), *Uncertainty Analysis with High Dimensional Dependence Modelling*. John Wiley and Sons, Chichester.
- LA BERGE, A. F. (1992), *Mission and method. The early nineteenth century French public health movement*. Cambridge University Press, Cambridge.
- LAHIDJI, R. (2010), 'Institutional and policy frameworks for climate change adaptation : Lessons from disaster risk management', contribution au Rapport spécial du Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC), *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation*, à paraître en 2012.
- LANGSETH, H. et L. Portinale (2007), 'Bayesian networks in reliability', *Reliability Engineering and System Safety*, 92(2), pp. 92-108.
- LAZARSFELD, P. F. (1961), 'Notes on the History of Quantification in Sociology - Trends, Sources and Problems', *Isis*, 52(2), pp. 277-333.

- LE GALL, J. (1953), *Le Tibre, fleuve de Rome, dans l'Antiquité*. Presses Universitaires de France, Paris.
- LE GOFF, J. (1962), *Marchands et banquiers au Moyen-Âge*. Presses Universitaires de France, Paris.
- LE PICHON, X. (2011), 'Erreurs de prévision et de gouvernance dans la gestion au Japon de la mitigation des séismes mises en lumière par le séisme de Tohoku du 11 mars 2011', in Académie des Sciences (2011), *Mégaséismes et mégatsunamis*. Rapport du sous-groupe sismique. Académie des sciences, Paris.
- LE PICHON, X., Mazzotti, S., Henry, P. et M. Hashimoto (1998), 'Deformation of the Japanese Islands and seismic coupling : an interpretation based on GSI permanent GPS observations', *Geophysical Journal International*, 134, pp. 501-4.
- LECOURT, D. (1999), 'Principe de causalité', in D. Lecourt (sous la direction de), *Dictionnaire d'histoire et de philosophie des sciences*, quatrième édition, 2006, pp. 173-7. Presses universitaires de France, Paris.
- LEIBNIZ, G. F. W. (1714), *Principes de la nature et de la grâce fondés en raison*, in *Œuvres*, éditions L. Prenant, Paris, 1972.
- LEWIS, D. (1973), *Counterfactuals*, Harvard University Press, Cambridge, Massachussetts.
- LEWIS, M. E. (2006), *The Flood Myths of Early China*. State University of New York Press, Albany.
- LEWIS, H. W. et al. (1978), *Risk Assessment Review Group Report to the U.S. Nuclear Regulatory Commission*, NUREG/CR-0400. Nuclear Regulatory Commission, Washington, D.C.
- LIBMANN, J. (1996), *Eléments de sûreté nucléaire*. Institut de protection et de sûreté nucléaire. Les éditions de physique, Paris.
- LINDLEY, D. V. (2000), 'The philosophy of statistics', *The Statistician*, 49(3), pp. 293-337.
- LUCAS, R. (1976), 'Econometric policy evaluation : a critique', in K. Brunner et A. Meltzer (sous la direction de), *The Phillips curve and labor markets*. Carnegie-Rochester conference series on public policy, American Elsevier, New York.
- MACCHERONI, F., Marinacci, M. et A. Rustichini (2006), 'Ambiguity Aversion, Robustness, and the Variational Representation of Preferences', *Econometrica*, 74, pp. 1447-98.
- MACHINA, M. J. (1989), 'Dynamic Consistency and Non-Expected Utility Models of Choice Under Uncertainty', *Journal of Economic Literature*, 27, pp. 1622-68.

- MACHINA, M. J. (2003), 'States of the World and the State of Decision Theory', in Meyer, D. (sous la direction de), *The Economics of Risk*. W.E. Upjohn Institute for Employment Research, Kalamazoo, Michigan.
- MACKIE, J. L. (1965), 'Causes and conditions', *American Philosophical Quarterly*, 2, pp. 245-64.
- MAJOR, J. A. (2002), *Advanced Techniques for Modeling Terrorism Risk*. Guy Carpenter & Co, New York.
- MARSCHAK, J. (1950), 'Statistical inference in economics', in T. Koopmans (sous la direction de), *Statistical inference in dynamic economic models*. Cowles Commission monograph n°10. Wiley, New York.
- MASSÉ, P. (1981), 'Autocritique des années soixante, vues par un Commissaire au Plan', *Bulletins de l'Institut d'Histoire du Temps Présent*, supplément n°1.
- MCCULLOUGH, R. C., Mills, M. M. et E. Teller (1955), 'The Safety of Nuclear Reactors', *Proceedings of the First International Conference on Peaceful Uses of Atomic Energy*, vol. XIII, pp. 79-87. Organisation des Nations Unies, New York, 1956.
- MECHLER, R., S. Hochrainer, J. Linnerooth-Bayer, J. Pflug (2006), Public sector financial vulnerability to disasters : The IIASA CATSIM model', in J. Birkmann (sous la direction de), *Measuring Vulnerability to Natural Hazards : Towards Disaster Resilient Societies*. United Nations University Press, Tokyo, pp. 380-99.
- MESERVE, R. (2001), 'The evolution of safety goals and their connection to safety culture'. Discours pour la réunion de l'Atomic Energy Society of Japan et de l'American Nuclear Society sur les objectifs et la culture de sûreté. Milwaukee, Wisconsin, 18 Juin. Accessible sur le site de la NRC (www.nrc.gov/reading-rm/doc-collections/commission/speeches/2001/s01-013.pdf, consulté le 4 novembre 2011).
- MINISTÈRE DE L'ÉCOLOGIE ET DU DÉVELOPPEMENT DURABLE et Ministère du transport, de l'équipement, du tourisme et de la mer (2005), *Plan de prévention des risques technologiques – Guide méthodologique*. Ministère de l'écologie, du développement durable, des transports et du logement.
- MINOURA, K., Imamura, F., Sugawara, D., Kono, Y. et T. Iwashita (2001), 'The 869 Jogan tsunami deposit and recurrence interval of large-scale tsunami on the Pacific coast of northeast Japan', *Journal of Natural Disaster Science*, 23(2), pp. 83-8.
- MIQUEL, J. (2006), *Hydrologie Statistique : Introduction à l'Etude des Processus Hydrométéorologiques et Application à la Prédétermination des Débits de Crues*. Cours d'hydrologie, chapitre 5. Ecole Nationale des Ponts et Chaussées, Paris.

- MISSION D'EXPERTISE SUR LES CRUES DE SEPTEMBRE 2002 (2003), *Retour d'expérience des crues de septembre 2002*. Ministère de l'écologie, du développement durable, des transports et du logement.
- MISSION D'INSPECTION SUR LA TEMPÊTE XYNTHIA (2010), *Tempête Xynthia : Retour d'expérience, évaluation et propositions d'action*. Ministère de l'écologie, du développement durable, des transports et du logement.
- MISSION D'INSPECTION SUR LES INONDATIONS DU VAR (2010), *Retour d'expérience des inondations survenues dans le Var les 15 et 16 juin 2010*. Ministère de l'écologie, du développement durable, des transports et du logement.
- MONGIN, P. (1995), 'Consistent Bayesian Aggregation', *Journal of Economic Theory*, 66, pp. 313-51.
- MONGIN, P. (1997), 'Spurious Unanimity and the Pareto Principle', présenté à la 6ème conférence de l'International Society for Utilitarian Studies, Nouvelle Orléans, mars 1997.
- MONGIN, P. et C. d'Aspremont (1998), 'Utility Theory and Ethics', in Barberà, S., P.J. Hammond et C. Seidl (éditeurs), *Handbook of Utility Theory*, volume 1 : Principles. Kluwer, Dordrecht.
- MORGAN, M. G. et M. Henrion (1990), *Uncertainty : A guide to dealing with uncertainty in quantitative risk and policy analysis*. Cambridge University Press, New York.
- MORRIS, W. E. (2009), Hume, *Stanford Encyclopedia of Philosophy*.
- NATIONAL RESEARCH COUNCIL (1983), *Risk assessment in the federal government : managing the process*. The National Academy Press, Washington DC.
- NATIONAL RESEARCH COUNCIL (1996), *Understanding Risk : Informing Decisions in a Democratic Society*. National Academy Press, Washington DC.
- NATIONAL RESEARCH COUNCIL (2008), *Department of Homeland Security Bioterrorism Risk Assessment: A Call for Change*. National Academy Press, Washington DC.
- NEAPOLITAN, R. E. (2004), *Learning Bayesian networks*. Prentice Hall, Upper Saddle River, New Jersey.
- NIETZSCHE, F. (1878), *Humain, trop humain*. Livre de poche, Paris.
- NOIVILLE, Ch. (2003), *Du bon gouvernement des risques*. Presses Universitaires de France - Les voies du droit, Paris.

- NOIVILLE, Ch. (2009), 'Science in Precautionary Measures. A synthesis of ECJ and WTO case law', rapport Co-Extra, in E. Truilhé-Marengo, *Expertise judiciaire en matière sanitaire et environnementale*, à paraître.
- NOTT, Jonathan (2006), *Extreme Events : A Physical Reconstruction and Risk Assessment*. The University Press, Cambridge.
- NOURBAKHS, H. P. et T. S. Kress (2005), 'Assessment of Phenomenological Uncertainties in Level 2 PRAs', *CSNI Workshop on Evaluation of Uncertainties in Relation to Severe Accidents and Level 2 PSA*. Agence pour l'Énergie Nucléaire.
- NOZICK, R. (1969), 'Newcomb's problem and two principles of choice', in N. Rescher (sous la direction de), *Essays in honor of Carl G. Hempel*. Reidel, Dordrecht.
- NRC (1975), *Reactor Safety Study : An Assessment of Accident Risks in Commercial Nuclear Power Plants*. WASH-1400, NUREG-75/014. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1979), 'Nuclear Regulatory Commission Issues Policy Statement on Reactor Safety Study and Review by the Lewis Panel'. *NRC Press Release*, 79-19. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1981), *Fault Tree Handbook*. NUREG-0492. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1983), *PRA Procedures Guide - A Guide to the Performance of Probabilistic Risk Assessments for Nuclear Power Plants*. NUREG/CR-2300. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1986), 'Safety Goals for the Operations of Nuclear Power Plants; Policy Statement'. 51 *Federal Register* 30028, 21 août 1986. US Government Printing Office, Washington, D.C.
- NRC (1988a), *Procedures for Treating Common-cause Failures in Safety and Reliability Studies*. NUREG/CR-4780. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1988b), 'Individual Plant Examination for Severe Accident Vulnerabilities - 10CFR50.54(f)', *Generic Letter 88-20*. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1988c), 'Individual Plant Examination of External Events for Severe Accident Vulnerabilities - 10CFR50.54(f)', *Generic Letter 88-20, Supplement 4*. Nuclear Regulatory Commission, Washington, D.C.
- NRC (1990), *Severe Accident Risks : An Assessment for Five U.S. Nuclear Power Plants*, NUREG-1150. Nuclear Regulatory Commission, Washington, D.C.

- NRC (1995), 'Use of Probabilistic Risk Assessment Methods in Nuclear Regulatory Activities ; Policy Statement'. *60 Federal Register 42622*, 26 août 1995. US Government Printing Office, Washington, D.C.
- NRC (1998), *An Approach for Using Probabilistic Risk Assessment in Risk-Informed Decisions on Plant- Specific Changes to the Licensing Basis, Regulatory Guide 1.174* (révisé en 2002 et 2009). Nuclear Regulatory Commission, Washington, D.C.
- NRC (2007), *An Approach for Determining the Technical Adequacy of Probabilistic Risk Assessment Results for Risk- Informed Activities, Regulatory Guide 1.200*. Nuclear Regulatory Commission, Washington, D.C.
- NRC (2009), *Guidance on the Treatment of Uncertainties Associated with PRAs in Risk-Informed Decision Making*. Nuclear Regulatory Commission, Washington, D.C.
- NRC (2011), *Recommendations for Enhancing Reactor Safety in the 21st Century - The Near-Term Task Force Review of Insights from the Fukushima Dai-Ichi Accident*. Nuclear Regulatory Commission, Washington, D.C.
- NRC (2011b), *Briefing on the Task Force Review of NRC Processes and Regulations Following the Events, Japan*. Retranscription des discussions lors de la réunion publique du 19 juillet 2011. Nuclear Regulatory Commission, Washington, D.C.
- NSC (2006), *Regulatory Guide for Reviewing Seismic Design of Nuclear Power Reactor Facilities*. Nuclear Safety Commission, Tokyo.
- NSC (2009), *2007–2008 Nuclear Safety White Paper*. Nuclear Safety Commission, Tokyo.
- OCDE (2009), *OECD Reviews of Risk Management Policies : Japan - Large-Scale Floods and Earthquakes*. Organisation pour la Coopération et le Développement Économiques, Paris.
- OKRENT, D. (1981), *Nuclear Reactor Safety : On the History of the Regulatory Process*. University of Wisconsin Press, Madison.
- PARKER, H. M. et J. W. Healey (1955), 'Environmental Effects of a Major Reactor Disaster', *Proceedings of the First International Conference on Peaceful Uses of Atomic Energy*, vol. XIII, pp. 106-9. Organisation des Nations Unies, New York, 1956.
- PARRY, M. L., Canziani, O. F., Palutikof, J. P., van der Linden, P. J. et C. E. Hanson (sous la direction de) (2007), *Rapport du Groupe de travail II - Conséquences, adaptation et vulnérabilité*. Quatrième rapport d'évaluation du Groupe d'experts intergouvernemental sur l'évolution du climat. Cambridge University Press, Cambridge, Royaume-Uni.

- PATÉ-CORNELL, M. E. (1996), 'Uncertainties in Risk Analysis : Six Levels of Treatment'. *Reliability Engineering and System Safety*, 54(2-3), pp. 95-111.
- PEARL, J. (1988), *Bayesian Networks*. Cambridge University Press, New York.
- PEARL, J. (1993), 'Comment: Graphical models, causality, and intervention'. *Statistical Science*, 8, pp. 266-9.
- PEARL, J. (2000), *Causality*. Cambridge University Press, New York.
- PEARSON, K. (1914-30), *The Life, Letters and Labours of Francis Galton*. 4 tomes, Cambridge University Press, Cambridge.
- PEARL, J. et T. Verma (1991), 'A theory of inferred causation', in *Principles of Knowledge Representation and Reasoning: Proceedings of the 2nd international conference*. Morgan Kauffman, San Mateo, Californie.
- PEDUZZI, P., H. Dao, C. Herold, et F. Mouton (2009), 'Assessing global exposure and vulnerability towards natural hazards: the Disaster Risk Index', *Natural Hazards and Earth System Sciences*, 9, pp. 1149-59.
- PHILIPSON, T. J. et E. Sun (2008), 'Is the Food and Drug Administration Safe and Effective?', *Journal of Economic Perspectives*, 22(1), p. 85-102.
- PIDGEON, N., R. E. Kasperson et P. Slovic (sous la direction de) (2003), *The Social Amplification of Risk*. Cambridge University Press, Cambridge.
- PIEGORSCH, W. W., Cutter, S. L. et F. Hardisty (2007), 'Benchmark Analysis for Quantifying Urban Vulnerability to Terrorist Incidents', *Risk Analysis*, 27 (6), pp. 1411-25.
- PNUD (2004), *La Réduction des Risques de Catastrophe: Un Défi pour le Développement*. Bureau pour la prévention des crises et le relèvement, Programme des Nations Unies pour le développement, New York.
- PRADIER P.-C. (1998), *Concepts et mesures du risque en théorie économique — essai historique et critique*. Thèse de doctorat, Ecole Normale Supérieure de Cachan.
- PUISEUX, L. (1982), 'Les bifurcations de la politique énergétique française depuis la guerre', *Annales. Économies, Sociétés, Civilisations*, 4, pp. 609-20.
- QUENET, G. (2005), *Les tremblements de terre aux XVIIe et XVIIIe siècles — La naissance d'un risque*. Champ Vallon, Seyssel.

- QUIGGIN, J. (1982), 'A Theory of Anticipated Utility', *Journal of Economic Behaviour and Organization*, 3, pp. 323-43.
- RAFFENSPERGER, C. et J. Tickner (sous la direction de) (1999), *Protecting Public Health and the Environment: Implementing the Precautionary Principle*. Island Press, Washington, D.C.
- RAIFFA, H. (1968), *Decision Analysis*. Addison-Wesley, Reading (Massachusetts).
- RAMADE, F. (1993), *Dictionnaire Encyclopédique de l'Ecologie et des Sciences de l'Environnement*. Ediscience International, Paris.
- RAMSEY, F. (1926), 'Truth and Probability', in R.B. Braithwaite (sous la direction de), *The Foundations of Mathematics and other Logical Essays*, 1931. Kegan, Paul, Trench, Trubner & co., Londres.
- REICHENBACH, H. (1956), *The direction of time*. University of California Press, Berkeley.
- RIDDEL, M. et W. D. Shaw (2006), 'A theoretically-consistent empirical model of non-expected utility: An application to nuclear-waste transport', *Journal of Risk and Uncertainty*, 32, pp.131-50.
- RMS (2011), *Terrorism Risk in the Post-9/11 Era: A 10-Year Retrospective*. Risk Management Solutions, Newark.
- ROGOVIN et al. (1980), 'Three Mile Island: A Report to the Commissioners and to the Public', NUREG/CR-1250. NRC, Washington, D.C.
- ROTHMAN, K. J. (1976), 'Causes', *American Journal of Epidemiology*, 104, pp. 587-92.
- ROTHMAN, K. J., S. Greenland et T. L. Lash (2008), *Modern Epidemiology*, troisième édition, Lippincott, Williams et Wilkins, Philadelphie.
- ROTHSCHILD, M. et J. E. Stiglitz (1970), 'Increasing Risk I : A Definition', *Journal of Economic Theory*, 2(3), pp. 225-43.
- ROUSSEAU, J.-J. (1756), *Lettre sur la Providence*.
- SAREWITZ, D., Pielke, R. et M. Keykhah (2003), 'Vulnerability and risk: some thoughts from a political and policy perspective', *Risk Analysis*, 23, pp. 805-810.
- SARIN, R. et P. Wakker (1994), 'Folding back in decision tree analysis', *Management Science*, 40, pp. 625-8.
- SAVAGE, L. (1954), *The foundations of statistics*. Dover Publications, New York.

- SCHELLING, T. (1960), *The strategy of conflict*. Harvard University Press, Cambridge.
- SCHERVISH, M. J., T. Seidenfeld et J. B. Kadane (1990), 'State-Dependent Utilities', *Journal of the American Statistical Association*, 85, pp. 840-7.
- SCHMEIDLER, D. (1989), 'Subjective probability and expected utility without additivity', *Econometrica*, 57, pp. 571-87.
- SCSIN (1977), *Lettre d'orientation SIN n°1076/77 du 11 juillet 1977*. Service Central de Sûreté des Installations Nucléaires, Ministère chargé de l'Industrie. Accessible sur le site de l'ASN (www.asn.fr, consulté le 31 octobre 2011).
- SCSIN (1978), *Lettre d'orientation SIN n°576/78 du 16 mars 1978*. Service Central de Sûreté des Installations Nucléaires, Ministère chargé de l'Industrie. Accessible sur le site de l'ASN (www.asn.fr, consulté le 31 octobre 2011).
- SCSIN (1984), *Règle fondamentale de sûreté n°I.2.e relative à la prise en compte du risque d'inondation d'origine externe*. Service Central de Sûreté des Installations Nucléaires, Ministère chargé de l'Industrie.
- SEGAL, U. (1990), 'Two-stage lotteries without the reduction axiom', *Econometrica*, 58, pp. 349-78.
- SEGAL, U. (1987), 'The Ellsberg paradox and risk aversion : an anticipated utility approach', *International Economic Review*, 28, pp. 175-202.
- SEO, K. (2009), 'Ambiguity and second-order belief', *Econometrica*, 77, pp. 1575-1605.
- SERQUIN, Y. (1998), *Gestion scientifique de la maintenance des grands systèmes. L'apport de l'aide à la décision par utilité multi-attribut généralisée*. Thèse de doctorat, Ecole Normale Supérieure de Cachan.
- SHACHTER, R. (1986), 'Evaluating influence diagrams', *Operations Research*, 34, pp. 871-82.
- SHACKLE, G. L. S. (1961), *Decision, Order and Time in Human Affairs*. Cambridge University Press, Cambridge.
- SHAFER, G. (1976), *A mathematical Theory of Evidence*. Princeton University Press, Princeton, New Jersey.
- SHAFER, G. (1986), 'Savage Revisited', *Statistical Science*, 1, pp. 463-86.
- SHAFER, G. (1996), *The Art of Causal Conjecture*. The MIT Press, Cambridge, Massachusetts.
- SHRADER-FRECHETTE, K. S. (1991), *Risk and Rationality - Philosophical Foundations for Populist Reforms*. University of California Press, Berkeley, Californie.

- SIDDAL, E. (1957), 'Safe and Reliable Reactor Protection', *Nucleonics*, pp. 124-9.
- SIMON, H. A. (1953), 'Causal ordering and identifiability', in W. C. Hood et T. Koopmans (sous la direction de), *Studies in econometric method*. Wiley, New York.
- SINNOTT-ARMSTRONG, W. (2006), Consequentialism, *Stanford Encyclopedia of Philosophy*.
- SKIADAS, C. (1998), 'Recursive utility and preferences for information', *Economic Theory*, 12, pp. 293-312.
- SKYRMS, B. (1980), *Causal necessity - A pragmatic investigation of the necessity of laws*. Yale University Press, New Haven.
- SLOVIC, P. (éditeur) (2000), *The Perception of Risk*. Earthscan, Londres.
- SMITH, C. A. B. (1961), 'Consistency in statistical inference and decision', *Journal of the Royal Statistical Society, série B*, 23, pp. 1-25.
- SOBEL, M. E. (1990), 'Effect analysis and causation in linear structural equation models', *Psychometrika*, 55, pp. 495-515.
- SORENSEN, J. N., Apostolakis, G. E., Kress, T. S. et D. A. Powers (1999), 'On the role of defence in depth in risk-informed regulation', in M. Modarres (sous la direction de), *PSA '99: Proceedings of the International Topical Meeting on Probabilistic Safety Assessment*. American Nuclear Society, La Grange Park, Illinois, pp. 408-13.
- SPIRITES, P., C. Glymour et R. Scheines (1993), *Causation, Prediction, and Search*. The MIT Press, Cambridge, Massachussets.
- STALNAKER, R. C. (1968), 'A theory of conditionals', in N. Rescher (sous la direction de), *Studies in logical theory*. Blackwell, Oxford. Réédité dans W. L. Harper, R. Stalnaker et G. Pearce (sous la direction de), *Ifs*. Reidel, Dordrecht.
- STARR, C. (1969), 'Social Benefit versus Technological Risk', *Science*, 165, pp. 1232-8.
- STEVENSON, J. D. (2003), 'Historical Development of the Seismic Requirements for Construction of Nuclear Power Plants in the U.S. and Worldwide and their Current Impact on Cost and Safety', *Transactions of the 17th International Conference on Structural Mechanics in Reactor Technology (SMiRT 17)*, Prague, République Tchèque, 17-22 août.
- STIGLER, S.M. (1986), *The History of Statistics. The Measurement of Uncertainty before 1900*. The Belknap Press of Harvard University Press, Cambridge.

- STOHL, A., P. Seibert, G. Wotawa, D. Arnold, J. F. Burkhart, S. Eckhardt, C. Tapia, A. Vargas et T. J. Yusunari (2011), 'Xenon-133 and caesium-137 releases into the atmosphere from the Fukushima Dai-ichi nuclear power plant : determination of the source term, atmospheric dispersion, and deposition', *Atmospheric Chemistry and Physics Discussions*, 11, pp. 28319-94.
- STRATÉGIE INTERNATIONALE DE RÉDUCTION DES CATASTROPHES (2009), *Réduction des risques de catastrophe : bilan mondial 2009 - Risques et pauvreté dans un climat en évolution*. Secrétariat de la Stratégie Internationale de Réduction des Catastrophes, Nations Unies, Genève.
- STROTZ, R. H. et H. O. A. Wold (1960), 'Recursive versus nonrecursive systems: An attempt at synthesis', *Econometrica*, 28, pp. 417-27.
- STRZALECKI, T. (2010), 'Temporal Resolution of Uncertainty and Recursive Models of Ambiguity Aversion', mimeo, Harvard University.
- SUPPES, P. (1970), *A probabilistic theory of causality*. North Holland, Amsterdam.
- SWAIN, A. D. et H. E. Guttman (1983), *Handbook of human reliability with emphasis on nuclear power plant application*. NUREG/CR-1278. Nuclear Regulatory Commission, Washington, D.C.
- TANGUY, P. (1988), 'Three decades of nuclear safety : Nuclear plant safety has not been a static concept', *IAEA Bulletin*, 2, pp. 51-7.
- TANGUY, P. et F. Cogné (1978), 'Utilisation des méthodes probabilistes dans l'évaluation de sûreté des installations nucléaires', Présentation orale à la conférence Nuclex (Bâle, 3-7 octobre), Rapport DSN n° 238 (f).
- TELLER, E. (1953), *Letter to Sterling Cole, United States Congressman, on Dangers and Military Applications of Nuclear Energy*. July 23, 1953. Accessible sur le site nuclearfiles.org (consulté le 26 octobre 2011).
- TEPCO (2011), *Fukushima Daiichi Nuclear Power Station Unit No. 1 - Report on Earthquake Response Analysis of the Reactor Building, Important Equipment and Piping System for Seismic Safety using Recorded Seismic Data of the Tohoku-Chihou-Taiheiyo-Oki Earthquake in the Year 2011*. Tokyo Electric Power Company, Tokyo.
- THOM, R. (1980), 'Halte au hasard, silence au bruit', *Le Débat*, 3.
- THORNTON, H. (1802), *An Enquiry into the Nature and Effects of the Paper Credit of Great Britain*. Édition de 1991. Kelley, New York.
- TREICH, N. (2010), 'The Value of a Statistical Life under Ambiguity Aversion', *Journal of Environmental Economics and Management*, 59(1), pp. 15-26.

- TUNC, A. (1990), *La responsabilité civile*. Deuxième édition, Economica, Paris.
- TVERSKY, A. et D. Kahneman (1973), 'Availability: a heuristic for judging frequency and probability', *Cognitive Psychology*, 5, pp. 207-232.
- TVERSKY, A. et D. Kahneman (1974), 'Judgment under uncertainty: Heuristics and biases', *Science*, 185, pp. 1124-31.
- TVERSKY, A. et D. Kahneman (1981), 'The Framing of Decisions and the Psychology of Choice', *Science*, 211, pp. 453-8.
- UNDRO (1979), *Natural disasters and vulnerability analysis*. Rapport du groupe d'experts, Bureau du coordonnateur des nations unies pour les secours en cas de catastrophe, Genève.
- UNISDR (2009), *Terminologie pour la prévention des risques de catastrophe*. Stratégie internationale de prévention des catastrophes des Nations Unies (UNISDR), Genève.
- VAN OORSCHOT, M. M. P. Kraan, B. C. P. van den Brink, R. M. M. Janssen, P. H. M. et R. M. Cooke (2003), *Uncertainty Analysis for NO_x Emissions from Dutch Passenger Cars in 1998*. RIVM report 550002004/2003, Université technique de Delft.
- VERMA, T. et J. Pearl (1987), 'Causal Networks: Semantics and Expressiveness', *Proceedings, 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, pp. 352-59. Réédité dans Shachter, R., T. S. Levitt et L. N. Kanal (sous la direction de), *Uncertainty in Artificial Intelligence*, vol. 4. Elsevier Science Publishers, Amsterdam.
- VERMA, T. et J. Pearl (1990), 'Equivalence and Synthesis of Causal Models', *Proceedings, 6th Conference on Uncertainty in Artificial Intelligence*, Cambridge, MA, pp. 220-7. Réédité dans Bonissone, P., M. Henrion, L. N. Kanal et J. F. Lemmer (sous la direction de), *Uncertainty in Artificial Intelligence*, vol. 6. Elsevier Science Publishers, Amsterdam.
- VERNE, J. (1873), *Le tour du monde en quatre-vingts jours*. Gallimard, Paris.
- VESELY, W. E. et D. M. Rasmuson (1984), 'Uncertainties in Nuclear Probabilistic Risk Analyses', *Risk Analysis*, 4(4), pp. 313-22.
- VIGDOR, J. (2008), 'The Economic Aftermath of Hurricane Katrina', *Journal of Economic Perspectives*, 22(4), pp. 135-54.
- VILLEMEUR, A. (1988), *Sûreté de fonctionnement des systèmes industriels. Fiabilité - Facteurs humains - Informatisation*. Éditions Eyrolles, Paris.
- VISCUSI, W. K. (1998), *Rational Risk Policy*. Oxford University Press.

- VISCUSI, W. K. (2007), 'Regulation of Health, Safety and Environmental Risks', in A. M. Polinsky et S. Shavell (sous la direction de), *Handbook of Law and Economics*, Vol. 1, Elsevier, North-Holland.
- VISCUSI, W. K., Magat, W. A. et J. Hubert (1991), 'Communication of ambiguous risk information', *Theory and Decision*, 31, pp.159-73.
- VOIDEY, N. (2005), *Le risque en droit civil*. Presses Universitaires d'Aix-Marseille. Texte remanié de *Le risque en droit civil*, thèse de doctorat en droit, Université de Strasbourg 3, 2003.
- VOLTAIRE (François Marie Arouet, dit) (1756), *Poème sur le désastre de Lisbonne*.
- VON KRIES, J. (1886), *Die Principien der Wahrscheinlichkeits-Rechnung: Eine logische Untersuchung*. Mohr, Tübingen.
- VON NEUMANN, J. et O. Morgenstern (1944), *Theory of Games and Economic Behavior*. Princeton University Press, Princeton.
- WALKER, J. S. (1992), *Containing the Atom: Nuclear Regulation in a Changing Environment, 1963-1971*. University of California Press, Berkeley.
- WALKER, J. S. (2000), *A Short History of Nuclear Regulation, 1946 -1999*. Rapport BR 0175, Nuclear Regulatory Commission, Washington D. C.
- WALLEY, P. (1991), *Statistical reasoning with imprecise probabilities*. Chapman and Hall, New York.
- WILLIS, H. H., Morral, A. R., Kelly, T. K. et J. J. Medby (2005), *Estimating Terrorism Risk*. Rand Center for Terrorism Risk Management Policy, the Rand Corporation, Santa Monica.
- WINKELSTEIN, W. (2004), 'Vignettes of the history of epidemiology: Three firsts by Janet Elizabeth Lane-Claypon', *American Journal of Epidemiology*, 160(2), pp. 97-101.
- WINKLER, R. L. (1994), 'Model uncertainty: probabilities for models?', in A. Mosleh et al. (sous la direction de), *Model Uncertainty: its characterization and quantification*, NUREG/CP-OI38, pp.107-16. Nuclear Regulatory Commission, Washington, D.C.
- WINKLER, R. L. (1996), 'Uncertainty in probabilistic risk assessment', *Reliability Engineering and System Safety*, 54(2-3), pp. 127-132.
- WOODWARD, J. (2003), *Making Things Happen - A Theory of Causal Explanation*. Oxford University Press, New York.
- WOODWARD, J. (2008), Causation and manipulability, *Stanford Encyclopedia of Philosophy*.

- ZADEH, L. (1978), 'Fuzzy sets as a basis for a theory of possibility', *Fuzzy Sets and Systems*, 1, pp. 3–28.

Glossaire

Accident de criticité : Accident nucléaire provoqué par une réaction en chaîne incontrôlée et une augmentation exponentielle de la puissance au sein d'un combustible.

Accident enveloppe : Accident le plus grave parmi une famille d'événements considérés pour le dimensionnement d'un aspect particulier d'un réacteur.

Agression externe : Événement extérieur à l'installation affectant son intégrité.

Agression interne : Événement intérieur à l'installation affectant l'intégrité de l'un de ses systèmes ou composantes.

Approche par états : Mode de conduite d'une tranche en situation accidentelle qui se fonde non pas sur l'identification des événements singuliers survenus dans l'installation, mais sur l'état général de celle-ci, défini à partir de grandeurs caractérisant le risque et la marche à suivre.

Arbre d'événements : Arbre de probabilités représentant les séquences accidentelles déclenchées par un événement initiateur, et dont chaque noeud porte deux branches représentant le succès et l'échec d'une mission.

Dénomination équivalente : arbre de conséquences

Arbre de défaillance : Diagramme booléen retraçant les conditions de défaillance d'un système à partir du comportement de ses composantes et d'éventuelles erreurs humaines.

Dénomination équivalente : arbre de causes

Arbre de déroulement des accidents graves : Arbre de probabilités représentant la progression d'un accident grave depuis la fusion du coeur jusqu'à ses conséquences en termes de rejets de matière radioactive en dehors de l'enceinte de confinement.

Equivalent anglais : containment tree

Conditions de fonctionnement : Catégories d'événements pris en compte dans l'élaboration des niveaux de la défense en profondeur. Initialement au nombre de quatre (fonctionnement normal, incidents, accidents, accidents graves mais maîtrisés), les conditions de fonctionnement ont été augmentées d'une cinquième catégorie pour les défaillances multiples et les accidents

très graves au cours des années 1970.

Dénomination équivalente : états de l'installation

Corium : Amas de combustible et d'éléments de structure fondus provenant de la fusion du coeur d'un réacteur nucléaire.

Dimensionnement : Détermination des caractéristiques d'une installation lors de sa conception afin de répondre à certains critères de sûreté préétablis.

Défaillance primaire : Défaillance d'une composante ou d'un système alors qu'il est placé dans des conditions de fonctionnement pour lesquelles il est normalement qualifié.

Défaillance secondaire : Défaillance d'une composante ou d'un système alors qu'il est placé dans des conditions de fonctionnement pour lesquelles il n'est pas qualifié.

Défense en profondeur : Ensemble de dispositions organisées en lignes successives et destinées à prévenir un accident nucléaire ou à atténuer ses conséquences. Initialement dotée de trois niveaux, la défense en profondeur a été remaniée après l'accident de Three Mile Island et en compte désormais cinq, dont une extérieure à l'installation.

Effet falaise : Situation où une faible variation d'une variable exogène ou d'un paramètre se traduit par une réponse disproportionnée du système considéré.

Fluide primaire : Liquide dont la circulation assure le refroidissement des éléments de combustible. Dans un réacteur à eau pressurisée, il s'agit d'eau borée.

Fonctions de sûreté : Fonctions dont le maintien permanent conditionne la sûreté d'une installation. Au niveau le plus fondamental, il y a trois fonctions de sûreté dans un réacteur : la maîtrise de la réaction en chaîne, le refroidissement des éléments de combustible et le confinement de la matière radioactive.

Ilotage : Séparation d'une tranche des réseaux électriques extérieurs, qui permet de consacrer l'énergie produite à l'alimentation de ses propres systèmes auxiliaires.

Initiateur : Evénement considéré comme le premier élément d'une séquence accidentelle.

Mitigation : Mesures mises en place afin d'atténuer les dommages causés par un aléa après que celui-ci s'est produit ou alors que son occurrence est imminente.

Radier : Revêtement en béton protégeant la base du bâtiment du réacteur dans une centrale nucléaire.

Terme source : Caractérisation des rejets de substances radioactives par une installation nucléaire, selon la quantité, la composition isotopique et la cinétique de la matière émise, dans des

conditions normales de fonctionnement ou dans un scénario d'accident grave avec ruine de l'enceinte de confinement.

Transitoires : Déviation des paramètres d'opération d'une centrale nucléaire (tels que la puissance, la pression, la température ou la circulation du réfrigérant) de leurs valeurs centrales qui conduit à un déséquilibre entre la chaleur produite dans le réacteur et celle qui en est extraite. On distingue les transitoires normaux (démarrages, montée en puissance, variations de charge) et les transitoires accidentels (consécutifs à des défaillances de matériels ou des erreurs humaines).

Table des matières

Remerciements	1
Organisation générale	2
Introduction générale	3
 I Une problématique du risque social	 17
1 Examen du concept de risque social	18
1.1 Introduction	18
1.2 Survol des domaines concernés	19
1.3 Proposition de synthèse	33
1.4 Conclusion	38
 2 L'action publique face au risque social	 39
2.1 Introduction	39
2.2 La société de la réglementation du risque ?	41
2.3 Décisions publiques ou collectives ? L'exemple des inondations	47
2.4 Le point de vue normatif	53
2.5 Conclusion : la nécessité d'un cadre d'analyse	58
 II Risques et sûreté nucléaires	 60
 3 L'analyse de la sûreté des centrales nucléaires	 61
3.1 Introduction	61
3.2 Préliminaires : risques et sûreté dans une centrale nucléaire	63
3.3 L'approche déterministe de la sûreté	66
3.4 L'approche probabiliste de la sûreté	80

3.5	Caractérisation des deux approches	110
3.6	Conclusion	116
4	Les politiques de sûreté nucléaire	118
4.1	Les décisions initiales (1945-1967)	119
4.2	La politique de la sûreté nucléaire aux États-Unis	132
4.3	La politique de la sûreté nucléaire en France	143
4.4	Conclusion	156
III	Une approche causale du risque	160
5	Une représentation par les réseaux bayésiens causaux	161
5.1	Introduction	161
5.2	Les réseaux bayésiens causaux	162
5.3	Une application à la sûreté	181
5.4	Conclusion	197
6	Les leçons de Fukushima	198
6.1	Introduction	198
6.2	Chronique des événements	199
6.3	Les causes d'une catastrophe	209
6.4	Après Fukushima	228
6.5	Conclusion	240
IV	Une approche causale de la décision en incertitude	242
7	Introduction aux réseaux décisionnels : un cas d'école	243
7.1	Introduction	243
7.2	Un problème de gestion du risque et son traitement causal	245
7.3	Analyse de quelques problèmes traditionnels	261
7.4	Conclusion	272
8	Réseaux décisionnels : cadre et résultats théoriques	273
8.1	Introduction	273
8.2	Un modèle de décision causale	275
8.3	Articulation avec la théorie de la décision	287
8.4	Décision dans l'ambiguïté	292

<i>TABLE DES MATIÈRES</i>	340
8.5 Conclusion	300
Conclusion générale	302
Bibliographie	304
Glossaire	335

Résumé

Incertitude, causalité et décision : le cas des risques sociaux et du risque nucléaire en particulier

La probabilité et la causalité sont deux outils indispensables à la prise en compte des situations de risque social. Les relations causales sont le fondement des représentations à partir desquelles on peut évaluer le risque et concevoir des actions de prévention, de mitigation ou d'indemnisation. La probabilité permet de quantifier cette évaluation et de calibrer ces actions. Dès lors, il semble non seulement naturel, mais nécessaire d'explicitier la place de la causalité et de la probabilité dans la définition d'un problème de décision en situation de risque social. C'est l'objet de cette thèse.

Un tour d'horizon de la terminologie du risque et des logiques d'intervention publique dans différentes catégories de risque social nous permettent de mieux comprendre la notion et les problèmes soulevés par sa représentation. Nous approfondissons notre analyse dans le cas de la sûreté nucléaire, en examinant en détail les méthodes et doctrines développées dans ce domaine et leur évolution au cours du temps, ce qui nous conduit à formuler différentes observations au sujet des évaluations de risque et de sûreté.

En généralisant la notion d'intervention dans les réseaux bayésiens, nous développons une forme de réseau bayésien causal qui répond à nos besoins. Nous parvenons, par son biais, à une définition du risque qui semble pertinente pour un grand nombre de situations. Nous proposons ensuite des applications simples de ce modèle à certains aspects de l'accident de Fukushima et d'autres problèmes de sûreté nucléaire. Outre certains enseignements spécifiques, ceci nous amène à souligner la nécessité d'une démarche systématique d'identification des incertitudes dans ce domaine.

Étendu en direction de la théorie de la décision, notre outil débouche naturellement sur un modèle de décision dynamique dans lequel les actes *causent* les conséquences et sont causalement liés entre eux. Il apporte en outre une interprétation causale au cadre conceptuel de Savage et permet d'en résoudre certains paradoxes et clarifier certains aspects. Il conduit enfin à envisager la question de l'ambiguïté comme incertitude concernant la structure causale d'un problème de décision, ce qui correspond à une vision courante du principe de précaution.

Mots clés : Risque, évaluation du risque, gestion du risque, causalité, réseaux bayésiens, ambiguïté, théorie de la décision, Fukushima, sûreté nucléaire, défense en profondeur

Summary

Uncertainty, causality and decision : the case of social risks and nuclear risk in particular

Probability and causality are two indispensable tools for addressing situations of social risk. Causal relations are the foundation for building risk assessment models and identifying risk prevention, mitigation and compensation measures. Probability enables us to quantify risk assessments and to calibrate intervention measures. It therefore seems not only natural, but also necessary to make the role of causality and probability explicit in the definition of decision problems in situations of social risk. Such is the aim of this thesis.

By reviewing the terminology of risk and the logic of public interventions in various fields of social risk, we gain a better understanding of the notion and of the issues that one faces when trying to model it. We further elaborate our analysis in the case of nuclear safety, examining in detail how methods and policies have been developed in this field and how they have evolved through time. This leads to a number of observations concerning risk and safety assessments.

Generalising the concept of intervention in a Bayesian network allows us to develop a variety of causal Bayesian networks adapted to our needs. In this framework, we propose a definition of risk which seems to be relevant for a broad range of issues. We then offer simple applications of our model to specific aspects of the Fukushima accident and other nuclear safety problems. In addition to specific lessons, the analysis leads to the conclusion that a systematic approach for identifying uncertainties is needed in this area.

When applied to decision theory, our tool evolves into a dynamic decision model in which acts *cause* consequences and are causally interconnected. The model provides a causal interpretation of Savage's conceptual framework, solves some of its paradoxes and clarifies certain aspects. It leads us to considering uncertainty with regard to a problem's causal structure as the source of ambiguity in decision-making, an interpretation which corresponds to a common understanding of the precautionary principle.

Keywords : Risk, risk assessment, risk management, causality, bayesian networks, ambiguity, decision theory, Fukushima, nuclear safety, defence in depth